Practical Guidelines for the Estimation and Inference of a Dynamic Logistic Model with Fixed-Effects☆

Romain Aeberhardt

CREST

Laurent Davezies*

CREST

Abstract

This paper shows how to simply compute one of the estimators proposed in Honoré and Kyriazidou (2000), as well as its variance, through a reshaping of the original dataset that is then used in a weighted logistic regression with clustering.

Keywords: conditional logit, state dependence
JEL classification: C23, C25, J62

1. Introduction

Disentangling state dependence and unobserved heterogeneity is an important problem in econometrics. The problem is particulary tedious when the model is not linear due to the endogeneity of the initial conditions and the incidental parameters problem. In the case of a discrete variable, Honoré and Kyriazidou (2000) (HK hereafter) have proposed an estimator that is consistent under a logistic error distribution and whatever the distribution of the individual fixed effects.[1] Despite this attractive property and even if this paper is an important reference in the literature, this estimator is rarely used. Applied econometricians often prefer to use the estimator

☆We thank Xavier D'Haultfœuille, Roland Rathelot, Philippe Zamora and an anonymous referee for their attentive reading and helpful suggestions.

*Corresponding author. Centre de Recherche en Économie et Statistique (CREST), 15 Boulevard Gabriel Péri 92245 Malakoff Cedex, France. Tel: +33 141176079. Fax: +33 141176046. E-mail: laurent.davezies@ensae.fr.

[1] HK also propose a sign-based estimator relaxing the logistic assumption, but it is out of the scope of this note.

proposed in Wooldridge (2005), although it is more restrictive, probably because this estimator is easily implementable with the standard procedures available in econometric softwares and allows them to make predictions and estimate marginal effects of time invariant variables. Indeed, the HK estimator seems to require some specific programming, in particular for the inference since it is non root-N consistent. The aim of this note is to show that estimation and inference can actually be done with a simple weighted logistic regression. For that, we only use an easy reshaping of the data. We also show that when there are more than four periods, there are two natural estimators of the asymptotic variance. Monte Carlo simulations on finite samples provide evidence that one of these two estimators clearly outperforms the other: the relative quality of the variance estimators does not depend much on n, the number of individuals, but the longer the time-dimension T, the larger the gap between the two estimators.

2. Theoretical results

Following HK, let us consider the fixed effect multinomial logit model with $M \geq 2$ alternatives depending on one lag of the choice variable and k exogeneous regressors. We assume that variables are observed for a sample of individuals during $T+1$ periods ($t = 0..T$):

$$P(y_{it} = m | x_i, \alpha_i, y_{it-1} = j) = \frac{\exp(x_{mit}\beta_m + \alpha_{mi} + \gamma_{jm})}{\sum_{h=1}^{M} \exp(x_{hit}\beta_h + \alpha_{hi} + \gamma_{jh})}. \tag{1}$$

Where y_{it} is the multinomial variable, α_{im} are fixed effects, and x_{mit} are the exogeneous regressors.

HK suggest the estimator[2]

$$\left(\widehat{\beta}_n, \widehat{\gamma}_n\right) = \widehat{\theta}_n =$$
$$\arg\max_{\theta} \sum_{\substack{i=1...n \\ 1\leq t<s\leq T-1 \\ m<l}} \mathbb{1}\left\{\{y_{it}, y_{is}\} = \{m, l\}\right\} K\left(\frac{x_{it+1} - x_{is+1}}{\sigma_n}\right) \ln\left(\frac{\exp(\mathbb{1}\left\{y_{it} = m\right\} Z_{itsml}\theta)}{1 + \exp(Z_{itsml}\theta)}\right) \tag{2}$$

[2]HK use the notation $m \neq l$ in the sums. However, the comparison with the binary case shows that this must be understood as a sum over the non-ordered pairs. To be more explicit we use the notation $m < l$ instead.

with $\theta = (b', g')'$ a vector of size[3] $kM + (M-1)^2$, $x_{it} = (x_{1it}, ..., x_{Mit})$ and Z_{itsml} is the vector of covariates such that:

$$\begin{aligned} Z_{itsml}\theta = (x_{mit} - x_{mis})b_m + (x_{lis} - x_{lit})b_l + g_{y_{it-1},m} + g_{l,y_{is+1}} - g_{y_{it-1},l} - g_{m,y_{is+1}} \\ +\mathbb{1}\{s-t=1\}(g_{m,l} - g_{l,m}) + \mathbb{1}\{s-t>1\}\left(g_{m,y_{it+1}} + g_{y_{is-1},l} - g_{l,y_{it+1}} - g_{y_{is-1},m}\right) \end{aligned} \quad (3)$$

HK show the asymptotic normality and compute the asymptotic variance of $\left(\widehat{\beta}, \widehat{\gamma}\right)$ when $M = 2$ and $T = 3$. We generalize their results to any M and T here.

Theorem 2.1 (Asymptotics of the multinomial logit case[4])**.**
Under the assumption that $x_{it} - x_{it'}$ has a positive density $f_{t,t'}$ in a neighborhood of 0 for all $t \neq t'$, σ_n is a positive sequence such that $\sqrt{n\sigma_n^{4+k}} \to \sigma$ and $K(.)$ is a smooth symmetric kernel, and assumptions (A1)-(A8) detailed in a supplementary material available on the authors' webpages:

$$\sqrt{n\sigma_n^k}\left(\widehat{\theta}_n - \theta_0\right) \to \mathcal{N}\left(B, J^{-1}VJ^{-1}\right)$$

With

$h_{itsml}(\theta) = \mathbb{1}\{\{y_{it}, y_{is}\} = \{m,l\}\} \ln\left(\frac{\exp(\mathbb{1}\{y_{it}=m\}Z_{i,t,s,m,l}\theta)}{1+\exp(Z_{i,t,s,m,l}\theta)}\right)$,

$h^{(1)}_{itsml}(\theta) = \frac{\partial h_{itsml}(\theta)}{(\partial\theta)}$,

$h^{(2)}_{itsml}(\theta) = \frac{\partial^2 h_{itsml}(\theta)}{(\partial\theta\partial\theta')}$

and

$B = O\left(\sqrt{n\sigma_n^{4+k}}\right)$,

$$J = -\sum_{\substack{1\leq t<s\leq T-1 \\ m<l}} f_{t+1,s+1}(0) E\left(h^{(2)}_{itsml} | x_{it+1} = x_{is+1}\right),$$

$$V = \sum_{\substack{1\leq t<s\leq T-1 \\ m<l}} f_{t+1,s+1}(0) E\left(h^{(1)}_{itsml} h^{(1)'}_{itsml} | x_{it+1} = x_{is+1}\right) \int K^2(u)du.$$

J can be consistently estimated by

$$\widehat{J}_n = -\frac{1}{n\sigma_n^k}\sum_{i=1}^{n}\sum_{\substack{1\leq t<s\leq T-1 \\ m<l}} K\left(\frac{x_{it+1} - x_{is+1}}{\sigma_n}\right) h^{(2)}_{tsmli}(\widehat{\theta}_n).$$

---

[3]Without normalisation on the distribution of $(\alpha_m)_{1\leq m\leq M}$, only $(M-1)^2$ component of γ are identified. Without loss of generality, one can assume that $\gamma_{1m} = \gamma_{m1} = 0$ for $1 \leq m \leq M$.

[4]The proof is included in a supplementary material available on the authors' webpages.

V can be consistently estimated by

$$\widehat{V}_n = \frac{1}{n\sigma_n^k}\sum_{i=1}^{n}\left[\sum_{\substack{1\le t<s\le T-1\\ m<l}} K\left(\frac{x_{it+1}-x_{is+1}}{\sigma_n}\right) h^{(1)}_{itsml}(\widehat{\theta}_n)\right] \times \left[\sum_{\substack{1\le t<s\le T-1\\ m<l}} K\left(\frac{x_{it+1}-x_{is+1}}{\sigma_n}\right) h^{(1)}_{itsml}(\widehat{\theta}_n)'\right]$$

or by

$$\widetilde{V}_n = \frac{1}{n\sigma_n^k}\sum_{i=1}^{n}\sum_{\substack{1\le t<s\le T-1\\ m<l}} K\left(\frac{x_{it+1}-x_{is+1}}{\sigma_n}\right)^2 h^{(1)}_{itsml}(\widehat{\theta}_n)h^{(1)}_{itsml}(\widehat{\theta}_n)'.$$

This theorem calls four remarks, which are at the basis of our simple estimation procedure.

First, following HK, by considering pairwise estimation, we are led back to maximizing the weighted likelihood of a binary logit, the weights being $K\left(\frac{x_{it+1}-x_{is+1}}{\sigma_n}\right)$.

Second, the "sandwich" structure of asymptotic covariance matrix looks like the structure of a robust covariance matrix.

Third, $\widehat{V}_n$ seems to be the "natural" candidate to estimate V, since $\widehat{V}_n$ appears in the Taylor expansion of the first order condition of (2). $\widehat{V}_n$ depends on the intra-individual correlations of $x_{it+1}-x_{is+1}$ and $h^{(1)}_{itsml}$, when s and t vary. $\widetilde{V}_n$ is a naive generalization of the estimator given by HK when $T=3$ and $M=2$. It does not depend on these intra-individual correlations. Despite that, both estimators have the same limit V. Indeed, after expansion of the products appearing in $\widehat{V}_n$, the terms depending on two distinct pairs $((s,t)\neq(s',t'))$ vanish when $\sigma_n\to 0$. In Section 4, the accuracy of $\widehat{V}_n$ and $\widetilde{V}_n$ will be compared using Monte Carlo simulations.

Fourth, most statistical softwares can calculate robust standard errors with a sandwich structure that is close to the one presented here. However, the pre-implemented formulas are developed in a parametric framework for root-N consistent estimators which is not the case here (the rate of consistency depends on σ_n). Despite this, the reported confidence intervals and p-values are still valid, as detailed below.

3. Simple computation and inference

To understand our implementation strategy, let us consider a weighted binary logit in the presence of clustering. Let $i=1..n$ denote the clusters, j a unit within the cluster, w_j a weight, y_j the outcome and X_j a set of exogeneous regressors. The weighted logit estimator is given by:

$$\widehat{\theta}_n = \arg\max_{\theta} \sum_{i=1}^{n} \sum_{j \in i} w_j \ln\left(\frac{\exp(y_j X_j \theta)}{1+\exp(X_j \theta)}\right) \tag{4}$$

$\widehat{\theta}_n$ is a consistent estimator of $\theta_0 = \arg\max_\theta E\left[w \ln\left(\frac{\exp(yX\theta)}{1+\exp(X\theta)}\right)\right]$, and following Binder (1983):

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \to \mathcal{N}\left(0, Q^{-1}GQ^{-1}\right)$$

With $h_j(\theta) = \ln\left(\frac{\exp(yX\theta)}{1+\exp(X\theta)}\right)$, $h_j^{(1)}(\theta) = \frac{\partial h_j(\theta)}{\partial \theta}$, $h_j^{(2)}(\theta) = \frac{\partial^2 h_j(\theta)}{\partial\theta\partial\theta'}$, $G = E\left(\sum_{j\in i} w_j h_j^{(1)}(\theta) \sum_{j'\in i} w_{j'} h_{j'}^{(1)}(\theta)'\right)$ and $Q = E\left(\sum_{j\in i} w_j h_j^{(2)}(\theta)\right)$, which can be consistently estimated by their empirical counterparts $\widehat{G}_n = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j\in i} w_j h_j^{(1)}(\theta) \sum_{j'\in i} w_{j'} h_j^{\prime(1)}(\theta)'\right]'$ and $\widehat{Q}_n = \sum_{i=1}^{n} w_j h_j^{(2)}$.

Such a "sandwich" formula of variance is frequent in econometrics, and is often referred to as the robust estimation of standard deviation. Most of current statistical or econometrics softwares offer procedures that quickly compute such estimators and provide t-statistics, confidence intervals and p-value.

A remarkable analogy exists between Programs 2 and 4. The programs will be the same if units j in Program 4 correspond to the 5-tuple (i, s, t, m, l) such that $1 \le t < s \le T-1$ and $m < l$ and $\mathbb{1}\{y_{it} = m\}\mathbb{1}\{y_{is} = l\} + \mathbb{1}\{y_{it} = l\}\mathbb{1}\{y_{is} = m\} = 1$ in the initial Program 2.

As mentioned above, V can be estimated by two asymptotically equivalent estimators $\widehat{V}_n$ or $\widetilde{V}_n$. Following the analogy, the asymptotic variance of the weighted logit estimate can be estimated under the assumption of clustering (identified by i) or not.

The forms of the asymptotic covariance matrix are the same except for the presence of a $\frac{1}{\sigma_n^k}$ factor in the three terms of the "sandwich" formula. More precisely $\widehat{Q}_n = \sigma_n^k \widehat{J}_n$ and $\widehat{G}_n = \sigma_n^k \widehat{V}_n$ (or $\sigma_n^k \widetilde{V}_n$). As a result, the estimated variance given by the software will be equal to $\frac{1}{n}\widehat{Q}_n^{-1}\widehat{G}_n\widehat{Q}_n^{-1} = \frac{1}{n\sigma_n^k}\widehat{J}_n^{-1}\widehat{V}_n\widehat{J}_n^{-1}$ (or $\frac{1}{n\sigma_n^k}\widetilde{J}_n^{-1}\widehat{V}_n\widetilde{J}_n^{-1}$) and then, the provided test and confidence intervals will be valid if we use the pre-implemented procedure to estimate the model.

As a result, we propose the following simple procedure for the estimation and inference of the model:

1. Create a new data set, in which each line corresponds to a 5-tuple (i, t, s, m, l) such that $1 \le t < s \le T-1$, $m < l$ and $\mathbb{1}\{y_{it} = m\}\mathbb{1}\{y_{is} = l\} + \mathbb{1}\{y_{it} = l\}\mathbb{1}\{y_{is} = m\} =$

1 in the original data set. Each line contains the binary variable $Y = \mathbb{1}_{\{y_{it}=m\}}$ and the variables Z_{itsml} as in Equation 3 (see also Table 1).

2. Choose a kernel K and a bandwidth σ_n and compute the weights[5] $W = K\left(\frac{x_{it+1}-x_{is+1}}{\sigma_n}\right)$. Examples of original and transformed datasets are presented in Table 2, for the case where $M = 3$ and $T = 5$, with a gaussian kernel.
3. Use a weighted binary logit procedure (the procedure must use a robust estimator of the covariance matrix) to regress Y on Z using weights W. The absence of intercept in the regressors should be specified.
   - Specify a clustering (identified by i), to estimate V by $\widehat{V}_n$.
   - Do not specify a clustering, to estimate V by $\widetilde{V}_n$.
4. Then the estimators of the parameters and standard deviations, as well as the t-statistics and p-values are asymptotically valid.

For Steps 1 to 3, an example of a SAS program is available on the authors' webpages.

4. Monte Carlo simulations

In this section, we use the previous method to estimate a logit model with state dependence on several data generating processes.

We draw the data according to (1), with $M = 3$, one exogeneous regressor ($k = 1$), $T = 3$ or $T = 7$ and $n = 250, 1000$ or 4000. We let $x_{it} \sim \mathcal{N}(0, 1)$, $\alpha_{i1} = 0$, $\alpha_{i2} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ and $\alpha_{i3} = -sgn(\sum_{t=1}^{T} x_{it}^3)\left(\frac{1}{T}\sum_{t=1}^{T} x_{it}^3\right)^{1/3}$. State 1 is chosen as the reference, so that $\beta_1 = \gamma_{1m} = \gamma_{m1} = 0$. We set: $\beta_2 = 0.7$, $\beta_3 = 0.7$, $\gamma_{22} = 0.6$, $\gamma_{23} = 0.3$, $\gamma_{32} = 0.2$, $\gamma_{33} = 0.2$. The initial conditions are drawn following the multinomial logit model (without state dependence).

$$P(y_{i0} = m|x_i, \alpha_i) = \frac{\exp(x_{mi0}\beta_m + \alpha_{mi})}{\sum_{h=1}^{3}\exp(x_{hi0}\beta_h + \alpha_{hi})}$$

We use the optimal rate $\sigma_n = cn^{-1/(k+4)} = cn^{-1/5}$, a gaussian kernel, and choose three values (0.1, 1 and 10) for the parameter c. For $T = 3$ there is at most one pair

[5] In theory, estimates are invariant to a multiplicative change of the weights. However, in practice large weights can cause computational problems. For this reason we recommend to use $K\left(\frac{x_{it+1}-x_{is+1}}{\sigma_n}\right)$ as weight instead of $\frac{1}{\sigma_n^k}K\left(\frac{x_{it+1}-x_{is+1}}{\sigma_n}\right)$.

Table 1: Construction of the regressors in the new dataset used for the estimation and inference. State 1 is chosen as the reference, $2 \leq m < l \leq M$, $k, k' \notin \{1, m, l\}$

Coefficient	Value of the associated covariate
β_m	$x_{imt} - x_{ims}$
β_l	$x_{ils} - x_{ilt}$
β_k	0
γ_{mm}	$\mathbb{1}_{\{y_{it-1}=m\}} - \mathbb{1}_{\{y_{is+1}=m\}} + \mathbb{1}_{\{s-t>1\}}\left(\mathbb{1}_{\{y_{it+1}=m\}} - \mathbb{1}_{\{y_{is-1}=m\}}\right)$
γ_{ll}	$\mathbb{1}_{\{y_{is+1}=l\}} - \mathbb{1}_{\{y_{it-1}=l\}} + \mathbb{1}_{\{s-t>1\}}\left(\mathbb{1}_{\{y_{is-1}=l\}} - \mathbb{1}_{\{y_{it+1}=l\}}\right)$
γ_{kk}	0
γ_{lm}	$\mathbb{1}_{\{y_{it-1}=l\}} + \mathbb{1}_{\{y_{is+1}=m\}} - \mathbb{1}_{\{s-t=1\}} - \mathbb{1}_{\{s-t>1\}}\left(\mathbb{1}_{\{y_{it+1}=m\}} - \mathbb{1}_{\{y_{is-1}=l\}}\right)$
γ_{ml}	$-\mathbb{1}_{\{y_{it-1}=m\}} - \mathbb{1}_{\{y_{is+1}=l\}} + \mathbb{1}_{\{s-t=1\}} + \mathbb{1}_{\{s-t>1\}}\left(\mathbb{1}_{\{y_{it+1}=l\}} - \mathbb{1}_{\{y_{is-1}=m\}}\right)$
γ_{km}	$\mathbb{1}_{\{y_{it-1}=k\}} - \mathbb{1}_{\{s-t=1\}}\mathbb{1}_{\{y_{is-1}=k\}}$
γ_{mk}	$-\mathbb{1}_{\{y_{is+1}=k\}} + \mathbb{1}_{\{s-t=1\}}\mathbb{1}_{\{y_{it+1}=k\}}$
γ_{kl}	$-\mathbb{1}_{\{y_{it-1}=k\}} + \mathbb{1}_{\{s-t=1\}}\mathbb{1}_{\{y_{is-1}=k\}}$
γ_{lk}	$\mathbb{1}_{\{y_{is+1}=k\}} - \mathbb{1}_{\{s-t=1\}}\mathbb{1}_{\{y_{it+1}=k\}}$
$\gamma_{kk'}$	0

of periods for each individual that is used in the estimation, so $\widehat{V}_n = \widetilde{V}_n$. For $T = 7$, even if $\widehat{V}_n$ and $\widetilde{V}_n$ are asymptotically equivalent, they are different in small samples.

We reshape the data as explained above and use the SURVEYLOGISTIC procedure in SAS[6], which gives an estimation of the parameters as well as the associated inference. The number of individuals used in the estimation is close to 54% of n for $T = 3$ and increases to 95% for $T = 7$. With our method the results are obtained almost instantaneously on a basic personal computer.

The main results of the simulation are the following (see Table 3). First, the small sample bias depends only weakly on the constant c (except for small c and sample size $n = 250$) and is negligible compared to the standard deviation.

Second, the level of the tests on small samples is close to the asymptotic level for the estimation using clustering. However, when we use the estimate $\widetilde{V}_n$, the actual level of the tests is widely above their nominal level. Additional simulations (not reproduced here) show that the disturbance increases with T, or, equivalently, with the ratio of the number of pairs of periods used to the number of individual observations used.

[6]Same results can be obtained using instruction LOGIT with options PWEIGHTS and CLUSTER in Stata or using functions SVYDESIGN and SVYGLM of package SURVEY in R.

Table 2: Rearrangements with 6 time periods ($T = 5$), $M = 3$ states, $k = 1$ regressor

(a) Original Dataset

i	T	y	x
Bo	0	1	0.21
Bo	1	2	1.44
Bo	2	2	0.94
Bo	3	1	−1.25
Bo	4	3	0.13
Bo	5	2	−0.15
Ekaterini	0	2	−0.83
Ekaterini	1	1	0.10
Ekaterini	2	2	0.25
Ekaterini	3	1	0.69
Ekaterini	4	1	0.71
Ekaterini	5	3	0.21

(b) Transformed dataset

i	t	s	m	l	$Z(\beta_2)$	$Z(\beta_3)$	$Z(\gamma_{22})$	$Z(\gamma_{23})$	$Z(\gamma_{32})$	$Z(\gamma_{33})$	y	W
Bo	1	3	1	2	−2.69	0.00	0	1	0	0	0	0.4463
Bo	1	4	2	3	1.31	−1.31	0	0	0	0	1	0.3691
Bo	2	3	1	2	−2.20	0.00	−1	1	0	0	0	0.2827
Bo	2	4	2	3	0.81	−0.81	0	−1	1	0	1	0.3630
Bo	3	4	1	3	0.00	1.39	0	−1	1	0	1	0.5486
Ekaterini	1	2	1	2	0.15	0.00	−1	0	0	0	1	0.5268
Ekaterini	2	3	1	2	0.44	0.00	0	0	0	0	0	0.5641
Ekaterini	2	4	1	2	0.46	0.00	0	1	0	0	0	0.5207

$$W = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_{t+1} - x_{s+1})^2}{2\sigma_n^2}} \quad \text{with} \quad \sigma_n = 1000^{-\frac{1}{k+4}} \simeq 0.25$$

Third, the actual levels of the tests using $\widetilde{V}_n$ converge dramatically slowly to the nominal levels as n increases. For $T > 3$ using clustering for inference clearly outperforms the inference without clustering, even with substantial n.

Table 3: Estimation and inference on finite sample

Parameter	c	n	4 Periods ($T=3$) Bias	Std.Dev.	level of the test ($\widehat{V}_n = \widetilde{V}_n$)	8 Periods ($T=7$) Bias	Std.Dev.	level of the test using $\widehat{V}_n$	level of the test using $\widetilde{V}_n$
	0.1	250	0.0987	0.5377	0.077	0.0065	0.1170	0.064	0.137
	0.1	1000	0.0170	0.2106	0.074	0.0042	0.0570	0.062	0.134
	0.1	4000	0.0048	0.0948	0.045	0.0005	0.0273	0.056	0.119
	1	250	0.0437	0.2733	0.061	0.0066	0.0907	0.049	0.216
$\beta_1 = 0,7$	1	1000	0.0175	0.1261	0.052	0.0053	0.0443	0.051	0.212
	1	4000	0.0107	0.0572	0.040	0.0025	0.0215	0.048	0.205
	10	250	0.0407	0.2283	0.056	0.0079	0.0843	0.049	0.242
	10	1000	0.0213	0.1082	0.056	0.0069	0.0416	0.046	0.256
	10	4000	0.0169	0.0499	0.057	0.0046	0.0205	0.053	0.247
	0.1	250	0.1958	0.5868	0.085	0.0104	0.1214	0.056	0.124
	0.1	1000	0.0349	0.2110	0.072	0.0037	0.0614	0.058	0.143
	0.1	4000	0.0097	0.0996	0.051	0.0018	0.0286	0.045	0.116
	1	250	0.0790	0.2851	0.075	0.0067	0.0945	0.054	0.203
$\beta_2 = 0,7$	1	1000	0.0200	0.1240	0.053	0.0037	0.0482	0.067	0.224
	1	4000	0.0092	0.0619	0.057	0.0028	0.0224	0.045	0.186
	10	250	0.0666	0.2399	0.065	0.0073	0.0883	0.056	0.226
	10	1000	0.0225	0.1083	0.057	0.0050	0.0448	0.055	0.259
	10	4000	0.0143	0.0541	0.077	0.0044	0.0210	0.051	0.222
	0.1	250	0.1621	1.1644	0.100	-0.0098	0.2554	0.059	0.190
	0.1	1000	0.0048	0.4208	0.059	-0.0044	0.1257	0.062	0.192
	0.1	4000	-0.0066	0.1951	0.051	-0.0004	0.0604	0.047	0.189
	1	250	0.0246	0.5956	0.068	-0.0144	0.2090	0.048	0.290
$\gamma_{11} = 0,6$	1	1000	-0.0132	0.2602	0.046	-0.0065	0.1042	0.051	0.297
	1	4000	-0.0073	0.1282	0.043	-0.0029	0.0502	0.046	0.285
	10	250	0.0020	0.5121	0.055	-0.0165	0.1991	0.051	0.325
	10	1000	-0.0182	0.2383	0.050	-0.0089	0.0989	0.053	0.318
	10	4000	-0.0098	0.1163	0.042	-0.0058	0.0487	0.043	0.350
	0.1	250	0.0983	1.0842	0.078	-0.0040	0.2577	0.044	0.186
	0.1	1000	-0.0078	0.3991	0.055	-0.0041	0.1254	0.044	0.185
	0.1	4000	-0.0035	0.1866	0.051	-0.0011	0.0650	0.048	0.208
	1	250	0.0099	0.5719	0.054	-0.0118	0.2192	0.049	0.299
$\gamma_{12} = 0,3$	1	1000	-0.0241	0.2626	0.058	-0.0085	0.1072	0.041	0.296
	1	4000	-0.0089	0.1263	0.054	-0.0047	0.0550	0.051	0.315
	10	250	0.0002	0.4969	0.055	-0.0183	0.2094	0.050	0.335
	10	1000	-0.0257	0.2405	0.071	-0.0137	0.1032	0.047	0.346
	10	4000	-0.0122	0.1146	0.044	-0.0095	0.0523	0.047	0.352
	0.1	250	0.0418	1.0408	0.096	-0.0065	0.2516	0.043	0.181
	0.1	1000	-0.0020	0.4054	0.069	-0.0052	0.1235	0.039	0.190
	0.1	4000	-0.0041	0.1795	0.048	-0.0020	0.0625	0.055	0.196
	1	250	0.0144	0.5582	0.048	-0.0104	0.2051	0.036	0.299
$\gamma_{21} = 0,2$	1	1000	-0.0074	0.2610	0.054	-0.0056	0.1048	0.039	0.312
	1	4000	-0.0062	0.1201	0.045	-0.0027	0.0528	0.049	0.298
	10	250	0.0000	0.4893	0.045	-0.0135	0.1964	0.037	0.335
	10	1000	-0.0117	0.2352	0.054	-0.0068	0.1007	0.042	0.342
	10	4000	-0.0082	0.1103	0.049	-0.0041	0.0509	0.050	0.328
	0.1	250	0.1081	1.1154	0.085	0.0085	0.2642	0.048	0.191
	0.1	1000	0.0373	0.4117	0.051	0.0003	0.1274	0.039	0.184
	0.1	4000	0.0324	0.1879	0.051	0.0032	0.0628	0.050	0.170
	1	250	0.0486	0.5757	0.055	0.0031	0.2216	0.050	0.311
$\gamma_{22} = 0,2$	1	1000	0.0230	0.2679	0.061	-0.0007	0.1066	0.045	0.304
	1	4000	0.0194	0.1255	0.063	0.0008	0.0520	0.038	0.298
	10	250	0.0377	0.5022	0.048	-0.0039	0.2119	0.053	0.337
	10	1000	0.0159	0.2429	0.049	-0.0048	0.1019	0.049	0.331
	10	4000	0.0140	0.1145	0.049	-0.0026	0.0498	0.043	0.311

Note : Computation obtained with 1000 simulations.
The level of test reported are the estimation of the actual tests for a nominal level of 5%.

Binder, D. A., December 1983. On the variances of asymptotically normal estimators from complex surveys. International Statistical Review / Revue Internationale de Statistique 51 (3), 279–292.

Honoré, B., Kyriazidou, E., July 2000. Panel data discrete choice models with lagged dependent variables. Econometrica 68 (4), 839–874.

Wooldridge, J. M., January 2005. Simple solutions to the initial conditions problem for dynamic, nonlinear panel data models with unobserved heterogeneity. Journal of Applied Econometrics 20, 39–54.