Efficiency of repeated-cross-section estimators in fixed-effects models

Montezuma Dumangane and Nicoletta Rosati

CEMAPRE and ISEG-UTL *

January 2009

Abstract

PRELIMINARY AND INCOMPLETE Exploiting across periods restrictions implied by the fixed-effects assumption, a lower bound on asymptotic variance is obtained to compare efficiency of estimators applied to panel data and repeated cross-sections. Small-sample efficiency is studied through simulation.

Asymptotically, panel data estimation is more efficient only in case of strong residual autocorrelation; in small samples, variances are comparable, but repeated cross-sections show larger bias for some sets of parameter values.

Keywords: asymptotic lower bound; conditional moment restrictions; fixed-effects model; repeated cross-sections.

JEL Classification: C13, C15.

*Corresponding author: Nicoletta Rosati, Mathematics Departament, ISEG, Technical University of Lisbon, Rua do Quelhas 6, 1200 - 781 Lisbon, Portugal. Email: nicoletta@iseg.utl.pt.

1 Introduction

The value of repeated cross-sections (RCS) to identifiability of structural parameters in fixed-effects models is known since Heckman and Robb (1985). At the same time, Deaton (1985) stressed that a measurement error problem, peculiar to RCS inference, arises in finite samples due to the need to estimate some nuisance parameters. He also showed that the resulting bias can be dealt with exploiting sample information. A line of research originated out of these seminal papers; among several contributions, see, for instance, Verbeek (1992), Moffit (1993), Verbeek and Nijman (1993), Collado (1997), Girma (2000), McKenzie (2004) and Verbeek and Vella (2005).

A still unsettled issue to our knowledge is the derivation of a benchmark for the amount of precision we might expect in an RCS based inference, which would also be useful to compare the potential of RCS to that of genuine panel information. Heckman and Robb (1985) made an informal contention that even if panel estimators are often claimed to be asymptotically more efficient than the RCS ones, in finite samples this might not happen since sample size is by far larger in cross-sectional surveys.

In this paper we derive a lower bound on the asymptotic variance of a RCS estimator by exploiting a set of moment restrictions implied by the fixed-effects assumption. Under suitable regularity conditions, we allow for a general nonlinear model, provided that unobservables enter the model additively. The lower bound is obtained by applying standard results of inference under conditional moment restrictions (see, for instance, Chamberlain, 1987). Small sample behaviour is studied through simulation.

2 The model

Applied sciences often deal with panel data, namely a vector of variables z_{it} observed at times $t = 1, \ldots, T$ on the units $i = 1, \ldots, N$. One of the advantages of panel data is the possibility to control for unobserved heterogeneity among individuals due to time-invariant effects, since repeated observations on each sampling unit are available. When panel data are not available, identification and estimation of the parameters is still

possible - under suitable conditions - based on repeated cross-sections, where sample averages are computed from individual observations within pre-defined time-invariant classes (cohorts), playing the role of "macro-individuals" in a pseudo-panel dataset. Estimating techniques are basically the same in both cases. In the case of panel data they are applied directly to individual observations, while in the case of repeated cross-sections they are applied to the class sample averages. The sample averages are considered as error-ridden measurements of the true class means, and therefore the estimators are corrected for the presence of response and covariate measurement error. When the sample size for each class is sufficiently large, the measurement error can be neglected, i.e. no correction is necessary.

To simplify notation, from now on let i indicate the sampling unit or the synthetic macro-individual, according to the sampling scheme (panel or pseudo-panel). For instance, y_i would indicate the response for individual i in a panel, but represents the sample average of the response for class i in a pseudo-panel (i-th macro-individual), the only difference being the presence of (possibly negligible) measurement error in the second case. Consider models where the unobservable component has the form:

$$u_{it} = g(z_{it}; \beta) \tag{1}$$

where $g(\ .\ )$ is known, $z_{it} = (y_{it}, x_{it})$ where y_{it} is the response variable and x_{it} is a vector of explanatory variables, and $E(u_{it}|x_{it}) = 0$. Our interest is in estimation of β, which is both time- and individual-invariant. The error u_{it} is assumed to be such that we can represent it as:

$$\begin{gathered} u_{it} = \eta_i + v_{it} \\ E\{u_{it}|w_i, x_{it}\} = 0 \qquad \forall w_i \in \mathcal{W} \end{gathered} \tag{2}$$

where η_i is an unobservable individual effect, w_i is a vector of observable time-invariant variables and $\mathcal{W}$ is its support. In the case of multiplicative effects other transformations can be used in order to eliminate the individual effect.

Notice that in the case of RCS the individual effects depend on time too, i.e. $\eta_i = \eta_{i(t)}$, being the sample average of the effects for the specific cohort sample at time t. However, its conditional expectation remains time-invariant within cohort i,

i.e. $E\{\eta_{i(r)}|w_i\} = E\{\eta_{i(s)}|w_i\}$, $\forall r \neq s$. This, together with assumption (2), implies among other things that for two time periods $r \neq s$ the following condition holds

$$E\{g(z_{ir};\beta) - g(z_{is};\beta)|w_i, x_{ir}, x_{is}\} = 0 \tag{3}$$

This equation specifies a conditional moment restriction upon which inference on β can rest. The lower bound we shall derive exploits such restriction.

3 Asymptotic efficiency

Suppose one is choosing whether to collect a panel data set or a sequence of cross-sections. Since sampling costs are rather different, it is often the case that the panel sample is of much smaller size than the cross-sectional samples (which we assume - for the sake of simplicity - to be all of the same size). Therefore, we aim at answering the following question: to estimate the parameter of interest, is it more convenient to use panel or RCS information?

An answer to this question can be obtained comparing the optimal asymptotic variance attainable in the two cases; such a comparison can be derived directly from the results in Chamberlain (1987), which we briefly recall here.

Consider a random sample of size N from variables z and w, a parameter β and a function $m(z,\beta)$ such that the following moment condition holds: $E\{m(z,\beta_0)|\, w\} = 0$; suppose also that function m satisfies the following regularity conditions:

(i) *β is in an open set $\mathcal{B} \subset \mathbb{R}^p$ such that $m(z,\beta)$ and $\partial m(z,\beta)/\partial\beta^\top$ are continuous for $(z,\beta) \in \mathcal{Z} \times \mathcal{B}$.*

(ii) *$E\{m(z,\beta_0)|w\} = 0 \;\; \forall \;\; w \in \mathcal{W}$.*

(iii) *$\Sigma(w) = E\{\; m(z,\beta)\; m(z,\beta)^\top|w\}$ is positive-definite for all $w \in \mathcal{W}$.*

(iv) *Let $D(w) = E\{\; \partial m(z,\beta)/\partial\beta^\top|w\}$ for $w \in \mathcal{W}$. The matrix $E\{D^\top(w)\Sigma^{-1}(w)D(w)\}$ is positive-definite.*

Then a lower bound on asymptotic variance for any regular consistent asymptotically normal estimator is given by

$$\Lambda = \left\{ \ E \ \left[D^{\top}(w)\Sigma^{-1}(w)D(w) \right] \ \right\}^{-1} \tag{4}$$

where matrices $D(w)$ and $\Sigma(w)$ are evaluated at $\beta = \beta_0$.

To apply Chamberlain's results to our case, consider the following function

$$m(z_i, \beta) = g(z_{i1}; \beta) - g(z_{i2}; \beta) \tag{5}$$

Equation (3) guarantees that this $m(z, \beta_0)$ satisfies the moment condition above: in the case of panel data this is true at individual level, while observing repeated cross-sections this holds at cohort level (macro-individuals). It follows that a lower bound on asymptotic variance of GMM estimators based on this moment condition can be obtained from (4), if the appopriate regularity conditions hold.

It is easy to verify that if the function m defined in (5) satisfies conditions (i)-(iv) above for panel data, then this happens for RCS too, so that the asymptotic lower bound can be applied to our problem. Asumptotics here refers to large N, where N represents the number of individuals in the panel data and the number of cohorts in the repeated cross-sectional data, respectively.

Substituting function m into the definition of $D(w)$ and $\Sigma(w)$, we obtain:

$$D(w_i) = E[\partial g(z_{i1}, \beta)/\partial \beta' | w_i] - E[\partial g(z_{i2}, \beta)/\partial \beta' | w_i]$$

and

$$\Sigma(w_i) = \mathrm{Var}\{ g(z_{i1}, \beta) - g(z_{i2}, \beta) |\ w_i \ \} = \mathrm{Var}\{ u_{i1} - u_{i2} |\ w_i \ \}$$

To specify the expression of the lower bound in the two sampling schemes, recall that in the case of RCS the variables represented the cohort sample average. Making this explicit, this yields:

$$\begin{aligned} D(w_i) &= E\left[\frac{\partial \bar{g}(z_{i1}, \beta)}{\partial \beta'} \middle| w_i \right] - E\left[\frac{\partial \bar{g}(z_{i2}, \beta)}{\partial \beta'} \middle| w_i \right] \\ &= E\left[\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(z_{i1}, \beta)}{\partial \beta'} \middle| w_i \right] - E\left[\frac{1}{N} \sum_{i=1}^{N} \frac{\partial g(z_{i2}, \beta)}{\partial \beta'} \middle| w_i \right] \\ &= E\left[\frac{\partial g(z_{i1}, \beta)}{\partial \beta'} \middle| w_i \right] - E\left[\frac{\partial g(z_{i2}, \beta)}{\partial \beta'} \middle| w_i \right] \end{aligned}$$

This expression coincides with the case of panel data, so that it is not affected by the sampling scheme. Consider now the difference $u_{i1} - u_{i2}$ appearing in the definition of $\Sigma(w_i)$. From assumption (2), in the case of panel data $u_{i1} - u_{i2} = v_{i1} - v_{i2}$ so that in case of homoskedasticity of the residuals

$$\begin{aligned}\Sigma_P(w_i) &= \mathrm{Var}(v_{i1}|w_i) + \mathrm{Var}(v_{i2}|w_i) - 2\,\mathrm{Cov}(v_{i1}, v_{i2}|w_i) \\ &= 2[\mathrm{Var}(v_i|w_i) - \mathrm{Cov}(v_{i1}, v_{i2}|w_i)]\end{aligned}$$

In the case of RCS, instead, $\bar{u}_{i1} - \bar{u}_{i2} = \bar{v}_{i1} - \bar{v}_{i2} + \bar{\eta}_{i1} - \bar{\eta}_{i2}$. This implies that, if v is orthogonal to the other variables in the model

$$\begin{aligned}\Sigma_{RCS}(w_i) &= \mathrm{Var}(\bar{v}_{i1}|w_i) + \mathrm{Var}(\bar{v}_{i2}|w_i) - 2\,\mathrm{Cov}(\bar{v}_{i1}, \bar{v}_{i2}|w_i) + \\ &\quad + \mathrm{Var}(\bar{\eta}_{i1}|w_i) + \mathrm{Var}(\bar{\eta}_{i2}|w_i) - 2\,\mathrm{Cov}(\bar{\eta}_{i1}, \bar{\eta}_{i2}|w_i) \\ &= \frac{2}{n^*}[\mathrm{Var}(v_i|w_i) + \mathrm{Var}(\eta_i|w_i)]\end{aligned}$$

where $n^* = 2n_{i1}n_{i2}/(n_{i1} + n_{i2})$, being n_{it} the sample size at time t for cohort i. If a cohort has always the same number of sampled individuals at all times, then this coincides with n^*. Notice that all covariances are zero, as the samples in the two time periods are independent.

In the linear case $g(z_{it}, \beta) = y_i - x'_{it}\beta$, so that $m(z_i, \beta) = (y_{i1} - y_{i2}) - (x_{i1} - x_{i2})'\beta$ and $\partial g(z_{it}, \beta)/\partial\beta' = -x'_{it}$. Quantity $\Sigma(w_i)$ is the same as in the general case, while $D(w_i) = E\{x'_{i2}|\ w_i\ \} - E\{x'_{i1}|\ w_i\ \} = E[\Delta x_i|w_i]$, where $\Delta x_i = x_{i2} - x_{i1}$. Note that, as $E[\Delta x_i|w_i]$ gets closer to zero (non-identifiable model), the variance grows larger and larger.

Analysing the expression for $\Sigma(w_i)$ is enough to compare the asymptotic lower bound under the two sampling schemes, since the term $D(w_i)$ is not affected. The lower bound for panel data is larger than that for RCS iff:

$$\mathrm{Var}(v_i|w_i) - \mathrm{Cov}(v_{i1}, v_{i2}|w_i) \geq \frac{1}{n^*}[\mathrm{Var}(v_i|w_i) + \mathrm{Var}(\eta_i|w_i)]$$

that is iff

$$\mathrm{Var}(v_i|w_i)\left(1 - \frac{1}{n^*}\right) \geq \mathrm{Cov}(v_{i1}, v_{i2}|w_i) + \frac{1}{n^*}\,\mathrm{Var}(\eta_i|w_i)$$

If the sample size for each cohort is sufficiently large, then the terms multiplied by $1/n^*$ become negligible, showing that panel is superior to RCS only if the serial correlation of the residuals v_{it} is positive and sufficiently large ($\text{Var}(v_i|w_i)/\text{Cov}(v_{i1}, v_{i2}|w_i) < 1$). Notice that this is never the case for both the $AR(1)$ and the $MA(1)$ specifications for v_{it}. In case of uncorrelated residuals, RCS is clearly better, showing that there is no potential efficiency gain using panel data.

The cases of small n^* with large N, and small n^* with small N are analysed through simulation, as shown in the next Section.

4 Simulation study

A simulation study has been conducted with the following data generating process:

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + v_{it} \quad i = 1, ..., N \;\; t = 1, 2$$

Let $y_{i0} = 0$ and $\eta_i \sim N(0, \sigma_\eta^2)$. The error follows an AR(1) model specified as $v_{it} = \theta_0 + \theta_1 v_{it-1} + \xi_{it}$, where $\xi_{it} \sim N(0, 1)$. The regressors are also AR(1) and are correlated with the individual effects, the generating process being specified as $x_{it} = \rho x_{it-1} + \gamma\eta_i + \varepsilon_{it}$, with $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ independent of η_i.

4.1 Experimental design

Denote the average number of individuals within each cohort by $\bar{n} = N^{-1}\sum_{i=1}^{N} n_{i1} = N^{-1}\sum_{i=1}^{N} n_{i2}$, where N is the number of cohorts (equal in period one and two) and n_{it} is the number of individuals sampled from cohort i in period t. Several cases are considered, for $N \in \{50, 100, 150, 200, 300\}$ and $\bar{n} \in \{8, 16, 24\}$.

The results are compared to panel data where $\bar{n} = 1$, that is a panel data-set with a number of individuals equal to the number of cohorts in the RCS data-set. Simulations were not performed for combinations where the total number of RCS observations $\bar{n} * N$ was very large (> 2400).

At each simulation, $\bar{n} * N$ observations are generated, split into cohorts according to a time invariant variable, and the cohort means are computed.

Table 1: Parameter values chosen for simulation and their influence on the variances and covariances of some design variables, for $\rho = 0.5$.

θ_1	σ_η^2	γ	**Var(x)**	**Var(y)**	**Corr(x, η)**	**Var(v)**	$\mathbf{r_1}^*$	$\mathbf{r_2}^*$	$\mathbf{R}^{2*}$
0.5	0.1	0.0	1.47	2.90	0.00	1.33	0.07	0.06	0.51
0.1	0.1	0.0	1.47	2.58	0.00	1.01	0.09	0.06	0.57
0.1	0.5	0.0	2.00	3.51	0.00	1.01	0.33	0.20	0.57
0.1	0.1	0.1	1.49	2.64	0.05	1.01	0.09	0.06	0.56
0.5	0.1	0.1	1.49	2.97	0.05	1.33	0.07	0.06	0.50
0.5	0.1	0.5	4.00	8.33	0.50	1.33	0.43	0.14	0.48
0.1	0.1	0.5	1.60	2.91	0.25	1.01	0.09	0.05	0.55
0.1	0.5	0.1	2.13	3.84	0.10	1.01	0.33	0.18	0.56

* $r_1 = \sigma_\eta^2/(\mathrm{Var}(v) + \sigma_\eta^2)$, $r_2 = \sigma_\eta^2/\mathrm{Var}(\beta x + \eta)$, $R^2 = \mathrm{Var}(\beta x)/\mathrm{Var}(y)$.

Table 2: Parameter values chosen for simulation and their influence on the variances and covariances of some design variables, for $\rho = 0$.

θ_1	σ_η^2	γ	**Var(x)**	**Var(y)**	**Corr(x, η)**	**Var(v)**	$\mathbf{r_1}^*$	$\mathbf{r_2}^*$	$\mathbf{R}^{2*}$
0.1	0.1	0.1	1.10	2.23	0.03	1.01	0.09	0.08	0.49
0.5	0.1	0.1	1.10	2.55	0.03	1.33	0.07	0.08	0.43
0.5	0.1	0.5	1.10	2.63	0.15	1.33	0.07	0.08	0.42
0.1	0.1	0.5	1.10	2.31	0.15	1.01	0.09	0.08	0.48
0.1	0.5	0.1	1.50	3.11	0.06	1.01	0.33	0.24	0.48

* $r_1 = \sigma_\eta^2/(\mathrm{Var}(v) + \sigma_\eta^2)$, $r_2 = \sigma_\eta^2/\mathrm{Var}(\beta x + \eta)$, $R^2 = \mathrm{Var}(\beta x)/\mathrm{Var}(y)$.

Estimation of β is then performed using the difference-based moment condition (3), where the g function is now given by $g(z_{it}; \beta) = y_{it} - \alpha y_{it-1} - \beta x_{it}$. Notice that this can be extended to estimators based on conditions in levels if $\gamma = 0$, since in this case they are more efficient.

The parameter of interest is β, which is set equal to one. Some of the other parameters are held fixed, with values respectively $\alpha = 0$, $\theta_0 = 1$ and $\sigma_\varepsilon^2 = 1$. Various sets of studies are performed, according to different values for θ_1, σ_η^2, ρ and γ as reported in Tables 1 and 2.

For the case of panel data, the first-differences OLS (FD-OLS) estimator is consistent for any set of parameter values. For RCS data, FD-OLS is inconsistent if both

Table 3: Simulation results for 100 replications, with $\rho = 0.5$, $\theta_1 = 0.5$, $\sigma_\eta^2 = 0.1$ and $\gamma = 0$. True value $\beta = 1$; standard errors in parentheses.

		$\bar{\mathbf{n}}$			
		1	**8**	**16**	**24**
N	**50**	1.0062 (0.1148)	0.9961 (0.1639)	0.9977 (0.1608)	0.9812 (0.1424)
	100	0.9888 (0.0728)	0.9889 (0.1085)	1.0035 (0.1091)	1.0025 (0.1068)
	150	0.9932 (0.0542)	0.9928 (0.0897)	0.9939 (0.0838)	–
	200	1.0019 (0.0524)	1.0050 (0.0829)	–	–
	300	1.0049 (0.0418)	1.0022 (0.0615)	–	–

γ and ρ are non-zero, although the bias is attenuated for smaller γ. In this case, efficiency should be compared using the mean squared error. Consistent estimates can be obtained through FD-IV estimation. If either $\rho = 0$ or $\gamma = 0$, FD-OLS becomes consistent for RCS, although a levels estimator would be more efficient for the second case.

4.2 Results

The first set of results refers to the case of non-zero ρ, simulated for the specific value $\rho = 0.5$. For the case of $\gamma = 0$ (see Tables 3–5) the panel and RCS estimators show comparable efficiency. This is not the case when we consider $\gamma \neq 0$ (Tables 6–10). This last case yielded biased estimates for the RCS case, the bias reaching values aroud 7-8%; however, the bias tended to disappear for small values of γ (for instance $\gamma = 0.1$). Comparison based on MSE shows superiority of panel data for this specific case.

In general, increasing the number of observations within each cohort does not improve efficiency, for a fixed number of cohorts, while increasing the number of cohorts helps reducing the variance.

Simulations for the case of $\rho = 0$ are yet to be performed, as well as consistent IV estimation for the case of $\rho \neq 0$ and $\gamma \neq 0$. Efficient levels estimation for the case of $\gamma = 0$ should also be performed.

Table 4: Simulation results for 100 replications, with $\rho = 0.5$, $\theta_1 = 0.1$, $\sigma_\eta^2 = 0.1$ and $\gamma = 0$. True value $\beta = 1$; standard errors in parentheses.

		$\bar{\mathbf{n}}$			
		1	**8**	**16**	**24**
	50	1.0078 (0.1358)	0.9967 (0.1425)	0.9985 (0.1502)	0.9744 (0.1289)
	100	0.9910 (0.0874)	0.9918 (0.1014)	1.0075 (0.0961)	0.9970 (0.1002)
N	**150**	0.9942 (0.0632)	0.9952 (0.0856)	0.9935 (0.0735)	–
	200	1.0041 (0.0644)	1.0069 (0.0697)	–	–
	300	1.0072 (0.0488)	0.9986 (0.0579)	–	–

Table 5: Simulation results for 100 replications, with $\rho = 0.5$, $\theta_1 = 0.1$, $\sigma_\eta^2 = 0.5$ and $\gamma = 0$. True value $\beta = 1$; standard errors in parentheses.

		$\bar{\mathbf{n}}$			
		1	**8**	**16**	**24**
	50	1.0078 (0.1358)	0.9926 (0.1698)	1.0002 (0.1712)	0.9762 (0.1492)
	100	0.9910 (0.0874)	0.9916 (0.1189)	1.0075 (0.0961)	0.9910 (0.1152)
N	**150**	0.9942 (0.0632)	0.9952 (0.0856)	0.9867 (0.0853)	–
	200	1.0041 (0.0644)	1.0050 (0.0819)	–	–
	300	1.0072 (0.0488)	0.9995 (0.0666)	–	–

Table 6: Simulation results for 100 replications, with $\rho = 0.5$, $\theta_1 = 0.1$, $\sigma_\eta^2 = 0.1$ and $\gamma = 0.1$. True value $\beta = 1$; standard errors in parentheses.

		$\bar{\mathbf{n}}$			
		1	**8**	**16**	**24**
	50	1.0078 (0.1359)	1.0115 (0.1420)	1.0134 (0.1504)	0.9810 (0.1291)
	100	0.9910 (0.0874)	1.0062 (0.1014)	1.0219 (0.0963)	1.0105 (0.1011)
N	**150**	0.9942 (0.0631)	1.0090 (0.0850)	1.0081 (0.0731)	–
	200	1.0041 (0.0644)	1.0215 (0.0692)	–	–
	300	1.0072 (0.0488)	1.0132 (0.0578)	–	–

Table 7: Simulation results for 100 replications, with $\rho = 0.5$, $\theta_1 = 0.5$, $\sigma_\eta^2 = 0.1$ and $\gamma = 0.1$. True value $\beta = 1$; standard errors in parentheses.

		$\bar{\mathbf{n}}$			
		1	**8**	**16**	**24**
	50	1.0062 (0.1148)	1.0102 (0.1629)	1.0124 (0.1614)	0.9971 (0.1428)
	100	0.9887 (0.0728)	1.0032 (0.1085)	1.0184 (0.1091)	1.0160 (0.1076)
N	**150**	0.9932 (0.0542)	1.0064 (0.0884)	1.0086 (0.0838)	–
	200	1.0019 (0.0524)	1.0198 (0.0825)	–	–
	300	1.0049 (0.0418)	1.0167 (0.0612)	–	–

Table 8: Simulation results for 100 replications, with $\rho = 0.5$, $\theta_1 = 0.5$, $\sigma_\eta^2 = 0.1$ and $\gamma = 0.5$. True value $\beta = 1$; standard errors in parentheses.

		$\bar{\mathbf{n}}$			
		1	**8**	**16**	**24**
	50	1.0063 (0.1148)	1.0627 (0.1542)	1.0667 (0.1609)	1.0566 (0.1416)
	100	0.9886 (0.0729)	1.0569 (0.1060)	1.0731 (0.1059)	1.0660 (0.1077)
N	**150**	0.9933 (0.0542)	1.0572 (0.0815)	1.0636 (0.0812)	–
	200	1.0020 (0.0524)	1.0744 (0.0787)	–	–
	300	1.0048 (0.0417)	1.0701 (0.0587)	–	–

Table 9: Simulation results for 100 replications, with $\rho = 0.5$, $\theta_1 = 0.1$, $\sigma_\eta^2 = 0.1$ and $\gamma = 0.5$. True value $\beta = 1$; standard errors in parentheses.

		$\bar{\mathbf{n}}$			
		1	**8**	**16**	**24**
	50	1.0078 (0.1360)	1.0665 (0.1360)	1.0684 (0.1468)	1.0488 (0.1272)
	100	0.9909 (0.0875)	1.0599 (0.0984)	1.0219 (0.0963)	1.0614 (0.1010)
N	**150**	0.9943 (0.0631)	1.0603 (0.0803)	1.0631 (0.0702)	–
	200	1.0042 (0.0645)	1.0752 (0.0655)	–	–
	300	1.0071 (0.0488)	1.0669 (0.0553)	–	–

Table 10: Simulation results for 100 replications, with $\rho = 0.5$, $\theta_1 = 0.1$, $\sigma_\eta^2 = 0.5$ and $\gamma = 0.1$. True value $\beta = 1$; standard errors in parentheses.

		$\bar{\mathbf{n}}$			
		1	**8**	**16**	**24**
N	**50**	1.0078 (0.1359)	1.0676 (0.1675)	1.0712 (0.1715)	1.0486 (0.1477)
	100	0.9910 (0.0874)	1.0619 (0.1174)	1.0812 (0.1177)	1.0605 (0.1149)
	150	0.9942 (0.0632)	1.0631 (0.0930)	1.0587 (0.0832)	–
	200	1.0041 (0.0644)	1.0766 (0.0798)	–	–
	300	1.0071 (0.0488)	1.0711 (0.0664)	–	–

5 Concluding remarks

It is well known that estimation of models with unobservable individual-specific effects is possible not only if repeated measurements on sampling units are available, but also with data collected in repeated cross-sections. This latter alternative, already investigated in the case of linear models, has been extended in this paper to more general models imposing only that unobservables enter the model additively.

The main result we obtain is a lower bound on the asymptotic variance of RCS estimators. This is derived from known results on estimation with conditional moment restrictions, which we build on to take into account that available information comes from two (or more) independent samples. Using the bound, asymptotic efficiency of RCS and panel data estimators is compared, showing that panel data are more efficient only in case of strong residual autocorrelation; a simulation study is performed in order to study the finite sample behaviour, finding comparable variances, but larger bias in repeated cross-sections for some sets of parameter values.

6 Acknowledgements

This study was supported in part by the Fundação para Ciência e a Tecnologia, program POCI, partially funded by FEDER.

7 References

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, **34**, 305-334.

Collado, M. D. (1997). Estimating dynamic models from time series of independent cross-sections. *Journal of Econometrics*, **82**, 37-62.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, **30**, 109-126.

Girma, S. (2000). A quasi-differencing approach to dynamic modelling from a time series of independent cross-sections. *Journal of Econometrics*, **98**, 365-383.

Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions. In *Longitudinal Analysis of Labor Market Data*, (eds. J. J. Heckman and B. Singer), Cambridge University Press, New York, 156-245.

McKenzie, D. J. (2004). Asymptotic theory for heterogeneous dynamic pseudo-panels. *Journal of Econometrics*, **120**, 235-262.

Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, **59**, 99-123.

Verbeek, M. (1992). Pseudo panel data. In *The Econometrics of Panel Data*, (eds. L. Mátyás and P. Sevestre), Kluwer Academic Publishers, Dordrecht, 303-315.

Verbeek, M. and Nijman, T. (1993). Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections. *Journal of Econometrics*, **59**, 125-136.

Verbeek, M. and Vella, F. (2005). Estimating dynamic models from repeated cross-sections. *Journal of Econometrics*, **127**, 83-102.