Lavoro e disoccupazione: questioni di misura e di analisi

Progetto di ricerca cofinanziato dal Ministero per l'Università
e la Ricerca Scientifica e Tecnologica - Assegnazione: 1998
Coordinatore: Ugo Trivellato

A lower bound on asymptotic variance of repeated cross-sections estimators in fixed-effects models

Nicoletta Rosati, Enrico Rettore, Guido Masarotto
Dip. di Scienze Statistiche, Univ. di Padova

Working Paper n. 2 agosto 1998

Unità locali del progetto:

Dip. di Economia Politica, Univ. Di Modena (coord. Michele Lalla)
Dip. di Economia "S. Cognetti De Martiis", Univ. di Torino (coord. Bruno Contini)
Dip. Di Statistica, Univ "Ca' Foscari" di Venezia (coord. Tommaso Di Fonzo)
Dip. di Metodi Quantitativi, Univ. di Siena (coord. Achille Lemmi)
Dip. di Scienze Statistiche, Univ. di Padova (coord. Ugo Trivellato)

Dip. di Scienze Statistiche
via S. Francesco 33, 35121 Padova

A LOWER BOUND ON ASYMPTOTIC VARIANCE OF REPEATED CROSS-SECTIONS ESTIMATORS IN FIXED-EFFECTS MODELS

Nicoletta Rosati   Enrico Rettore   Guido Masarotto

Dipartimento di Scienze Statistiche
Università di Padova
Via S. Francesco, 33
35121 Padova (Italy)

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

JEL classification: C23, C14.

ABSTRACT

In this paper fixed-effects models are investigated. The class of models considered is very general and includes as a special case the linear one, the only crucial assumption being that the standard fixed-effects assumption applies with unobservables additively entering the model. A lower bound on asymptotic variance is derived for the case of two periods of observation. The bound is obtained exploiting across periods conditional moment restrictions implied by the fixed-effects assumption. If the distribution is multinomial with known support, the problem becomes parametric and the bound can be obtained by straightforwardly modifying standard results. Since a general distribution satisfying the restrictions can always be approximated by a multinomial distribution satisfying the same restrictions, the multinomial bound applies to the general case. The bound allows to compare asymptotic efficiency of longitudinal data to that of repeated cross-sections.

1 Introduction

The value of repeated cross-sections (RCS) to identifiability of structural parameters in fixed-effects models is known since Heckman, 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 (see the review in Verbeek, 1992).

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, against which to check the asymptotic behavior of actual estimators. As usual, checking against such standard is nothing but a starting point, albeit crucial, to devise an estimator suiting the particular problem.

In particular, such benchmark would be useful to compare the potential of RCS to that of genuine longitudinal information. Heckman, Robb (1985) made an informal contention that even if longitudinal estimators outperform RCS ones in terms of asymptotic variance, in finite samples this needs not happen since sample size is by far larger in cross-sectional surveys. To give content to this intuition one needs to know the optimal asymptotic precision attainable exploiting both types of information.

In this paper we derive a lower bound on the asymptotic variance of a RCS estimator for the parameters of a fixed-effects model by exploiting a set of restrictions on conditional moments implied by the fixed-effects assumption. We do not confine ourselves to the standard linear-in-the-parameters model. Instead, under suitable regularity conditions, we allow for a general nonlinear model as well as for general heteroskedastic unobservables, provided that unobservables enter the model additively. The lower bound is obtained by properly

modifying the line of reasoning in Chamberlain (1987).

As a by-product, our results allow comparing RCS to longitudinal estimators conditional on few key parameters.

2 The model

Applied sciences often deal with **longitudinal data**, namely a vector of variables v_{it} observed at times $t = 1, \ldots, T$ on the units $i = 1, \ldots, n$. The structure of data is similar to both time series and cross-sections, but clearly includes wider information and therefore allows for more complex analyses. One of the advantages of longitudinal 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.

In this paper we shall focus on models where the unobservable component has the form:

$$\varepsilon_{it} = \eta(v_{it}; \theta) \quad , \tag{1}$$

$\eta(\ .\ )$ known. Our interest is in estimation of θ, which is both time- and individual-invariant. ε_{it} is assumed to be such that we can represent it as:

$$\begin{aligned} \varepsilon_{it} &= \alpha_i + \xi_{it} \quad , \\ E\{\xi_{it}|w_i\} &= 0 \qquad \forall \ \ w \in \mathcal{W} \ \ , \end{aligned} \tag{2}$$

where w_i is a vector of observable time-invariant variables and $\mathcal{W}$ is its support; clearly, $E\{\varepsilon_{it}|w_i\}$ turns out to be time-invariant as a consequence of such assumptions. A special case emerges if longitudinal information is available; in such case one can choose w_i to be the identity of each unit. As a result $E\{\varepsilon_{it}|w_i\} = \alpha_i$.

When longitudinal data are not available, identification and estimation of θ is still possible - under suitable conditions - based on **repeated cross-sections**. In this case the role played by repeated observations on each sampling unit is replaced by observations on the auxiliary time-invariant variables, that

allow - in a sense - to match units coming from different samples to control for unobserved heterogeneity across individuals.

Estimating techniques are basically the same in both cases. In the case of longitudinal data they are applied directly to individual observations; in the case of repeated cross-sections, instead, the sampling units in each cross-section are previously grouped into classes according to the values of the auxiliary time-invariant variables, and then the class means are treated as individual observations yielding a pseudo-longitudinal data set, to which the techniques known for longitudinal data can be applied.

2.1 Linear models

Specification (1)-(2) allows for a variety of models, among which the linear one. This last arises when $\eta(\, . \,)$ is linear with respect to θ:

$$\eta(v_{it};\theta) = y_{it} - \theta^\top x_{it} \quad , \tag{3}$$

with θ and x_{it} conformable.

Linear models have been widely studied in this context, both in the case of longitudinal data and of repeated cross-sections. For longitudinal data see, among others, Hsiao (1986) and Baltagi (1995); for repeated cross-sections see Deaton (1985) and Verbeek (1992).

2.2 Restrictions on conditional moments

Specification (1)-(2) places across-periods restrictions on the first order moments of ε_{it} conditional on w_i. In fact, according to the model we know that:

$$\begin{aligned} E\{\eta(v_{it};\theta)|w_i\} &= E(\varepsilon_{it}|w_i) \qquad (4) \\ &= E(\alpha_i|w_i) + E(\xi_{it}|w_i) \\ &= E(\alpha_i|w_i) \ . \end{aligned}$$

Consider two periods r and s, $r \neq s$. Identity (4) implies:

$$E\{\eta(v_{lr};\theta)|w\} = E\{\eta(v_{js};\theta)|w\} \tag{5}$$

for any $l(r)$ and $j(s)$ such that $w_{l(r)} = w_{j(s)} = w$, where $i(t)$ denotes i-th unit in t-th period. Since it is clear that in (5) the index of individuals is not important, we can omit it. Moreover, from now on we shall focus on the case $T = 2$. Then, (5) becomes:

$$E\{\eta(v_1;\theta)|w\} = E\{\eta(v_2;\theta)|w\} \quad , \quad w \in \mathcal{W} \ , \tag{6}$$

that is the regression of $\eta(\ .\ )$ on w is the same in both periods. Equation (6) specifies a set of conditional moment restrictions upon which inference on θ can rest. The lower bound we shall derive exploits such restrictions.

Example 1 (Repeated cross-sections) Consider the case of linear regression with data from two repeated cross-sections. At individual level the following identity holds:

$$y_{it} = \theta^\top x_{it} + \varepsilon_{it} \qquad \begin{array}{c} i = 1, \ldots, n \ , \\ t = 1, 2 \ , \end{array} \tag{7}$$

where $\varepsilon_{it} = \alpha_i + \xi_{it}$ is the unobservable component.

If we take expectations on both sides of (7) conditional on the auxiliary time-invariant set of variables, w, we obtain:

$$E\{y_{it}|w_i\} = \theta^\top E\{x_{it}|w_i\} + E\{\alpha_i|w_i\} + E\{\xi_{it}|w_i\} \qquad \begin{array}{c} i = 1, \ldots, n \ , \\ t = 1, 2 \ . \end{array}$$

Since $E\{\varepsilon_{it}|w_i\}$ is time-invariant, then:

$$E\{y_{i1} - \theta^\top x_{i1}|w_i = w\} = E\{y_{j2} - \theta^\top x_{j2}|w_j = w\} \tag{8}$$

for any unit i in the first sample and unit j in the second such that $w_i = w_j = w$, for any value w in $\mathcal{W}$.

Hence, with obvious notation:

$$E(y_1|w) - E(y_2|w) = \theta^\top \{E(x_1|w) - E(x_2|w)\} \ .$$

As it is clear, estimation of regressions of y and x on w in the two periods allows to estimate θ. $\square$

Notice that, if w is continuous and is arbitrarily discretised to give rise to the so called cohorts, grouping individuals by cohorts to estimate cohorts' means of x_{it} and y_{it} amounts to approximate $E\{x_t|w\}$ and $E\{y_t|w\}$ by step functions. It will turn out in the following that the quality of such approximation is crucial for attaining the lower bound on asymptotic covariance.

3 Estimation with conditional moment restrictions

Consider a random sample of size n from variable v, a parameter θ and a function $\rho(v, \theta)$; consider also a conditioning variable w. The case of **inference with conditional moment restrictions** occurs when in the reference population the following condition holds:

$$E\ \{\rho(v, \theta_0)|\ w\ \} = 0\ , \tag{9}$$

where θ_0 is the true value of the parameter.

Estimation with conditional moment restrictions has been studied by Chamberlain (1987), in particular with reference to asymptotic efficiency; instrumental variables estimators for this inference problem are obtained in Newey (1990, 1993).

If condition (9) holds, then a restriction on *un*conditional moments holds as well, that is:

$$E\ \{\rho(v, \theta_0)\ g(w)\} = 0 \tag{10}$$

for any function $g(w)$ of conditioning variable. In fact, the law of iterated expectations yields:

$$E\ \{\rho(v, \theta_0)\ g(w)\} = E\{g(w)\ E[\rho(v, \theta_0)|\ w\ ]\} = 0\ .$$

Condition (10) sets up an infinite set of unconditional moment restrictions and can be used to estimate θ. Following Newey (1990, 1993), define a vector of functions of w, say $G(w)$, and consider:

$$\hat{\theta} = \text{argmin}_\theta \; \{ \; \hat{f}_n(\theta)^\top A_n \; \hat{f}_n(\theta) \; \} \; , \tag{11}$$

where:

$$\hat{f}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} G(w_i) \rho(v_i, \theta) \; ,$$

and A_n converges with probability one to A, a positive-definite matrix.

Estimator (11) is based on the **instrumental variables** $G(w)$; it is a very general estimator, since it includes several known estimators as special cases.

Asymptotic variance of $\hat{\theta}$ is minimized by the instruments:

$$\tilde{G}(w) = C \; D^\top(w) \; \Sigma^{-1}(w) \; , \tag{12}$$

where C is any non singular matrix and:

$$D(w) \;\; = \;\; E \left\{ \frac{\partial \rho(v, \theta)}{\partial \theta^\top} \; \middle| \; w \right\} \; , \tag{13}$$

$$\Sigma(w) \;\; = \;\; E \left\{ \rho(v, \theta) \; \rho(v, \theta)^\top \mid w \right\} \; ; \tag{14}$$

the variance results in:

$$\Lambda = \left\{ \; E \; \left[D^\top(w) \Sigma^{-1}(w) D(w) \right] \; \right\}^{-1} \; , \tag{15}$$

where matrices $D(w)$ and $\Sigma(w)$ are evaluated at $\theta = \theta_0$.

The estimator $\hat{\theta}$ based on $\tilde{G}(w)$ is asymptotically optimal among all instrumental variables estimators.

Chamberlain (1987) showed that, under suitable regularity conditions, the bound Λ in (15) is a lower bound on asymptotic variance in a wider sense, namely not only among instrumental variables estimators. In fact, under the semiparametric model only specifying the conditional moment restriction (9),

Λ is an asymptotic minimax bound. Hence, Λ is a lower bound on asymptotic variance for any regular consistent asymptotically normal estimator.

In the next Section we show how results by Chamberlain (1987) apply to the estimation of θ in (1) under the restriction (2), based on RCS. For the sake of easing reference to that paper, we will stick to Chamberlain's notation as close as possible.

4 A lower bound on asymptotic variance with across periods conditional moment restrictions

Let the playground be defined as follows.

Two random samples are available, independent from each other, each of size n, drawn from the random vector $v = (u_1, u_2, w)$. In the first sample only $v_1 = (u_1, w)$ is observed, while in the second it is observed only $v_2 = (u_2, w)$.

Denote the distributions involved as follows:

$$\begin{array}{ll} v \sim F(v;\theta) & v_1 \sim F_1(v_1;\theta) \\ w \sim F^*(w;\theta) & v_2 \sim F_2(v_2;\theta) \end{array}$$

Denote the conditional distributions of v given w by $F(v|w;\theta)$; hence $F(dv;\theta) = F(dv|w;\theta)\ F^*(dw;\theta)$. Similarly, $F_i(dv_i;\theta) = F_i(dv_i|w;\theta)\ F^*(dw;\theta)$, $i = 1, 2$.

Let the support of v be denoted by $\mathcal{V} = \mathcal{U} \times \mathcal{U} \times \mathcal{W}$, where $\mathcal{U}$ and $\mathcal{W}$ are the supports of u_i, $i = 1, 2$ and w, respectively.

According to model (1)-(2), we are given a function $\eta : \mathcal{V} \times \Theta \longmapsto \mathrm{I\!R}^{\mathrm{m}}$ such that the equation:

$$E_F\ \{\ \eta(v_1,\theta) - \eta(v_2,\theta)\ |\ w\ \} = 0 \qquad \forall\ \ w \in \mathcal{W} \tag{16}$$

has a unique zero at $\theta = \theta_0$, the true value of the parameter.

Our interest is in estimating θ_0. The estimator for θ will be obtained starting from condition (16).

By setting:

$$\rho(v,\theta) = \eta(v_1,\theta) - \eta(v_2,\theta) \tag{17}$$

results by Chamberlain (1987) we mentioned in the previous section can be exploited provided $\rho(\ .\ )$ meets the following regularity conditions (see Chamberlain's "Condition (C_2)"):

(i) *θ is in an open set $\Theta \subset \mathrm{I\!R}^p$ such that $\eta(v_i, \theta)$ and $\partial\eta(v_i,\theta)/\partial\theta^\top$ are continuous for $(v,\theta) \in \mathcal{V} \times \Theta$.*

(ii) *$E_F\{\eta(v_1,\theta_0) - \eta(v_2,\theta_0)|\ w\ \} = 0\ \ \forall\ \ w \in \mathcal{W}$.*

(iii) *$\Sigma(w) = E_F\{\ [\eta(v_1,\theta) - \eta(v_2,\theta)]\ [\eta(v_1,\theta) - \eta(v_2,\theta)]^\top\ |\ w\ \}$ is positive-definite for all $w \in \mathcal{W}$.*

(iv) *Let $D(w) = E_F\{\ \partial[\eta(v_1,\theta) - \eta(v_2,\theta)]/\partial\theta^\top\ |\ w\ \}$ for $w \in \mathcal{W}$. Then the matrix $E_{F^*}\{D^\top(w)\ \Sigma^{-1}(w)\ D(w)\}$ is positive-definite.*

Notice that: condition (ii) is met as a result of the postulated model; condition (iv) rules out the case in which $E_F\{\ \partial\eta(v_t,\theta)/\partial\theta^\top|w\}$ is constant over time.

Chamberlain's reasoning goes as follows.

1) A multinomial distribution G with finite support can be found "close" to F, with the following conditional moments the same as F:

$$E_G\ [\ \rho(v,\theta_0)|\ w\ ] = 0 \qquad \forall\ \ w \in \mathcal{W}$$

$$E_G\ [\ \rho(v,\theta)\ \rho^\top(v,\theta)|\ w\ ] = \Sigma(w)$$

$$E_G\ [\partial\rho(v,\theta)/\partial\theta^\top|\ w\ ] = D(w)$$

$$E_{G^*}[D^\top(w)\ \Sigma^{-1}(w)\ D(w)] = E_{F^*}[D^\top(w)\ \Sigma^{-1}(w)\ D(w)]$$

where G^* denotes the marginal distribution of G corresponding to the component w. (see Chamberlain's "Lemma 4".)

2) Let $\{\tau_1,\ldots,\tau_\ell\} \subset \mathcal{W}$ be the support of G^*. Let $h_j : \mathcal{W} \longmapsto \mathrm{I\!R}$ be any continuous function such that $h_j(\tau_i) = 1$ if $i = j$, $h_j(\tau_i) = 0$ if $i \neq j$, $j = 1,\ldots\ \ell$.

Define:

$$B(w) = \begin{bmatrix} h_1(w)\ I_m \\ \vdots \\ h_\ell(w)\ I_m \end{bmatrix} ,$$

and: $\psi(v, \theta) = B(w)\ \rho(v, \theta)$.

Since:

$$E_G\ \{\psi(v, \theta)\} = \begin{bmatrix} G^*\{\tau_1\}\ E_G\ [\ \rho(v, \theta)\ |\ \tau_1\ ] \\ \vdots \\ G^*\{\tau_\ell\}\ E_G\ [\ \rho(v, \theta)\ |\ \tau_\ell\ ] \end{bmatrix} , \tag{18}$$

then $E_G\ [\ \rho(v, \theta)\ |\ w\ ] = 0 \quad \forall \ \ w \in \mathcal{W}$ if and only if $E_G\ \{\psi(v, \theta)\} = 0$.

That is, the conditional moment restriction on $\rho(\ .\ )$ can be equivalently written as an unconditional moment restriction on $\psi(\ .\ )$, a major implication being that in the multinomial case we can recover the lower bound on asymptotic variance for our problem from standard theory. Such lower bound turns out to be:

$$\begin{aligned} \Lambda \ &= \ \left\{ \sum_{j=1}^{\ell} G^*\{\tau_j\}\ D^\top(\tau_j)\ \Sigma^{-1}(\tau_j) D(\tau_j) \right\}^{-1} \\ &= \ \Big[\ E_{G^*} D^\top(w)\ \Sigma^{-1}(w) D(w)\ \Big]^{-1} \ . \end{aligned} \tag{19}$$

Since any F satisfying the conditional moment restriction in (ii) can be approximated arbitrarily well by a multinomial distribution satisfying the same restriction, the previous result extends to the general case (see Chamberlain's "Theorem 3").

Notice that due to (17) the conditional moment restriction (ii) is additively separable:

$$E_F\{\rho(v, \theta)|w\} = E_{F_1}[\eta(v_1, \theta)|w] - E_{F_2}[\eta(v_2, \theta)|w]\ .$$

As a consequence, it depends on F only through its marginal distributions F_1, F_2.

By the same token, its unconditional counterpart written with reference to the approximating multinomial G:

$$\begin{aligned} E_G\ \{\psi(v,\theta)\} &= \begin{bmatrix} G^*\{\tau_1\}\ E_{G_1}[\eta(v_1,\theta)|\tau_1] \\ \vdots \\ G^*\{\tau_\ell\}\ E_{G_1}[\eta(v_1,\theta)|\tau_\ell] \end{bmatrix} - \begin{bmatrix} G^*\{\tau_1\}\ E_{G_2}[\eta(v_2,\theta)|\tau_1] \\ \vdots \\ G^*\{\tau_\ell\}\ E_{G_2}[\eta(v_2,\theta)|\tau_\ell] \end{bmatrix} \\ &= E_{G_1}\ \{\psi_1(v_1,\theta)\} - E_{G_2}\ \{\psi_2(v_2,\theta)\} \end{aligned} \tag{20}$$

depends on G only trough G_1, G_2, with G_i the distribution of (v_i, w), $i = 1, 2$.

This fact implies that two alternative sampling schemes are conceivable to gather information on θ: one might draw a random sample from F; alternatively, one might draw a random sample from F_1 and an independent sample from F_2. With reference to the problem we are considering here, the first alternative amounts to collect longitudinal information, while the second one amounts to collect RCS information.

As we have seen, the central result by Chamberlain (1987) is that a restriction on conditional moments can be rewritten as a finite set of restrictions on unconditional moments *with the same informational content on* θ.

The particular sampling scheme chosen *does not make any difference* to the derivation of such result, since it only depends on features of F and $\rho(\ .\ )$.

Instead, the sampling scheme *does make the difference* with respect to two other issues. First, as we pointed out in Section 2, with longitudinal data one can choose w to be the individual identity, which is not available with RCS. The second point is that, even conditioning on a common w, the lower bound depends on the sampling scheme. To see this, it suffices to note that if a single sample is randomly drawn from G then $\psi_1(v_1,\theta)$ and $\psi_2(v_2,\theta)$ in (20) need not be independent, while if two independent samples are randomly drawn from G_1 and G_2, respectively, then $\psi_1(v_1,\theta) \perp \psi_2(v_2,\theta)$.

This fact bears consequences on Λ in (19). $D(w)$ is the same under both

sampling schemes, since:

$$E_G\{\partial\rho(v,\theta)/\partial\theta^\top|w\} = E_{G_1}[\partial\eta(v_1,\theta)/\partial\theta^\top|w] - E_{G_2}[\partial\eta(v_2,\theta)/\partial\theta^\top|w]$$

whether or not $\eta(v_1,\theta)$ is independent of $\eta(v_2,\theta)$.

On the other hand,

$$\begin{aligned}\Sigma(w) &= \mathrm{Var}\{\eta(v_1,\theta) - \eta(v_2,\theta)|\ w\ \} \\ &= \mathrm{Var}\{\eta(v_1,\theta)|w\} + \mathrm{Var}\{\eta(v_2,\theta)|w\} - 2\,\mathrm{Cov}\{\eta(v_1,\theta),\eta(v_2,\theta)|w\}\end{aligned}$$

is not the same under the two sampling schemes, in that if G_1 and G_2 are independently investigated then the covariance term vanishes by construction, while if a single sample is drawn from G, that is if available information is longitudinal, the covariance term vanishes if and only if ξ_t is serially uncorrelated. As a result, under the RCS scheme:

$$\Sigma(w) = \mathrm{Var}\{\varepsilon_1|w\} + \mathrm{Var}\{\varepsilon_2|w\}\ .$$

4.1 A related issue

Estimation of θ in (1) via RCS information is an example of inference based on splicing complementary data sources. A formally identical problem is the following (see Angrist, Krueger, 1992, and Arellano, Meghir, 1992). Let the model be:

$$y - \beta x = \varepsilon\ ,$$

$$E\{\varepsilon|w\} = 0\ .$$

The restriction on the conditional moment of ε can be written in terms of observables and the unknown parameter β:

$$E\{y - \beta x|w\} = E\{y|w\} - \beta E\{x|w\} = 0\ .$$

Due to the additive separability of $(y - \beta x)$, inference on β can be based either on sampling from (y, x, w) or on independently sampling from (y, w) and (x, w), respectively.

The lower bound on the asymptotic variance does depend on the chosen sampling scheme.

5 Towards an efficient estimator

How the efficient estimator could be obtained for our inference problem it is an open question, beyond the scope of this paper. We just give some informal hints and illustrate them by means of a special case.

Since $E_F\ \{\eta(v_1,\theta)-\eta(v_2,\theta)|w\}=0$ at $\theta=\theta_0$, $\forall\ w\in\mathcal{W}$, in the case of two independent samples one could build the following function:

$$\hat{f}_n(\theta)=\sum_{i=1}^{n}\sum_{j=1}^{n} M(\ w(w_i,w_j)\ )\ [\ \eta(v_1,\theta)_i-\eta(v_2,\theta)_j\ ]\ K(w_i,w_j)\ , \tag{21}$$

where index i runs over the units of first sample and index j over the units of second sample; the function $K(w_i,w_j)$ weights pairs (w_i,w_j) according to the degree of similarity between i-th unit of first sample and j-th unit of second sample.

A generalized method of moments estimator would be obtained as:

$$\hat{\theta}=\text{argmin}_\theta\ \left\{\hat{f}_n(\theta)^\top A_n\ \hat{f}_n(\theta)\right\}\ ,$$

where $A_n \xrightarrow{\text{p}} A$, a positive definite matrix.

Consider the case in which w has finite support $\{\tau_1,\ldots,\tau_\ell\}$, and let n_{kj} be the number of units in the j-th sample ($j=1,2$) for which $w=\tau_k$, $k=1,\ldots,\ell$. Assume that $n_{k1}=n_{k2}=m,\ k=1,\ldots,\ell$. Let:

$$K(w_i,w_j)=\begin{cases}1 & \text{if}\ \ w_i=w_j\\ 0 & \text{otherwise}\end{cases}$$

Then (21) becomes:

$$\hat{f}_n(\theta)\propto\sum_{k=1}^{\ell}\ M(\tau_k)\,[\ \bar{\eta}(1)_k-\bar{\eta}(2)_k\ ]\,, \tag{22}$$

where $\bar{\eta}(t)_k = \frac{1}{m} \sum_{\{i\,:\,w_i=\tau_k\}} \ \eta(v_{it}, \theta)$.
Asymptotic variance of $\hat{\theta}$ is minimized by the following choice of instruments: $M^*(w) = D^\top(w)\ \Sigma^{-1}(w)$.

To specialize further, let:

$$\eta(v_t, \theta) = y_t - x_t^\top \theta \ .$$

Then, in this case, we have $\rho(v, \theta) = (y_1 - y_2) - (x_1 - x_2)^\top \theta$; moreover $\partial \eta(v_t, \theta)/\partial \theta^\top = -x_t^\top$, and the quantities $D(w)$ and $\Sigma(w)$ are given by, respectively, $D(w) = E_F \left\{ \partial \rho(v, \theta)/\partial \theta^\top \middle|\ w \right\} = E\{x_2^\top |\ w\ \} - E\{x_1^\top |\ w\ \}$ and $\Sigma(w) = \mathrm{Var}\{\eta(1, \theta)|\ w\ \} + \mathrm{Var}\{\eta(2, \theta)|\ w\ \} = \sigma^2_{\varepsilon_1}(w) + \sigma^2_{\varepsilon_2}(w)$.

Hence, the lower bound for the asymptotic variance of an estimator for the slope parameter of a linear model with individual-specific effects is given by:

$$\Lambda = \left\{ E \left(\frac{[E(x_2|w) - E(x_1|w)]\ [E(x_2|w) - E(x_1|w)]^\top}{\sigma^2_{\varepsilon_1}(w) + \sigma^2_{\varepsilon_2}(w)} \right) \right\}^{-1} . \qquad (23)$$

Note that, as $E(\Delta x|w)$ gets closer to zero, the variance grows larger and larger.

The efficient instruments are:

$$M^*(w) = \frac{[E(x_2|w) - E(x_1|w)]}{\sigma^2_{\varepsilon_1}(w) + \sigma^2_{\varepsilon_2}(w)} \ .$$

Replacing population values with their sample analogues and substituting in (22), the estimating equation becomes:

$$\sum_{k=1}^{\ell} \frac{(\bar{x}_{2k} - \bar{x}_{1k})[(\bar{y}_{2k} - \bar{y}_{1k}) - (\bar{x}_{2k} - \bar{x}_{1k})^\top \theta]}{\hat{\sigma}^2_{\varepsilon_1}(\tau_k) + \hat{\sigma}^2_{\varepsilon_2}(\tau_k)} = 0$$

(where $\bar{x}_{1k} = \frac{1}{m} \sum_{\{i\,:\,w_i=\tau_k\}} \ x_{i1}$, and the other quantities are defined similarly), which yields the usual pseudo-panel within-groups estimator.

If w is not multinomial, then in the current practice a partition of $\mathcal{W}$ is chosen, $\mathcal{W} = \bigcup_{i=1}^{p} \mathcal{W}_i, \quad \mathcal{W}_i \bigcap \mathcal{W}_j = \emptyset$ if $i \neq j$, and the cohort means

$E\{x_t|w \in \mathcal{W}_i\}$, $t = 1, 2$, $i = 1, \ldots, p$, are estimated. Notice that this practice amounts to using the instruments:

$$M^{+}(w) = \frac{[E(x_2|w \in \mathcal{W}_i) - E(x_1|w \in \mathcal{W}_i)]}{\sigma^2_{\varepsilon_1}(w \in \mathcal{W}_i) + \sigma^2_{\varepsilon_2}(w \in \mathcal{W}_i)} ,$$

which are not optimal unless each $\mathcal{W}_i$ becomes somehow thinner and thinner as the sample size grows.

6 Efficiency comparison between longitudinal data and repeated cross-sections

Suppose one is choosing whether to collect a longitudinal data set or a sequence of cross-sections. Since sampling costs are rather different, it is often the case that the longitudinal 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 to answer the following question: to estimate the parameter of interest, will it be more convenient to use longitudinal information 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 Section 4.

Consider the linear model case with individual effects recalled in Section 5, and assume that homoskedasticity holds:

$$\mathrm{Var}\{\varepsilon_j|W = w\} = \sigma^2_{\varepsilon,W} \quad \forall \;\; w \in \mathcal{W} \;, \;\; j = 1, 2 \;. \tag{24}$$

If ξ is orthogonal to the other variables in the model, then:

$$\sigma^2_{\varepsilon,W} = \sigma^2_{\alpha,W} + \sigma^2_{\xi} \;.$$

Therefore, the lower bound (23) becomes:

$$\Lambda = 2 \; (\sigma^2_{\alpha,W} + \sigma^2_{\xi}) \; \Big\{ \; E \; \Big[\; E(\Delta x|w) \; E(\Delta x|w)^{\top} \; \Big] \; \Big\}^{-1} \;, \tag{25}$$

where $\Delta x = x_2 - x_1$.

If longitudinal information is available, under the supplementary assumption that the ξ_{it} in (2) are serially uncorrelated, the lower bound on the asymptotic variance can be derived from (25) with w identifying individuals, namely $w_i = i$. Then we have $E\ (\Delta x|w) = E\ (x_2 - x_1|i) = x_{i2} - x_{i1}$; therefore $E\ \left[\ E(\Delta x|w)\ E(\Delta x|w)^\top\ \right] = E\ [\ (x_{i2} - x_{i1})\ (x_{i2} - x_{i1})^\top\ ] = E\ (\ \Delta x\, \Delta x^\top\ )$. Further:

$$\mathrm{Var}\{\varepsilon_{ij}|i\} = \mathrm{Var}(\alpha_i|i) + \mathrm{Var}(\xi_{ij}|i) = \sigma_\xi^2\ ,$$

since the variance of the individual effect conditional on the individual is zero.

If RCS information is available, then $0 < \sigma^2_{\alpha,W} \leq \sigma^2_\alpha$, since conditioning on w does not increase the variance of α, but neither annihilates it (unless w is the individual identity).

These considerations give:

- in the case of RCS:

$$\Lambda_{RCS} = 2\ (\ \sigma^2_{\alpha,W} + \sigma^2_\xi\ )\ [\ \mathrm{Var}\{E(\Delta x|w)\} + E(\Delta x)\ E(\Delta x)^\top\ ]^{-1}\ ; \tag{26}$$

- in the case of longitudinal data:

$$\begin{aligned} \Lambda_{LD} &= 2\ \sigma^2_\xi\ [\ \mathrm{Var}(\Delta x) + E(\Delta x)\ E(\Delta x)^\top\ ]^{-1} \\ &= 2\ \sigma^2_\xi\ [E\{\mathrm{Var}(\Delta x|w)\} + \mathrm{Var}\{E(\Delta x|w)\} + E(\Delta x)E(\Delta x)^\top]^{-1} \end{aligned} \tag{27}$$

  by the usual decomposition of variance, where the time-invariant w we are conditioning on is the same one appearing in (26).

Hence, it is clear that the lower bound in (27) is smaller than the one in (26), since the numerator is smaller and the denominator is bigger.

If $\hat{\theta}_n$ is an estimator for θ_0 based on a sample of size n for which $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\text{p}} \mathcal{N}(0, s)$ holds, then we have that $s \geq \Lambda$; therefore, longitudinal data allow, asymptotically, to reach a higher efficiency, since $\Lambda_{RCS} > \Lambda_{LD}$.

If sample size is large enough to allow for an asymptotic approximation of the finite sample distribution, then the lower bound for RCS and longitudinal estimators are:

$$\frac{\Lambda_{RCS}}{n_{RCS}} \quad , \quad \frac{\Lambda_{LD}}{n_{LD}} ,$$

respectively, where n_{RCS} denotes the size of each sectional sample and n_{LD} the size of the longitudinal sample. Since the typical case is that $n_{RCS} \gg n_{LD}$, it is not obvious which of the two variances is the smallest, and in the applications we have to evaluate case by case which sampling scheme is more convenient to obtain the desired precision.

On the other hand, we had to expect such a result, which is already evident at intuitive level from simple remarks - for example - about the costs of the surveys, which are among the reasons for the strong differences between the sampling sizes. A forewarning in this sense has already been given by Heckman and Robb (1985), who - among the first to consider the possibility of estimating by repeated cross-sections models born for longitudinal data - noticed that it is not possible to establish in advance which sampling scheme yields more information on the parameter of interest.

7 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. Links to inference problems based on splicing complementary data sources are made explicit.

Further, we show that in the linear case the usual pseudo-panel within-groups estimator attains the lower bound if the time-invariant characteristic on which cohorts are built is multinomial.

Lastly, we compare RCS estimators to longitudinal ones in terms of lower bound on variance. As intuition suggests, it turns out that the latter outperform the former asymptotically, while this needs not happen in finite samples due to the typical much smaller sample size of longitudinal surveys.

ACKNOWLEDGMENTS

Financial support from the MURST (grant no. 020902124) is gratefully acknowledged by the first two authors.

REFERENCES

Angrist, J. and Krueger, A. B. (1992). The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, **87**, 328-336.

Arellano, M. and Meghir, C. (1992). Female labour supply and the on-the-job search: an empirical model estimated using complementary data sets. *Review of Economic Studies*, **59**, 537-557.

Baltagi, B. H. (1995). *Econometric Analysis of Panel Data.* Wiley, New York.

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

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

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.

Hsiao, C. (1986). *Analysis of Panel Data.* Cambridge University Press, New York.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica*, **58**, 809-837.

Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In *Handbook of Statistics*, vol. **11**, Econometrics, (eds. G. S. Maddala, C. R. Rao and H. D. Vinod), Elsevier Science Publishers, North Holland, 419-454.

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

Working Papers già pubblicati

1. E. Battistin, A. Gavosto e E. Rettore, *Why do subsidized firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy*, Agosto 1998.
2. N. Rosati, E. Rettore e G. Masarotto, *A lower bound on asymptotic variance of repeated cross-sections estimators in fixed-effects models*, Agosto 1998.

Le richieste di working papers vanno indirizzate alla Sig.ra Patrizia Piacentini, Dipartimento di Scienze Statistiche, via C. Battisti 241-243, 35121 Padova, oppure all'indirizzo e-mail patrizia.piacentini@unipd.it.