BC BOOTSTRAP CONFIDENCE INTERVALS FOR RANDOM EFFECTS PANEL DATA MODELS

Andrés Carvajal
Brown University
Department of Economics

Abstract:

We study the application of bootstrap procedures to the problem of constructing confidence intervals for the coefficients of random effects panel data models, based on GLS point estimation. The central problem is the one of adequately resampling from the estimated residuals of the model, avoiding violations of the structural features of the random shocks.

January 14, 2000

1 Introduction

One of the most important tools in microeconometrics, as well as in other fields of econometrics, is the use of models that combine time series and cross-sectional data, or panel data models. On the other hand, bootstrap procedures to evaluate the accuracy of summary statistics, or for inference problems in general, have gained popularity. Although they are computationally more expensive than standard methods, they can be applied to almost any statistical problem, do not pose problems when the statistician transforms her or his parameters, are usually more accurate than the standard intervals[1] and do not require having to assume particular probability distributions[2].

In this paper, we study the application of bootstrap procedures to the problem of constructing confidence intervals for the coefficients of random effects panel data models, based on GLS point estimation. The central problem is the one of adequately resampling from the estimated residuals of the model, avoiding (important) violations of the structural features of the random processes.

The paper is organized as follows: in the following section we introduce the random effects panel data model; then, we study the generalities of the bootstrap procedures, paying particular attention to the resampling problem for a random effects panel data model. In particular we concentrate in the problem of resampling the estimated residuals in a coherent way that avoids the imposition of false restrictions on the structure of the random shocks. We propose four alternative resampling plans; after that, we introduce an experiment that tests the proposed plans; once we analyze the results of our experiment, we state some conclusions that are to be taken as a preliminary approach to the problem.

2 The Random Effects Panel Data Model

2.1 Model and Assumptions

Consider the following canonical model:

$$y_{it} = \beta' x_{it} + v_i + \varepsilon_t + w_{it} \tag{1}$$

where $i \in \{1, 2, ..., N\}$, $t \in \{1, 2, ..., T\}$, the dimensions of both x_{it} and β are $K \times 1$ and all the other terms in the equation are scalars. v_i, ε_t and w_{it} are random disturbances[3], and x_{it} is the vector of explanatory variables. We assume that the model includes a constant, so that

$$x_{it} = \begin{pmatrix} 1 \\ x_{it}^S \end{pmatrix}$$

[1] Efron (1987), p. 171.

[2] Efron and Tibshirani (1993), p. 160.

[3] v_i is an *individual-specific*, or *idiosincratic*, shock, while ε_t is *time-specific*. Since w_{it} is not particualr of an individual or time period, we will refer to it as a *unspecific* shock.

where x_{it}^S is a $(K-1)\times 1$ vector.

We also assume that the following conditions are satisfied:

Condition 1 *:* $E(v_i) = E(\varepsilon_t) = E(w_{it}) = 0$

Condition 2 *:* $\forall\ i \in \{1,2,...,N\},\ \forall t \in \{1,2,...,T\},\ E(v_i\varepsilon_t) = E(v_i w_{it}) = E(\varepsilon_t w_{it}) = 0$

Condition 3 *:*

$$\begin{aligned} E(v_i v_j) &= \begin{cases} \sigma_v^2 & \textit{if}\ \ i=j \\ 0 & \textit{otherwise} \end{cases} \\ E(\varepsilon_t \varepsilon_s) &= \begin{cases} \sigma_\varepsilon^2 & \textit{if}\ \ t=s \\ 0 & \textit{otherwise} \end{cases} \\ E(w_{it} w_{js}) &= \begin{cases} \sigma_w^2 & \textit{if}\ \ i=j \ \text{ and }\ t=s \\ 0 & \qquad \textit{otherwise} \end{cases} \end{aligned}$$

Condition 4 *:* $\forall\ i \in \{1,2,...,N\}$ *and* $\forall t \in \{1,2,...,T\}$, $E(v_i x_{it}) = E(\varepsilon_t x_{it}) = E(w_{it} x_{it}) = 0$

Under these conditions, it is straightforward that

$$Var(y_{it} \mid x_{it}) = \sigma_v^2 + \sigma_\varepsilon^2 + \sigma_w^2 \tag{2}$$

Now, denote

$$l_A = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{A\times 1} \quad \text{for any integer } A, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix}_{T\times 1},$$

$$w_i = \begin{pmatrix} w_{i1} \\ \vdots \\ w_{iT} \end{pmatrix}_{T\times 1}, \quad y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}_{T\times 1} \quad \text{and } X_i = \begin{pmatrix} x_{i1}' \\ \vdots \\ x_{iT}' \end{pmatrix}_{T\times K}$$

and let $u_{it} = v_i + \varepsilon_t + w_{it}$, $u_i = v_i l_T + \epsilon + w_i$ and

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}_{NT\times 1}, \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_T \end{pmatrix}_{NT\times K} = \begin{pmatrix} l_{NT} & X_S \end{pmatrix},$$

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}_{NT\times 1} \quad \text{and} \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_T \end{pmatrix}_{NT\times 1}$$

where X^S is a $NT \times (K-1)$ matrix. Then, we can reexpress (3) as

$$y = X\beta + v \otimes l_T + (l_N \otimes I_T)\,\varepsilon + w \tag{3}$$

2.2 Generalized Least Squares (GLS) Estimation

Denoting

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}_{NT \times 1} = v \otimes l_T + (l_N \otimes I_T)\,\varepsilon + w$$

one gets

$$\Omega = E\,(uu' \mid X) = \sigma_v^2\,(I_N \otimes l_T l_T') + \sigma_\varepsilon^2\,(l_N l_N' \otimes I_T) + \sigma_w^2 I_{NT} \tag{4}$$

and the GLS estimator

$$\widehat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y \tag{5}$$

is BLUE. Denote

$$\beta = \begin{pmatrix} \beta^C \\ \beta^S \end{pmatrix} \text{ and } \widehat{\beta}_{GLS} = \begin{pmatrix} \widehat{\beta}_{GLS}^C \\ \widehat{\beta}_{GLS}^S \end{pmatrix}$$

where β^C and $\widehat{\beta}_{GLS}^C$ are scalars and β^S and $\widehat{\beta}_{GLS}^S$ are $(K-1)\times 1$ vectors. Since Ω is usually hard to invert, some algebra –Judge et al, (1985)– shows that one can also obtain the slope coefficients, $\widehat{\beta}_{GLS}^S$ as

$$\begin{aligned} \widehat{\beta}_{GLS}^S \;=\; & \left(\frac{X_S'Q_1X_S}{\sigma_1^2} + \frac{X_S'Q_2X_S}{\sigma_2^2} + \frac{X_S'QX_S}{\sigma_w^2}\right)^{-1} \\ & \left(\frac{X_S'Q_1X_S}{\sigma_1^2}\widehat{\beta}_1^S + \frac{X_S'Q_2X_S}{\sigma_2^2}\widehat{\beta}_2^S + \frac{X_S'QX_S}{\sigma_w^2}\widehat{\beta}^S\right) \end{aligned} \tag{6}$$

where

$$\begin{aligned} Q_1 &= I_N \otimes \frac{l_T l_T'}{T} - \frac{l_{NT} l_{NT}'}{NT} \\ Q_2 &= \frac{l_N l_N'}{N} \otimes I_T - \frac{l_{NT} l_{NT}'}{NT} \\ Q &= I_{NT} - I_N \otimes \frac{l_T l_T'}{T} - \frac{l_N l_N'}{N} \otimes I_T + \frac{l_{NT} l_{NT}'}{NT} \\ \sigma_1^2 &= \sigma_w^2 + T\sigma_v^2 \\ \sigma_2^2 &= \sigma_w^2 + N\sigma_\varepsilon^2 \\ \widehat{\beta}_1^S &= \left(X_S'Q_1X_S\right)^{-1} X_S'Q_1 y \\ \widehat{\beta}_2^S &= \left(X_S'Q_2X_S\right)^{-1} X_S'Q_2 y \\ \widehat{\beta}^S &= \left(X_S'QX_S\right)^{-1} X_S'Q y \end{aligned}$$

While the constant can be obtained as

$$\widehat{\beta}_{GLS}^C = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \left(\widehat{\beta}_{GLS}^S\right)' x_{it}^S\right) \tag{7}$$

3 Bootstrap Confidence Intervals

3.1 Bootstrap Procedures

Any bootstrap procedure follows two basic concepts:

Definition 5 *Let the sequence* $\langle z_m \rangle_{m=1}^{M}$ *represent a random sample of size* M*,* $0 < M < \infty$*, of a random variable* Z*, which has distribution function* $F : R \longrightarrow [0,1]$*. The* empirical distribution function*, denoted* $\widehat{F} : R \longrightarrow [0,1]$*, is the (simple) function that assigns to each* z_m *a mass* M^{-1}*. Thus,* $\widehat{F}(z) = M^{-1}\text{Card}\{z_m \in (-\infty, z]\}$*.*

Definition 6 *Let* $\mathcal{F}$ *represent the set of all distribution functions, and consider a mapping* $\theta : \mathcal{F}_\theta \longrightarrow \Theta$ *where* $\mathcal{F}_\theta \subseteq \mathcal{F}$*. If* $F \in \mathcal{F}_\theta$*, we refer to* $\theta(F)$ *as a parameter of the distribution* F*. The* plug-in estimate *of such parameter is* $\widehat{\theta} = \theta\left(\widehat{F}\right)$*, whenever* $\widehat{F} \in \mathcal{F}_\theta$*.*

In particular, consider the case of a regression model

$$y_m = \beta' x_m + u_m$$

for $m \in \{1, 2, ..., M\}$. The bootstrap confidence intervals procedure consists of using the empirical distribution of the sequence of estimated residuals $\langle \widehat{u}_m \rangle_{m=1}^{M}$, to resample sequences $\langle \widehat{u}_m^b \rangle_{m=1}^{M}$, for $b \in \{1, 2, ..., B\}$[4]; then, the sequence of the dependent variable (y) is recreated, conditional on the independent variables (x) and the estimates of the regression coefficients ($\widehat{\beta}$), using $\langle \widehat{u}_m^b \rangle_{m=1}^{M}$(more clearly: $y_m^b = \widehat{\beta}' x_m + \widehat{u}_m^b$); then, for each b, a new $\widehat{\beta}_b$ is estimated using the observed sequence $\langle x_m \rangle_{m=1}^{M}$ and the simulated $\langle y_m^b \rangle_{m=1}^{M}$; with this, one obtains the sequence $\left\langle \widehat{\beta}_b \right\rangle_{b=1}^{B}$, and can calculate its (simple) distribution function $\widehat{G} : R \longrightarrow [0,1]$, which is the empirical analogous of the (possibly unknown) distribution function, G, of the random variable $\widehat{\beta}$; Finally, one uses the plug-in principle and, through the inverse function $\widehat{G}^{-1}$, obtains the confidence interval[5].

3.2 Resampling from the GLS residuals

The crux of the problem in the case of panel data models is that there is no obvious way to resample the estimated residuals. The whole point is that when we obtain the estimated residuals, what we are getting (retaking the notation introduced in section 1) is the sequence $\left\langle \langle \widehat{u}_{it} \rangle_{t=1}^{T} \right\rangle_{i=1}^{N}$, where $u_{it} = v_i + \varepsilon_t + w_{it}$, while, ideally, we would like to resample independently from each of the following sequences: $\langle \widehat{v}_i \rangle_{i=1}^{N}$, $\langle \widehat{\varepsilon}_t \rangle_{t=1}^{T}$ and $\left\langle \langle \widehat{w}_{it} \rangle_{t=1}^{T} \right\rangle_{i=1}^{N}$.

[4] B is a "large" integer number about whose determination we will later talk.

[5] How exactly to determine the bounds will be explained later on.

Before proceeding, we introduce the following definitions:

Definition 7 *A resampling plan is* time-coherent *if* $\forall b$

$$\widehat{u}_{it}^{b}=\widehat{v}_{i'}+\widehat{\varepsilon}_{s}+\widehat{w}_{i''s'}\Longrightarrow\widehat{u}_{jt}^{b}=\widehat{v}_{j'}+\widehat{\varepsilon}_{s}+\widehat{w}_{j''s''}\quad\forall j\in\{1,...,N\}$$

where $i,i',i'',j',j''\in\{1,...,N\}$ *and* $t,s,s',s''\in\{1,...,T\}$.

Definition 8 *A resampling plan is* individual-coherent *if* $\forall b$

$$\widehat{u}_{it}^{b}=\widehat{v}_{j}+\widehat{\varepsilon}_{t'}+\widehat{w}_{j't''}\Longrightarrow\widehat{u}_{is}^{b}=\widehat{v}_{j}+\widehat{\varepsilon}_{s'}+\widehat{w}_{j''s''}\quad\forall s\in\{1,...,T\}$$

where $i,j,j',j''\in\{1,...,N\}$ *and* $t,t',t'',s',s''\in\{1,...,T\}$.

Definition 9 *A resampling plan is* dynamically over-restrictive *if* $\exists b$ *such that*

$$\widehat{u}_{it}^{b}=\widehat{v}_{i'}+\widehat{\varepsilon}_{t'}+\widehat{w}_{i''t''}\Longrightarrow\exists s\in\{1,...,T\},s\neq t:\widehat{u}_{is}^{b}=\widehat{v}_{j}+\widehat{\varepsilon}_{s'}+\widehat{w}_{j't''}$$

where $i,i',i'',j,j'\in\{1,...,N\}$ *and* $t,t',t'',s'\in\{1,...,T\}$.

Definition 10 *A resampling plan is* cross-sectionally over-restrictive *if* $\exists b$ *such that*

$$\widehat{u}_{it}^{b}=\widehat{v}_{i'}+\widehat{\varepsilon}_{t'}+\widehat{w}_{i''t''}\Longrightarrow\exists j\in\{1,...,N\},j\neq i:\widehat{u}_{jt}^{b}=\widehat{v}_{j'}+\widehat{\varepsilon}_{s}+\widehat{w}_{i''s'}$$

where $i,i',i'',j'\in\{1,...,N\}$ *and* $t,t',t''s,s'\in\{1,...,T\}$.

Since the notation is cumbersome, these definitions deserve further comment. A plan is time-coherent if, during the simulations, if at time t an individual receives the time-specific shock corresponding to time s ($\widehat{\varepsilon}_s$), then all the other individuals should receive that same time-specific shock at that same time period. A plan is individual-coherent if, during the simulations, at some period an individual i receives the idiosyncratic shock corresponding to individual j ($\widehat{v}_j$), implies then that at all other time periods, that same individual (i) should receive that same individual-specific shock. One would like to use a plan that is both time- and individual-coherent.

On the other hand, a plan is dynamically over-restrictive if it happens that, during some simulation, the fact that at time t one individual receives an unspecific shock corresponding to some time period t'' ($\widehat{w}_{i''t''}$) suffices to imply that the same individual will (at some other point) receive another unspecific shock corresponding to that same time period (t''). Since w is unspecific, one would like to have a plan where that does not happen. In some sense, a dynamically over-restrictive imposes to the empirical distribution of w a dynamic correlation that we have ruled out form the features of the true distribution. Similarly, a plan is cross-sectionally over-restrictive if, during some simulation, the fact that at some time period individual i receives an unspecific shock corresponding to some individual i'' implies that at that same time period someone else will receive a unspecific shock also corresponding to i'' (although, maybe at some other time). Again, this amounts to empirically imposing to w a cross-sectional correlation that it does not have. One would like to avoid such imposition.

One choice that the researcher has is to try and study the possibility of decomposing between the three components[6]. We take a different approach. What we do is to present different resampling plans for the sequence $\left\langle \langle \widehat{u}_{it} \rangle_{t=1}^{T} \right\rangle_{i=1}^{N}$ and see their advantages and disadvantages in terms of the features we just defined.

In order to keep things simple, consider the sequence $\left\langle \langle \widehat{u}_{it} \rangle_{t=1}^{T} \right\rangle_{i=1}^{N}$ organized in a $T \times N$ matrix as follows

$$\begin{pmatrix} \widehat{u}_{11} & \widehat{u}_{21} & \cdots & \widehat{u}_{N1} \\ \widehat{u}_{12} & \widehat{u}_{22} & \cdots & \widehat{u}_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{u}_{1T} & \widehat{u}_{2T} & \cdots & \widehat{u}_{NT} \end{pmatrix} \tag{8}$$

where the rows are the time dimension and the columns are the cross-sectional dimension[7]. The problem of resampling is simply to construct, for each b, a matrix

$$\begin{pmatrix} \widehat{u}_{11}^{b} & \widehat{u}_{21}^{b} & \cdots & \widehat{u}_{N1}^{b} \\ \widehat{u}_{12}^{b} & \widehat{u}_{22}^{b} & \cdots & \widehat{u}_{N2}^{b} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{u}_{1T}^{b} & \widehat{u}_{2T}^{b} & \cdots & \widehat{u}_{NT}^{b} \end{pmatrix} \tag{9}$$

which will be used, in a consistent manner, to create the sequence $\left\langle \langle \widehat{u}_{it}^{b} \rangle_{t=1}^{T} \right\rangle_{i=1}^{N}$

3.2.1 Incoherent Resampling Plan (IRP)

One first approach would be to ignore the coherence problem altogether and resample from the sequence $\left\langle \langle \widehat{u}_{it} \rangle_{t=1}^{T} \right\rangle_{i=1}^{N}$ giving a probability mass equal to $1/NT$ to each and all of its elements. In terms of the matrices we introduced beforehand, this means to fill each of the positions of the matrix 9 by random selection (with replacement) of the elements of matrix 8 with probability $1/NT$. The advantage of this plan is that it is not over-restrictive, neither dynamically nor cross-sectionally. The cost it implies is, however, that our resampling will be both time- and individual-incoherent.

[6] The most appealing way probably being to use the estimators $\widehat{\beta}_1^S$ and $\widehat{\beta}_2^S$ to obtain $\langle \widehat{v}_i \rangle_{i=1}^{N}$ and $\langle \widehat{\varepsilon}_t \rangle_{t=1}^{T}$ respectively and then obtain $\left\langle \langle \widehat{w}_{it} \rangle_{t=1}^{T} \right\rangle_{i=1}^{N}$. This is interesting but is not free of problems. For example, it amounts to assuming that $\forall i \ \frac{1}{T} \sum_{t=1}^{T} w_{it} = E(w_{it}) = 0$, $\forall t \ \frac{1}{N} \sum_{i=1}^{N} w_{it} = E(w_{it}) = 0$, and that $\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t = E(\varepsilon_t) = 0$ and $\frac{1}{N} \sum_{i=1}^{N} v_i = E(v_i) = 0$ and, nonetheless, introduces theoretical problems regarding the variances.

[7] One must be cautious and notice that, contrary to what is usual, $\widehat{u}_{it}$ is placed in the (t, i) entry of the matrix in the sense that it occupies the t^{th} row and the i^{th} column.

3.2.2 (Fully) Coherent Resampling Plan (CRP)

On the other extreme, we can use a plan that fully ensures the coherence of matrix 9. The plan consists of two steps, the order of which is immaterial:

- First, one constructs a matrix whose rows are randomly selected (with replacement) from the rows of matrix 8, with probability $1/T$. The key fact at this step, is that one preserves the row of matrix 8 when constructing the new matrix. This fact ensures the time-coherence of the plan.
- Then, based on the new matrix we just created, we create matrix 9. To do it coherently, one selects the columns of 9 randomly (and with replacement) from the columns of the matrix previously constructed, with probability $1/N$. This ensures the individual-coherence of the plan.

The advantages of the plan are, as we just said, its full coherence. Its disadvantage should also be clear: resampling in this way is both dynamically and cross-sectionally over-restrictive.

3.2.3 Time-Coherent Resampling Plan (TCRP)

Another alternative combines the ideas of the (extreme) previous plans. Again, there are two steps:

- As in the CRP, one first constructs a matrix whose rows are randomly selected (with replacement) from the rows of matrix 8, with probability $1/T$. Again, this suffices to ensure the time-coherence of the plan.
- Now, in order to avoid being cross-sectionally over-restrictive, one constructs matrix 9 by randomly choosing for each element of its rows from the elements of the corresponding row of the previously created matrix, with probability $1/N$. This is independently repeated for each of the rows.

The difference between the last steps of the TCRP and the CRP is simple: while in the CRP one resamples the whole columns, in the TCRP one resamples, for each row, element by element. The independence that the second step of the TCRP has implies that it is not cross-sectionally over-restrictive while the dependence that the first step has implies that it is time-coherent. The disadvantages are clear, the TCRP is not individual-coherent and is dynamically over-restrictive.

3.2.4 Individual-Coherent Plan (ICRP)

The fourth alternative is the "transpose" of the concept behind the TCRP:

- First, one creates a matrix whose columns are randomly selected (with replacement) from the columns of matrix 8, with probability $1/N$. The fact that one preserves the whole column suffices to imply that the plan is individual-coherent.

- Then, one constructs matrix 9 by randomly choosing for each element of its columns one element of the corresponding column of the matrix created in the first step, with probability $1/T$. This is done independently for each of the columns, which ensures that the plan is not $1/N$ dynamically over-restrictive.

We already mentioned the advantages of the ICRP plan: it is individual-coherent and not dynamically over-restrictive. Its disadvantages are also clear: it is time-incoherent and cross-sectionally over-restrictive.

3.2.5 Choosing the right plan

It must be clear now that, without identifying each of the components of $\widehat{u}_{it}$ one cannot get a perfect plan: time-coherence implies dynamic over-restrictiveness, and individual-coherence implies cross-sectional over-restrictiveness. What we later do is to perform an experiment to show how the relative magnitudes of σ_v^2, σ_ε^2 and σ_w^2 imply "better" or "worse" confidence intervals in each of the resampling plans. Before that, however, we introduce the concepts that we will need for the construction of the confidence intervals.

3.3 The BC Confidence Intervals

For each b, the sequences $\left\langle \langle \widehat{u}_{it}^b \rangle_{t=1}^T \right\rangle_{i=1}^N$ and $\left\langle \langle x_{it}' \rangle_{t=1}^T \right\rangle_{i=1}^N$, as well as $\widehat{\beta}_{GLS}$, are used to create a new sequence $\left\langle \langle y_{it}^b \rangle_{t=1}^T \right\rangle_{i=1}^N$. Then, $\left\langle \langle y_{it}^b, x_{it}' \rangle_{t=1}^T \right\rangle_{i=1}^N$ is used to estimate, also by GLS, a new $\widehat{\beta}_b$. Repeating the process B times, one gets the sequence $\left\langle \widehat{\beta}_b \right\rangle_{b=1}^B$. Suppose, for simplicity[8], $K = 1$. Then, one constructs the mapping $\widehat{G} : R \longrightarrow [0,1]$, which is, as we had previously said, the (empirical) cdf of $\widehat{\beta}_b$. The determination of the $1-2\alpha$ confidence interval[9] reduces now to simply determining some critical points of such cdf.

Efron (1987) introduced a bootstrap confidence interval which proved to have reduced bias and high accuracy. It was called the $\mathrm{BC_a}$ confidence interval. Let $\Phi : R \longrightarrow [0,1]$ represent the standard normal cdf and let $\underline{\beta}^{BC_a}(\alpha)$ and $\overline{\beta}^{BC_a}(\alpha)$ be the lower and upper bounds of the $\mathrm{BC_a}$ interval, respectively. They are given by

$$\left[\underline{\beta}^{BC_a}(\alpha), \overline{\beta}^{BC_a}(\alpha)\right] = \left[\widehat{G}^{-1}\left(\Phi\left(z\left(\alpha\right)\right)\right), \widehat{G}^{-1}\left(\Phi\left(z\left(1-\alpha\right)\right)\right)\right] \tag{10}$$

where the function $z : [0,1] \longrightarrow R$ is defined as

$$z\left(\gamma\right) = z_0 + \frac{z_0 + \Phi^{-1}\left(\gamma\right)}{1 - a\left(z_0 + \Phi^{-1}\left(\gamma\right)\right)} \tag{11}$$

[8] In the case $K > 1$, one does the following for the element of $\widehat{\beta}_b$ corresponding to the one of β on whose confidence interval one is interested.

[9] i.e. one which leaves $(100\alpha)\%$ of the probability mass below its lower bound and $(100\alpha)\%$ of it above its upper bound, approximately.

given two constants:

$$z_0 = \Phi^{-1}\left(\widehat{G}\left(\widehat{\beta}\right)\right) \tag{12}$$

$$a \doteq \frac{1}{6}\text{SKEW}\left(\dot{\ell}_\beta\right)_{\beta=\widehat{\beta}} \tag{13}$$

where equation 13 is only an approximation and $\dot{\ell}_\beta$is the score function of the random variable $\widehat{\beta}$ under the parameter β[10].

If one forces $a = 0$, so that equation 11 becomes

$$z(\gamma) = 2z_0 + \Phi^{-1}(\gamma) \tag{14}$$

and still uses 12, then 10 can be used to find $\underline{\beta}^{BC}(\alpha)$ and $\overline{\beta}^{BC}(\alpha)$, defining the BC bootstrap confidence interval.

4 Our Experiment

We carried out an experiment to evaluate the performance of our resampling plans in what has to do with the bias and width of the BC (and BC$_\text{a}$, we think) confidence interval. The design of the experiment was as follows.

We let $K = 2$ and $N = T = 20$ and construct a matrix X (400×2) as defined in section 2.1[11]. Then, we simulated random shocks as follows:

$$\begin{aligned}
\forall i &\in \{1, ..., N\}, \; v_i \sim i.i.d.\ N\left(0, \sigma_v^2\right) \\
\forall t &\in \{1, ..., T\}, \; \varepsilon_t \sim i.i.d.\ N\left(0, \sigma_\varepsilon^2\right) \\
\forall (i,t) &\in \{1, ..., N\} \times \{1, ..., T\}, \; w_{it} \sim i.i.d.\ N\left(0, \sigma_w^2\right)
\end{aligned}$$

using different configurations for the variances.

With the sequences of shocks, and the matrix X, we constructed y (400×1) series using, as Judge et al (1985) propose[12],

$$\beta = \begin{pmatrix} 10 \\ 1 \end{pmatrix}$$

Then, we constructed BC confidence intervals[13] for the "slope" coefficient

[10] That is, the gradient of the log-likelihood function.

[11] Actually, the first 10 observations of the first 4 individuals were taken from Judge et al (1985), exercise 13.8.2, pp. 553-553. However since we wanted a large N=T, we extended the series. The series, of course, are available upon request.

[12] Exercise 13.8.2, pp. 553-553.

[13] Since we are using normally distributed shocks, we claim that the BC and BC$_\text{a}$ confidence intervals coincide (at least to the degree to which equation 13 is a good approximation). To see why, one just notices that, given equation 4, upon diffrentiation one finds that

$$\dot{\ell}_\beta = -X'\Omega^{-1}(y - X\beta) \sim N\left(0 \cdot l_2, X'\Omega^{-1}X\right)$$

so that, $\forall\beta$, $\text{SKEW}\left(\dot{\ell}_\beta\right) = 0$ and $a \doteq 0$ according to equation 13.

(β_{21}), using each of the resampling plans and using $B = 1000$[14], for $1 - 2\alpha$ levels of 0.9, 0.95 and 0.99. We calculated both the width of the interval and its bias, defined as the absolute value of the difference between the midpoint of the interval and the true value ($\beta_{21} = 1$).

We repeated the whole experiment 25 times[15], and calculated the averages of bias and width across the 25 experiments, for each of the confidence levels, resampling plans and variance configurations.. The results we obtained are the material of the next section.

5 Results of our Experiment[16]

5.1 Configuration C1 ($\sigma_v^2 = 8$, $\sigma_\varepsilon^2 = 8$, $\sigma_w^2 = 16$)

Since $\sigma_v^2 = \sigma_\varepsilon^2$ and $\sigma_w^2 = \sigma_v^2 + \sigma_v^2$, it seems hard to say *a priori* whether coherence or over-restrictiveness should concern us more. What our results showed was that, as expected, the CRP gave narrower confidence intervals than any other plan. The ICRP and TCRP gave intervals with approximately the same width, which was always lower than the one of the intervals produced by the IRP.

As for the bias, the results turned out to be less clear, although they also seem to favor the CRP. Differences in the average bias were low, but for "low" levels of α the CRP obtained the lowest average bias. In all the cases, the IRP exhibited the largest average bias.

Obviously, one should use a *formal criterion*, defined *ex-ante*, to decide which of the plans did perform best under this variance profile. Without such criterion, however, it seems that in this case the CRP gave the best results in the sense of narrow intervals with a low bias. The fact that the IRP produced the least satisfactory results seems less controversial [17].

5.2 Configuration C2 ($\sigma_v^2 = 8$, $\sigma_\varepsilon^2 = 16$, $\sigma_w^2 = 8$)

This configuration implies higher variance for the time-specific shock than for any other. Thus, a conjecture would be that time-coherence should be a major concern.

What we found was that the TCRP gave in average the second narrowest confidence intervals[18]. On the other hand, however, TCRP produced the least biased results. Again, without a formal criterion, any conclusion has to be taken carefully. It seems, however, that the combination of relatively narrow intervals

[14] There are methods to determine B endogenously –e.g. Andrews and Buchinsky (1999). However, for reasons of computational costs, we followed B=1000 as the rule of thumb proposed by Efron (1987, p. 173 and section 9).

[15] This number may seem low and indeed it is. However, it was the largest feasible number, given the computational constraints.

[16] A summary of these results is given in a table attached at the end of this paper. The Gauss program with which we performed the experiment is, of course, available upon request.

[17] One must recognize, nonetheless, that (by a little) the IRP is less computationally costly that the others.

[18] The ascending order, according to average width was: CRP, TCRP, ICRP, IRP.

with the lowest bias favors, under this variance configuration, the performance of the TCRP.

5.3 Configuration C3 ($\sigma_v^2 = 16$, $\sigma_\varepsilon^2 = 8$, $\sigma_w^2 = 8$)

Again, based on the magnitude of the variance of the individual-specific shock, in this case one should be particularly concerned about individual coherence. Once again, the results conform to that conjecture: on average, the ICRP provides us with the second narrowest but least biased BC confidence intervals. The CRP gives us narrower but more biased results (to the 95 and 90%, the CRP gives the most biased intervals).

Consistently with the results under configuration C2, in this case, the ICRP seems to exhibit the best performance.

5.4 Configuration C4 ($\sigma_v^2 = 1$, $\sigma_\varepsilon^2 = 1$, $\sigma_w^2 = 16$)

Under this variance configuration, the magnitude of the variance of the unspecific shock would lead one to be especially concerned about the over-restrictiveness problem. and, in effect, our results are consistent with that, in the sense that it is the IRP the plan that seems to give optimal results[19]. In general, the width of the intervals was very similar across all the plans[20]. On the other hand, the IRP showed the lowest bias to the 99 and 90% and the second lowest to the 95% confidence levels.

5.5 Configuration C5 ($\sigma_v^2 = 8$, $\sigma_\varepsilon^2 = 8$, $\sigma_w^2 = 1$)

Again, this is a case in which coherence, in both dimensions of the panel, seems to be most important concern. Accordingly, the plan that seems to give the most adequate results is the CRP. It gives intervals far narrower than the ones obtained through other plans, with levels of bias that in one case are the lowest of all, and in the others are no much larger (never being the largest) than the IRP. This case is less conclusive, but a fair conjecture seems to be that the CRP would be the optimal plan under this configuration.

6 Final Remarks

In order to build bootstrap confidence intervals for a random effects panel data model, we would like to have a resampling plan that is coherent and does not impose restrictions that do not exist (we assume) on the true random shocks. We have argued that that may be an impossible task whenever we do not want

[19] In which case, the pejorative name "incoherent", that we gave to this plan, presents itself as particularly unfair. One could better use, for example, "adequately restrictive". But we will not.

[20] This is actually a very strong result for the IRP, which is designed to have higher variance as a resampling plan. In this case, the IRP usually gave the second narrowest intervals, but the differences with the narrowest were small.

to try and estimate series for each of the shocks. Nonetheless, we believe that by adequately choosing the resampling plan, we can minimize the problems of incoherences and/or over-restrictiveness. The results of our experiment suggest that the variances of each of the shocks may provide the econometrician with an idea of which of the problems (time- or individual-coherence, or over-restrictiveness) should be the major concern. Consistently a more adequate resampling plan may be used.

Of course, the results that we did obtain constitute only a particular experiment. Further considerations should add to the decision. However, we believe that one experiment similar to the one we performed here, in the case of a particular applied work, may be helpful in the sense of orienting the econometrician towards more precise confidence intervals.

7 References

Andrews, D. and M. Buchinsky. "On the Number of Bootstrap Repetitions for BC_a Confidence Intervals." Brown University. Department of Economics. *Working Paper* No 99-17. 1999.

Efron, B. "Better Bootstrap Confidence Intervals." *Journal of the American Statistical Association*, Volume 82, Issue 397 (March, 1987), pp. 171-185. 1987.

Efron, B and R.J. Tibshirani. "An Introduction to the Bootstrap." *Monographs on Statistics and Applied Probability,* 57. Chapman & Hall. 1993.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T. Lee. "The Theory and Practice of Econometrics." Second Edition. Wiley. 1985.

Hsiao, C. "Analysis of Panel Data." *Econometric Society Monographs, 11.* Cambridge University Press. 1986.

Summary of results: Bias and width of the BC confidence intervals,

by resampling plan, variances configuration and confidence coefficient $(1-2\alpha)$

Configuration	Confidence Coefficient	IRP		CRP		TCRP		ICRP	
		Bias	Width	Bias	Width	Bias	Width	Bias	Width
C1	0.99	0.0284	0.2740	0.0258	0.1888	0.0272	0.2375	0.0240	0.2439
	0.95	0.0271	0.2086	0.0252	0.1439	0.0268	0.1785	0.0258	0.1808
	0.90	0.0268	0.1747	0.0252	0.1202	0.0265	0.1507	0.0255	0.1514
C2	0.99	0.0268	0.2805	0.0255	0.1420	0.0247	0.2047	0.0254	0.2435
	0.95	0.0261	0.2128	0.0260	0.1061	0.0254	0.1517	0.0266	0.1812
	0.90	0.0266	0.1785	0.0261	0.0887	0.0252	0.1266	0.0261	0.1523
C3	0.99	0.0235	0.2863	0.0206	0.1385	0.0200	0.2417	0.0191	0.2064
	0.95	0.0199	0.2142	0.0204	0.1033	0.0200	0.1840	0.0198	0.1535
	0.90	0.0201	0.1780	0.0207	0.0874	0.0204	0.1510	0.0199	0.1291
C4	0.99	0.0313	0.1735	0.0320	0.1738	0.0343	0.1772	0.0323	0.1694
	0.95	0.0312	0.1305	0.0325	0.1286	0.0324	0.1336	0.0311	0.1255
	0.90	0.0311	0.1091	0.0324	0.1081	0.0326	0.1119	0.0314	0.1055
C5	0.99	0.0091	0.2037	0.0080	0.0495	0.0085	0.1424	0.0095	0.1505
	0.95	0.0077	0.1518	0.0078	0.0372	0.0077	0.1065	0.0085	0.1143
	0.90	0.0076	0.1273	0.0079	0.0310	0.0078	0.0887	0.0086	0.0970