BOOTSTRAP FOR PANEL REGRESSION MODELS WITH RANDOM EFFECTS

Bertrand Hounkannounon*

Université de Montréal, CIREQ

February 05, 2009

Abstract

This paper considers bootstrap methods for panel models with random effects. It is shown that the resampling only in the cross section dimension is not valid in the presence of temporal heterogeneity. The block resampling only in the time series dimension is not valid in the presence of cross section heterogeneity. The double resampling that combines the two previous resampling methods, is valid for panel data models with cross section and/or temporal heterogeneity, with or without spatial dependence. This approach also avoids multiple asymptotics. Simulations confirm these theoretical results.

*Address for correspondance : Université de Montréal, Département de Sciences Economiques, C.P. 6128, succ. Centre-Ville, Montréal, Qc H3C 3J7, Canada. E-mail : b.hounkannounon@umontreal.ca.

1 Introduction

The true probability distribution of a test statistic is rarely known. Generally, its asymptotic law is used as approximation of the true law. If the sample size is not large enough, the asymptotic behavior of that statistic could lead to a poor approximation of the true one. Using bootstrap methods, under some regularity conditions, it is possible to obtain a more accurate approximation of the distribution of the test statistic. The original bootstrap procedure has been proposed by Efron (1979) for statistical analysis of independent and identically distributed (i.i.d.) observations. It is a powerful tool for approximating the distribution of complicated statistics based on i.i.d. data. Since Efron (1979) there has been an extensive research to extend the bootstrap to statistical analysis of non i.i.d. data. Several bootstrap procedures have been proposed for time series. The two most popular approaches are the sieve bootstrap and the block bootstrap. For an overview of bootstrap methods for dependent data, see Lahiri (2003). Application of bootstrap methods to multiple indices data is an embryonic research field. *Multiple indices data* include clustered data, multilevel data, and panel data.

The term *"panel data"* refers to the pooling of observations on a cross-section of statistical units over several periods. Because of their two dimensions, cross section and time series, panel data have the important advantage to allowing the researcher to control for unobservable heterogeneity, that is systematic difference across cross-sectional units or periods. For an overview of panel data models, see for example Baltagi (1995) or Hsiao (2003). There is an abounding literature on asymptotic theory for panel data models. Some recent developments treat large panels, when temporal and cross section dimensions are both important. However, the theoretical literature about bootstrap methods for panel data is rather recent. Kapetanios (2008) presents theoretical results when the cross-sectional dimension goes to infinity, under the assumption that cross-sectional vectors of regressors and errors terms are i.i.d.. Gonçalves (2008) shows the first order asymptotic validity of the moving blocks bootstrap for fixed effects OLS estimators of panel linear regression models with individual fixed effects. Analyzing the sample mean, Hounkannounon (2008) explores the validity of several the bootstrap resamplings methods for panel data. The main result of this paper is to provide the double resampling bootstrap that combines resampling in cross-sectional dimension and block resampling in temporal dimension. This special method is valid in the presence of cross-sectional and temporal heterogeneity, and also in the presence of spatial dependence. This paper aims to extend these results to linear regression model. The paper is organized as follows. In the second section, different panel data models are presented. Section 3 presents three bootstrap resampling methods for panel data. The fourth section presents theoretical results, analyzing validity of each resampling method. In section 5, simulation results are presented to confirm theoretical results. The sixth section concludes. Proofs of propositions are given in the appendix.

2 Panel Data Models

It is practical to represent panel data as a matrix. By convention, in this document, rows correspond to the cross-sectional units and columns represent time periods. A panel dataset with N cross-sectional units and T time periods is represented by a matrix Y of N rows and T columns. Thus Y contains NT elements. y_{it} *is the cross-sectional* $i's$ *observation at period* t.

$$\underset{(N,T)}{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & \cdots & y_{2T} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & .. & \cdots \\ y_{N1} & y_{N2} & .. & \cdots & y_{NT} \end{pmatrix} \tag{2.1}$$

Multiple Asymptotics

The double indexes imply several ways to analyze asymptotic with panel . One index can be fixed and the other goes to infinity. In the second case, how N and T go to infinity, is not always without consequence. Hsiao (2003 p. 296) distinguishes three approaches : *sequential limit, diagonal path limit* and *joint limit.* A sequential limit is obtained when an index is fixed and the other passes to infinity, to have intermediate result. The final result is obtained by passing the fixed index to infinity. In case of diagonal path limit, N and T pass to infinity along a specific path, for example $T = T(N)$ and $N \to \infty$. With joint limit, N and T pass to infinity simultaneously without a specific restrictions. In some times it can be necessary to control relative expansion rate of N and T. It is obvious that joint limit implies diagonal path limit. For equivalence conditions between sequential and joint limits, see Phillips and Moon (1999). In practice, it is not always clear how to choose among these multiple asymptotic distributions which may be different.

Panel Model Specifications

Consider a panel linear model

$$y_{it} = \theta + V_i\tau + W_t\gamma + X_{it}\zeta + \nu_{it} = Z_{it}\beta + \nu_{it} \tag{2.2}$$

$$\underset{(K,1)}{\beta} = \begin{pmatrix} \theta \\ \tau \\ \gamma \\ \zeta \end{pmatrix} \tag{2.3}$$

Three kinds of variables are considered : cross-section varying variables V_i, time varying variable W_t and double dimensions varying variables X_{it}. β is an unknown vector of parameters. The allow the identification of θ, the regressors must be centered. Inference will be about these parameters and consists in building confidence intervals for each component β_k of β.

Assumptions about ν_{it} define different panel data models. The general specification is *(2.4)* under assumptions below :

$$\nu_{it} = \mu_i + f_t + \lambda_i F_t + \varepsilon_{it} \tag{2.4}$$

Assumptions A

A1 : $(\mu_1, \mu_2, .., \mu_N) \sim i.i.d.\left(0, \sigma_\mu^2\right),\ \sigma_\mu^2 \in (0, \infty)$

A2 : $\{f_t\}$ *is a stationary and strong* α*-mixing process :*
$\alpha(j) = \sup\{|P(A \cap B) - P(A)P(B)|\}, j \in N$ where $A \in \sigma\langle\{f_t : t \leq k\}\rangle,\ B \in \sigma\langle\{f_t : t \geq k + j + 1\}\rangle,$
$k \in$

Z , with $\alpha(j) \to 0$ as $j \to \infty$. $E(f_t) = 0$ *and* $\{f_t\}$ *verifies Ibragimov's conditions :* $\exists,\ \delta \in (0, \infty),\ E|f_t|^{2+\delta} < \infty,\ \sum_{j=1}^{\infty} \alpha(j)^{\delta/(2+\delta)} < \infty,$
$V_f^\infty = \sum_{h=-\infty}^{\infty} Cov(f_t, f_{t+h}) \in (0, \infty)$.

A3 : $\{F_t\}$ *is a stationary and strong* α*-mixing process with* $E(F_t) = 0$ *verifying Ibragimov's conditions.*

A4 : $(\lambda_1, \lambda_2, .., \lambda_N) \sim i.i.d.\left(0, \sigma_\lambda^2\right),\ \sigma_\lambda^2 \in (0, \infty)$

A5 : $\{\varepsilon_{it}\}_{i=1...N,\ t=1..T} \sim i.i.d.\left(0, \sigma_\varepsilon^2\right),\ \sigma_\varepsilon^2 \in (0, \infty)$ *A6 :* $\sqrt{N}\overline{\overline{\varepsilon}} \underset{N,T\to\infty}{\overset{P}{\to}} 0$

A7 : $\sqrt{T}\overline{\overline{\varepsilon}} \underset{N,T\to\infty}{\overset{P}{\to}} 0.$

A8 : The time varying processes are independent of the cross-section varying process.

(2.4) is a two-way error component model (ECM) with spatial dependence. The term ECM comes from the structure of error terms. μ_i and f_t are respectively systematic differences across units and time periods. Classical papers on error component models include Balestra and Nerlove (1966), Fuller and Battese (1974) and Mundlak (1978). It is important to emphasize that the unobservable heterogeneity here is a random variable, not a parameter to be estimated. The alternative is to use the *fixed effects model* in that the heterogeneities are parameters[1]. The product $\lambda_i F_t$ allows the common factor F_t to have differential effects on cross-section units. This specification is used by Bai and Ng (2004), Moon and Perron (2004) and Phillips and Sul (2003). It is a way to introduce dependence among cross-sectional units. *(2.4)* uses a single

[1]Fixed effect in one dimension has a immediate consequence : parameters associated to the regressors varying only in this dimension become unidentified

factor. a multiple factor model is also possible : $\sum_{l=1}^{L} \lambda_i^L F_t^L$ with L finite or infinite. Another approach to introduce dependence a consists to use a spatial model in which the structure of the dependence can be related to geographic, economic or social distance (see Anselin (1988)).

Assumption *A1* is usual for cross-sectional heterogeneity : each unit has his specificity that does not depend on the others unobservable heterogeneity. *A3* allows weak dependence for $\{f_t\}$ and $\{F_t\}$. The fact to impose the loadings λ_i and the factor F_t to have expectations equals to zero, is not restrictive. If $E(F_t) = F \neq 0$ and $E(\lambda_i) = \lambda \neq 0$, it is possible to introduce centered variables $F_t^c = F_t - F$, $\lambda_i^c = \lambda_i - \lambda$ and $\lambda_i F_t + \varepsilon_{it} = (\lambda_i^c + \lambda)(F_t^c + F) = \lambda F + \lambda_i^c F + \lambda F_t^c + \lambda_i^c F_t^c = \eta + U_i + \Gamma_t + \lambda_i^c F_t^c + \varepsilon_{it}$: a two-way ECM with spatial dependence. In order to have expectation zero for the error term, the constant $\eta\,(= \lambda F)$ can be add to θ in the general model specification *(2.2)*. The identification of θ obliges thus the loading or the factor to have $E(F_t)\,E(\lambda_i) = 0$ in the final specification. The previous transformation show that temporal random variables can be linked, the same thing is possible with the cross-section errors. This fact supports assumption *A8* that assume independence only between the two dimensions. Considering different combinations of processes in *(2.4)*, gives the cross-sectional one-way ECM $(\mu_i + \varepsilon_{it})$, the temporal one-way ECM $(f_t + \varepsilon_{it})$, the two-way ECM $(\mu_i + f_t + \varepsilon_{it})$ and the factor model $(\lambda_i F_t + \varepsilon_{it})$ that can be used to introduce spatial dependence in the two previous. The framework of specification *(2.4)* can be compared to Driscoll and Kraay's framework that can be perceived as a one-way temporal ECM with spatial dependence vanishing when N and T grow up. Specification *(2.4)* does better adding cross-sectional heterogeneity. The possibility of the presence of random effect in the two dimensions, place our asymptotic framework in the case when N and T go to infinity. In addition to the assumptions about the errors, the followings assumptions are made about the regressors.

Assumptions B *(Regressors)*

B1 : The regressors are fixed or strictly exogenous and centered.

B2 :

$$\frac{\widetilde{Z}'\widetilde{Z}}{NT} \underset{NT\to\infty}{\to} \underset{(K,K)}{Q} > 0 \tag{2.5}$$

B3 :

$$\frac{\overline{Z}'\overline{Z}}{N} \underset{N\to\infty}{\to} \overline{Q} \tag{2.6}$$

B4 :

$$\frac{\underline{Z}'\underline{Z}}{T} \underset{T\to\infty}{\to} \underline{Q} \tag{2.7}$$

B5 :

$$\frac{\underline{Z}'\underline{Z}}{T} \underset{T\to\infty}{\to} \underline{Q} \tag{2.8}$$

B6 :

$$\lim_{NT\to\infty} \max_{j,s} Z^{/}_{(js)} \left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1} Z_{(js)} = 0 \tag{2.9}$$

B7 :

$$\lim_{N\to\infty} \max_{1\leq i\leq N} \overline{Z}^{/}_{(i)} \left(\overline{Z}^{/}\overline{Z}\right)^{-1} \overline{Z}_{(i)} = 0 \tag{2.10}$$

B8 :

$$\lim_{T\to\infty} \max_{1\leq t\leq T} \underline{Z}^{/}_{(t)} \left(\underline{Z}^{/}\underline{Z}\right)^{-1} \underline{Z}_{(t)} = 0 \tag{2.11}$$

B9 :

$$Var\left(\widehat{\beta}\right) = \left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1} \widetilde{Z}^{/}\Omega\widetilde{Z}\left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1} \to \Omega^{\infty}_{\beta} > 0 \tag{2.12}$$

The centering assumption in *A1* is necessary to identify θ. B6 to B7 are called Lindeberg's condition : asymptotically, no observation dominates the others. *B9* is important for confidence intervals : it avoids non-invertibility of the covariance matrix. The next sub-section present some classical asymptotic distributions under the assumptions above.

Classical Asymptotic Distributions

Proposition 1 *: Two-way ECM*

1 - Assume that A1 to A4, A6, B1 to B8 hold. When $N, T \to \infty$, *with* $\frac{N}{T} \to \delta \in [0, \infty)$

$$\sqrt{N}\left(\widehat{\beta} - \beta\right) \Longrightarrow N\left(0, \sigma^2_{\mu}\left(Q^{-1}\overline{Q}Q^{-1}\right) + \delta.\Sigma^{\infty}_{f}\right) \tag{2.13}$$

where $\Sigma^{\infty}_{f} = \lim \left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1} \widetilde{Z}^{/}\Omega_f\widetilde{Z}\left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1} = Q^{-1}\Omega^{\infty}_{f\widetilde{Z}}Q^{-1}$.

2 - Assume that A1 to A4, A7, B1 to B8 hold. When $N, T \to \infty$, *with* $\frac{N}{T} \to \infty$

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) \Longrightarrow N\left(0, \Sigma^{\infty}_{f}\right) \tag{2.14}$$

The relative convergence rate between the two indexes N and T, δ defines a continuum of asymptotic distributions.

Proposition 2 *: Factor Model :*

Assume that $\nu_{it} = \lambda_i F_t + \varepsilon_{it}$. *Assume also that A2, A4, A5, B1 to B8 hold. When* $N, T \to \infty$,

$$\sqrt{NT}\left(\widehat{\beta} - \beta\right) \Longrightarrow \left[N\left(0, \sigma^2_{\lambda}\left(Q^{-1}\overline{Q}Q^{-1}\right)\right)\right]\left[N\left(0, \Sigma^{\infty}_{F}\right)\right] + N\left(0, \sigma^2_{\varepsilon}Q^{-1}\right) \tag{2.15}$$

The convergence rate with $\nu_{it} = \lambda_i F_t + \varepsilon_{it}$ is greater than when there is cross-section or temporal heterogeneities. Thus in the presence of random effect in one of the two dimension, the weak spatial dependence becomes negligible .

3 Bootstrap Methods

This section presents the bootstrap method.

Methodology

From initial data (Y, Z) , create pseudo data (Y^*, Z^*) by resampling with replacement elements of (Y, Z) . This operation must be repeated B times in order to have $B+1$ pseudo-samples $\{Y_b^*, X_b^*\}_{b=1..B+1}$. Statistics are computed with these pseudo-samples in order to make inference. In this paper, inference is about β and consists in building confidence intervals for each component of the vector β. There are two main bootstrap approaches with regression models : the pairs bootstrap and the residual-based bootstrap .This paper analyzes the second one. The idea is to estimate β and to resample the residuals to create pseudo data. Several estimators are available : pooled regression estimator, within estimator, between estimator and FGLS estimator. Within estimator estimates only a subvector of β. Then inference is possible only with parameters that are not drooped by the centering. Between estimation consists on averaging the data in one dimension to make inference to have one dimension model before estimation. The drawback of this approach is to reduce drastically the number of observations. Inference became impossible for coefficient associated with variables in averaged dimension. FGLS estimation uses an estimated variance-covariance matrix. A non parametric estimator would be very useful. Driscoll and Kraay (1998) provides non parametric estimator for panel data unfortunately, its framework does not cover the specification cross-section heterogeneity. Even if a more general non parametric estimator exists, it would be asymptotically valid and would not necessarily provide good inference in small samples. In This paper, the choice is to use an unbiased and consistent estimator, even if it is not efficient : the pooled OLS estimator. The $\tilde{}$ is used to quoted vectors obtained pooling the elements of matrices. Subbar and upbar refer respectively to the average in the cross-section dimension and the temporal dimension. The different steps of the residuals based bootstrap are the followings :

Step 1 : Run the pooling regression to obtain OLS estimator $\widehat{\beta}$ and the residuals $\widehat{\nu}_{it}$

$$\widehat{\beta} = \left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1}\widetilde{Z}^{/}\widetilde{Y} \tag{3.1}$$

$$\widehat{\nu}_{it} = y_{it} - Z_{it}\widehat{\beta} \tag{3.2}$$

Step 2 : Center and Rescale the residuals in order to have better properties in small samples. By the OLS properties, the residuals have zero mean for model with constant thus the centering is not necessary.

$$u_{it} = \xi_{it}\widehat{\nu}_{it} \tag{3.3}$$

The rescaling factors available in the literature, can be easily accomodated to panel data. Using the nomenclature of MacKinnon and White (1985) :

$$HC_0 : \xi_{it} = 1 \qquad HC_1 : \xi_{it} = \sqrt{\frac{NT}{NT-K}} \tag{3.4}$$

$$HC_2 : \xi_{it} = \frac{1}{\sqrt{1-h_{it}}} \qquad HC_3 : \xi_{it} = \frac{1}{1-h_{it}} \tag{3.5}$$

where $h_{it} \equiv \widetilde{Z}_{it}\left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1}\widetilde{Z}^{/}_{it}$. The $N \times T$ matrix of rescaled residuals is U. With HC_0 there is no rescaling. HC_1 takes into account the number of degree of freedom, HC_2 and HC_3 take into account the leverage effect.

Step 3 : Use a resampling method to create pseudo-sample of residuals U^*.udo-values of the dependent variable.

$$y^*_{it} = Z_{it}\widehat{\beta} + u^*_{it} \qquad or \qquad \widetilde{Y}^* = \widetilde{Z}\widehat{\beta} + \widetilde{U}^* \tag{3.6}$$

Run pooling regression with (Y^*, Z)

$$\widehat{\beta}^* = \left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1}\widetilde{Z}^{/}\widetilde{Y}^* \tag{3.7}$$

$$\widehat{Var}^*\left(\widehat{\beta}^*\right) = \left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1}\widetilde{Z}^{/}\widehat{\Omega}^*\widetilde{Z}\left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1} \tag{3.8}$$

Step 4 : Repeat *step 3* B times in order to have $B+1$ realizations of Y^*, Z,and $\widehat{\beta}^*$: $\left\{Y^*_b, Z, \widehat{\beta}^*_b\right\}_{b=1..B+1}$

The probability measure induced by the resampling method conditionally on U is noted P^*. $E^*()$ and $Var^*()$ are respectively expectation and variance associated to P^*. The resampling methods used to compute pseudo-panel-data are exposed below.

Cross-sectional Resampling

For a $N \times T$ matrix U, *cross-sectional resampling* is the operation of constructing a $N \times T$ matrix U^* with rows obtained by resampling with replacement rows of U. Conditionally on U, the rows of U^* are independent and identically distributed. u^*_{it} cannot take any value. u^*_{it} can just take one of the N values $\{u_{it}\}_{i=1,...N}$.

Block Resampling Bootstrap

It is a direct accommodation of block bootstrap methods designed for time series. Non-overlapping block bootstrap (NMB) (Carlstein (1986)), moving block bootstrap (MBB) (Kunsch (1989), Liu and Singh (1992)), circular block bootstrap (CBB) (Politis and Romano (1992)) and stationary block bootstrap (SB) (Politis and Romano (1994)) can be adapted to panel data. The idea is to resample in time dimension, blocks of consecutive periods in order to capture temporal dependence. The Block bootstrap resampling is the operation of constructing a $N \times T$ matrix U^* with columns obtained by resampling with replacement, blocks of columns of U.

Double Resampling Bootstrap

For a $N \times T$ matrix U, the double resampling is the operation of constructing a $N \times T$ matrix U^{**}[2]. with columns and rows obtained by resampling blocks of columns and rows of Y.The term *double* comes from the fact that the resampling can be made in two steps. In a first step, one dimension is taken into account : from U, an intermediate matrix U^* is obtained. Another resampling is then made in the second dimension : from U^* the final matrix U^{**} is obtained. It is a combination of the two previous resampling methods. Carvajal (2000) and Kapetanios (2008) suggest the double resampling in the special case when the length of blocks is one. They also improve this resampling method by Monte Carlo simulations, but give no theoretical support.

[2]Double stars $**$ are used to distinguish the estimator, the probability measure, the expectation and the variance induced by the double resampling.

Bootstrap Confidence Intervals

In the literature, there are several bootstrap confidence interval. The intervals commonly used are the *percentile interval* and the *percentile-t interval. Percentile Interval*

With each pseudo-sample Y_b^*, compute $\widehat{\beta}_b^*$ and the K statistics $r_k^{b*} = \widehat{\beta}_k^{b*} - \widehat{\beta}_k$. The empirical distribution of these $(B+1)$ realizations is :

$$R_k^*(x) = \frac{1}{B+1} \sum_{b=1}^{B+1} I\left(r_k^{b*} \leq x\right) \tag{3.9}$$

The *percentile* confidence interval of level $(1-\alpha)$ for the parameter $\widehat{\beta}_k$ is:

$$CI_{1-\alpha,k}^* = \left[\widehat{\beta}_k - r_{k,1-\alpha/2}^*; \widehat{\beta}_k + r_{k,\alpha/2}^*\right] \tag{3.10}$$

where $r_{k,\alpha/2}^*$ and $r_{k,1-\alpha/2}^*$ are respectively the $\alpha/2$-percentile and $(1-\alpha/2)$-percentile of R_k^*. B must be chosen so that $\alpha(B+1)/2$ is an integer. When R_k^*. is symmetric, $r_{k,\alpha/2}^* = -r_{k,1-\alpha/2}^*$ and the confidence interval becomes $CI_{1-\alpha,k}^* = \left[\widehat{\beta}_{k,\alpha/2}^*; \widehat{\beta}_{k,1-\alpha/2}^*\right]$ where $\widehat{\beta}_{k,\alpha/2}^*$ and $\widehat{\beta}_{k,1-\alpha/2}^*$ are respectively the $\alpha/2$-percentile and $(1-\alpha/2)$-percentile of the empirical distribution of $\left\{\widehat{\beta}_{k,b}^*\right\}_{b=1..B+1}$.

Percentile-t Interval With each pseudo-sample Y_b^*, compute $\widehat{\beta}_b^*$ and the K statistics

$$t_k^{b*} = \frac{\widehat{\beta}_k^{b*} - \widehat{\beta}_k}{\sqrt{\widehat{V}ar^*\left(\widehat{\beta}_k^{b*}\right)}} \tag{3.11}$$

The empirical distribution of the $(B+1)$ realizations of t_k^{b*} is :

$$G_k^*(x) = \frac{1}{B+1} \sum_{b=1}^{B+1} I\left(t_k^{b*} \leq x\right) \tag{3.12}$$

The *percentile-t* confidence interval of level $(1-\alpha)$ for the parameter $\widehat{\beta}_k$ is:

$$CI_{1-\alpha,k}^* = \left[\widehat{\beta}_k - \sqrt{\widehat{Var}^*\left(\widehat{\beta}_k^{b*}\right)}.t_{k,1-\frac{\alpha}{2}}^*; \widehat{\beta}_k - \sqrt{\widehat{Var}^*\left(\widehat{\beta}_k^{b*}\right)}.t_{k,\frac{\alpha}{2}}^*\right] \tag{3.13}$$

where $t_{k,\alpha/2}^*$ and $t_{k,1-\alpha/2}^*$ are respectively the $\alpha/2$-percentile and $(1-\alpha/2)$-percentile of G_k^*. The strength of the percentile-t interval is that it permits theoretical demonstrations about asymptotic refinements.

4 Theoretical Results

This section presents theoretical results about resampling methods exposed in section 3, using models specified in section 2.

Mimic Analysis

Davidson (2007) argues that a bootstrapping procedure must respect two golden rules. The first one being that the bootstrap Data Generating Process (DGP) must respect the null hypothesis when testing hypothesis. The second is that unless the test statistic is pivotal, the bootstrap DGP should be an estimate of the true DGP as possible. This means that the bootstrap data must *mimic* as possible the behavior of the original data.. To understand this finite sample property approach, we must bear in mind that bootstrap procedure was originally designed for small samples. A good resampling method for panel data models must mimic very well the behavior of the components of ν_{it}. The error terms takes the form of four matrices. This formal decomposition allows one to appreciate the impact of each resampling method.

$$U = \underset{[\mu]}{\begin{pmatrix} \widehat{\mu}_1 & .. & \widehat{\mu}_1 \\ \widehat{\mu}_2 & .. & \widehat{\mu}_2 \\ .. & .. & .. \\ \widehat{\mu}_N & .. & \widehat{\mu}_N \end{pmatrix}} + \underset{[f]}{\begin{pmatrix} \widehat{f}_1 & .. & \widehat{f}_T \\ \widehat{f}_1 & .. & \widehat{f}_T \\ .. & .. & .. \\ \widehat{f}_1 & .. & \widehat{f}_T \end{pmatrix}} + \underset{[\lambda]}{\begin{pmatrix} \widehat{\lambda}_1 \\ \widehat{\lambda}_2 \\ .. \\ \widehat{\lambda}_N \end{pmatrix}} \underset{[F]}{\begin{pmatrix} \widehat{F}_1 & .. & \widehat{F}_T \end{pmatrix}} + \underset{[\varepsilon]}{\begin{pmatrix} \widehat{\varepsilon}_{11} & .. & \widehat{\varepsilon}_{1T} \\ \widehat{\varepsilon}_{21} & .. & \widehat{\varepsilon}_{2T} \\ .. & .. & .. \\ \widehat{\varepsilon}_{N1} & .. & \widehat{\varepsilon}_{NT} \end{pmatrix}} \quad (4.1)$$

$$U^*_{cross} = [\mu]^*_{cross} + [f] + [\lambda]^*_{cross} [F] + [\varepsilon]^*_{cross} \quad (4.2)$$

$$U^*_{bl} = [\mu] + [f]^*_{bl} + [\lambda] [F]^*_{bl} + [\varepsilon]^*_{bl} \quad (4.3)$$

$$U^{**} = [\mu]^*_{cross} + [f]^*_{bl} + [\lambda]^*_{cross} [F]^*_{bl} + [\varepsilon]^{**} \quad (4.4)$$

Each line of $[\mu]$ contains T times the same value. Resampling $[\mu]$ on the cross-section dimension is equivalent to an i.i.d. resampling on $(\widehat{\mu}_1, ..., \widehat{\mu}_N)$. The cross-sectional resampling is also equivalent to i.i.d. resampling on $\left(\widehat{\lambda}_1, .., \lambda_N\right)$. The rows of $[f]$ and $[F]$ are identical, the cross-sectional resampling has no impact on $[f]$ and $[F]$. It treats $\left(\widehat{f}_1, ..., \widehat{f}_T\right)$ and $\left(\widehat{F}_1, ..., \widehat{F}_T\right)$ as constants. For *the temporal block resampling,* the analysis is symmetrical to the first case. It is equivalent to block resampling on $\left(\widehat{f}_1, ..., \widehat{f}_T\right)$ and $\left(\widehat{F}_1, ..., \widehat{F}_T\right)$. It treats $(\widehat{\mu}_1, .., \widehat{\mu}_N)$ and $\left(\widehat{\lambda}_1, .., \widehat{\lambda}_N\right)$ as constants. *The double resampling* is the resultant of the two previous methods. It is equivalent to i.i.d. resampling on $(\widehat{\mu}_1, .., \widehat{\mu}_N)$ *and* $\left(\widehat{\lambda}_1, .., \widehat{\lambda}_N\right)$ and block resampling on $\left(\widehat{f}_1, ..., \widehat{f}_T\right)$ and $\left(\widehat{F}_1, ..., \widehat{F}_T\right)$. The extension to mutiple factors is obvious. The strength of the double resampling is to replicate the behavior this error terms, without having to separate them. No obligation to know for example, the number of factors.

The impact of the double resampling is known for $[\mu]$, $[f]$, $[\lambda]$ and $[F]$. It is important to analyze the impact on the distribution of $[\varepsilon]$. It is commonly assumed that ε_{it} is an idiosyncratic shock. It is good idea to compare the properties of ε_{it}^{**} to the properties of ε_{it}^{*} obtained by a direct accommodation of the classical i.i.d. bootstrap resampling. Conditionally on $[\varepsilon]$, the elements of $[\varepsilon^{**}]$ are not all independent. In fact each element ε_{it}^{**} depends on the elements in its column and on its row. This link exists because elements in the same line belong to the same unit i and elements in the same column refer to the same period t. The next proposition analyzes the impact of this dependence.

Proposition 3 *: $\forall$ N, T, using CBB or i.i.d. resampling in time dimension, the double resampling bootstrap-variance is greater than the i.i.d. bootstrap-variance :*

$$Var^{**}\left(\widetilde{\varepsilon}^{**}\right) \geq Var^{*}\left(\widetilde{\varepsilon}^{*}\right) \tag{4.5}$$

I.i.d. and CBB resamplings in time dimension avoids the edge effects appearing with MMB or SB. It is important to mention two things about inequality *(4.5)*. First, no particular assumptions have been made about $[\varepsilon]$ thus the inequality holds also for $\widetilde{U}$ Second, *(4.5)* is a finite sample property : it holds for any sample size. The equality holds in *(4.5)* when $T = 1$ (cross-section data), $N = 1$ (time series), or $\left[Cov^{**}\left(\varepsilon_{it}^{**}, u_{jt}^{**}\right), Cov^{**}\left(\varepsilon_{it}^{**}, \varepsilon_{is}^{**}\right)\right] = (0, 0)$. If ε_{it} are i.i.d. *(A5)*, the best way to resample then is to use i.i.d. resampling. Instead of this, the double resampling gene

Consistency Analysis

There are several ways to prove consistency of a resampling method. For an overview, see Shao and Tu (1995, chap. 3). The method commonly used is to show that the distance between the cumulative distribution function on the classical estimator and the bootstrap estimator goes to zero when the sample grows-up. Different notions of distance can be used : sup-norm, Mallow's distance.... Sup-norm is the commonly used. The notations used for one dimension data must be to panel data, in order to be more formal. Because of multiple asymptotic distributions, there are several consistency definitions. A bootstrap method is said *consistent* for β if :

$$\sup_{x \in R^K} \left| P^*\left(\sqrt{NT}\left(\widehat{\beta}^* - \widehat{\beta}\right) \leq x\right) - P\left(\sqrt{NT}\left(\widehat{\beta} - \beta\right) \leq x\right)\right| \underset{NT \to \infty}{\overset{P}{\to}} 0 \tag{4.6}$$

or

$$\sup_{x \in R^K} \left| P^*\left(\sqrt{N}\left(\widehat{\beta}^* - \widehat{\beta}\right) \leq x\right) - P\left(\sqrt{N}\left(\widehat{\beta} - \beta\right) \leq x\right)\right| \underset{NT \to \infty}{\overset{P}{\to}} 0 \tag{4.7}$$

or

$$\sup_{x\in R^K}\left|P^*\left(\sqrt{T}\left(\widehat{\beta}^*-\widehat{\beta}\right)\leq x\right)-P\left(\sqrt{T}\left(\widehat{\beta}-\beta\right)\leq x\right)\right|\underset{NT\to\infty}{\overset{P}{\to}}0 \tag{4.8}$$

Definitions *4.6, 4.7* and *4.8* are given with convergence in probability ($\overset{P}{\to}$). This case implies a *weak consistency.* The case of almost surely ($a.s.$) convergence provides a *strong consistency.* These definitions of consistency does not require that the bootstrap estimator or the classical estimator has asymptotic distribution. The idea behind it, is the mimic analysis : when the sample grows, the bootstrap estimator *mimics* very well the behavior of the classical estimator. In the special when the sample mean asymptotic distribution is available, consistency can be established by showing that bootstrap-sample mean has the same distribution. The next proposition expresses this idea.

Proposition 4
Assume $\sqrt{NT}\left(\widehat{\beta}-\beta\right)\Longrightarrow L$ *and* $\sqrt{NT}\left(\widehat{\beta}^*-\widehat{\beta}\right)\overset{*}{\Longrightarrow}L^*$. *If* L^**and* L *are identical and continuous, then*

$$\sup_{x\in R^K}\left|P^*\left(\sqrt{NT}\left(\widehat{\beta}^*-\widehat{\beta}\right)\leq x\right)-P\left(\sqrt{NT}\left(\widehat{\beta}-\beta\right)\leq x\right)\right|\underset{NT\to\infty}{\overset{P}{\to}}0$$

Similar propositions similar can be formulated for definitions *4.7* and *4.8*. Following Proposition 1, for each resampling method, the methodology adopted is to find $\left(\widehat{\beta}^*-\widehat{\beta}\right)$ and to deduce its asymptotic distribution, using the appropriate scaling factor. The comparison of theses distributions with their classical asymptotic counterpart permits to find consistent and inconsistent bootstrap resampling methods for the different panel models.

$$\left(\widehat{\beta}^*-\widehat{\beta}\right)=\left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1}\widetilde{Z}^{/}\left[\widetilde{Y}^*-Y\right]=\left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1}\widetilde{Z}^{/}\left[\widetilde{U}^*-\widetilde{U}\right] \tag{4.9}$$

To appreciate the impact of the resampling methods on $\left[\widetilde{U}^*-\widetilde{U}\right]$, it is easier to see it before the pooling.

$$U^*_{cross}-U=\left([\mu]^*_{cross}-[\mu]\right)+\left([\lambda]^*_{cross}-[\lambda]\right)[F]+\left([\varepsilon]^*_{cross}-[\varepsilon]\right) \tag{4.10}$$

$$U^*_{bl}-U=\left([f]^*_{bl}-[f]\right)+[\lambda]\,[F]^*_{bl}+\left([\varepsilon]^*_{bl}-[\varepsilon]\right) \tag{4.11}$$

$$U^{**}-U=\left([\mu]^*_{cross}-[\mu]\right)+\left([f]^*_{bl}-[f]\right)+\left([\lambda]^*_{cross}[F]^*_{bl}-[\lambda]\,[F]\right)+\left([\varepsilon]^{**}-[\varepsilon]\right) \tag{4.12}$$

The centering drops[3] out $\left(\widehat{f}_1,...,\widehat{f}_T\right)$ in the case of the cross-sectional resampling. $(\widehat{\mu}_1,..,\widehat{\mu}_N)$ is dropped out with the temporal block resampling. The double resampling deals well with $[\mu]$, $[f]$, $[\lambda]$ and $[F]$.

[3]The results are given in the case of no rescaling (HC_0). With the others rescaling factors, the argument is asymptotic.

Proposition 5 *:*

1 - Assume that A1 to A4, A6,A7, B1 to B8 hold and also (A9) : $Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right) \underset{N\rightarrow\infty}{\overset{P}{\rightarrow}} 0$ *and* $\frac{N}{T}Cov^{**}\left(\varepsilon_{it}^{**},\varepsilon_{jt}^{**}\right) \underset{N,T\rightarrow\infty}{\overset{P}{\rightarrow}} 0$

When N and T go to infinity. in $[\varepsilon]^{**}$ *is negligible in the presence of cross-section random effect.*

2 - Assume that A7 holds and also (A10) : $Cov^{**}\left(\varepsilon_{it}^{**},\varepsilon_{jt}^{**}\right) \underset{N,T\rightarrow\infty}{\overset{P}{\rightarrow}} 0,$ *and*

$\frac{T}{N}Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right) \underset{N,T\rightarrow\infty}{\overset{P}{\rightarrow}} 0$ *.When N and T go to infinity* $[\varepsilon]^{**}$ *is negligible in the presence of temporal random effect.*

The additional assumptions limit the dependence that can be incorporated in $[\varepsilon]$. $Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right)$ converging to zero means there is no cross-section heterogeneity in $[\varepsilon]$. With CBB, $Cov^{**}\left(\varepsilon_{it}^{**},\varepsilon_{jt}^{**}\right) = Var^{*}\left(\overline{\varepsilon}_{.t}^{*}\right)$, then there must be no temporal heterogeneity in $[\varepsilon]$. Proposition 6 implies that in the presence of heterogeneity, the researcher can disregard the impact of the double resampling on $[\varepsilon]$. when N and T go to infinity. Assumption *A5* implies *A9* and *A10.* The consistency of the NMB, MBB or CBB need a assumption about the block l size.

B10 : $l^{-1} + lT^{-1} = o(1)$ as $T \rightarrow \infty$. *B10* The condition about the convergence of l has a heuristic interpretation. If l is bounded, the block bootstrap method fails to capture the real dependence among the data. On the other hand, if l goes to infinity at the same rate as T, there are not enough blocks to resample.

Proposition 6 *: Consistency of the cross-sectional bootstrap*

1 - In the presence of temporal heterogeneity, the cross-sectional bootstrap is inconsistent when $\frac{N}{T} \rightarrow \delta \in [0,\infty)$

2 - Assume that A1 to A4, A7, B1 to B8 hold. In the presence of cross-sectional and temporal heterogeneities, the cross-sectional bootstrap is consistent when $N,T \rightarrow \infty$ *with* $\frac{N}{T} \rightarrow \infty$

Proposition 7 *: Consistency of the block temporal bootstrap*

1 - In the presence of cross-sectional heterogeneity, the block bootstrap methods are inconsistent when $N,T \rightarrow \infty$ *with* $\frac{N}{T} \rightarrow \delta \in (0,\infty]$

2 - Assume that A1 to A4, A7, B1 to B8 hold. In the presence of cross-sectional and temporal heterogeneities, the block bootstrap methods are consistent when $N,T \rightarrow \infty$ *with* $\frac{N}{T} \rightarrow 0$

Proposition 8 *: Consistency of the double resampling bootstrap*

Assume that A1 t0 A4, A6 to A10 hold. In the presence of cross-sectional and/or temporal heterogeneity, the double resampling bootstrap is consistent when N and T go to infinity.

Confidence Interval Consistency

The definitions of consistency are given for the vector of parameters, but confidence intervals are given for each component of this vector. For a parameter θ, a bootstrap confidence interval of level $1-\alpha$ is consistent if $Plim\big(\theta \in CI^{*}_{1-\alpha}\big) = 1-\alpha$. This interval is *conservative* if this probability is strictly larger than $1-\alpha$. The definitions of consistency exposed above, imply the consistency of of the percentile confidence interval in the same asymptotic framework. If the asymptotic law is continuous, strictly increasing and symmetric, confidence interval using directly the percentile of $\left\{\widehat{\beta}^{*}_{k}\right\}$ is also consistent. The consistency of the percentile-t confidence interval requires the consistency of the bootstrap-variance estimator $Var^{*}\left(\widehat{\beta}^{*}_{k}\right)$.

5 Simulations

Data Generating Process for errors is the following : $\mu_i \sim i.i.d.N\,(0,1)$, $\lambda_i \sim i.i.d.N\,(0,1)$, $f_t \sim i.i.d.N\,(0,1)$, $\varepsilon_{it} \sim i.i.d.N\,(0,1)\,, F_t = \rho F_{t-1} + \eta_t$, $\eta_t \sim i.i.d.N\,\big(0,\big(1-\rho^2\big)\big)$ $\rho = 0.25$. Data Generating Process for data for regressors is the following : $\theta = 1$, $U_i \sim i.i.d.N\,(0,1)$, $W_t \sim i.i.d.N\,(0,1)\,, X_{it} \sim i.i.d.N\,(0,1)$. For each bootstrap resampling method, *999 Replications and 1000 Simulations* are used. four sample sizes are considered : $(N,T) = (10,10)\,,(20,20)\,,(30,30)\,,(50,50)$. The block sizes are respectively $l = 2,\ 2,\ 3,\ 5$ for $T = 10, 20, 30$ and 50. Tables 1 and 2 give rejection rates, for theoretical level $\alpha = 5\%$. The simulations confirm the theoretical results. The cross-sectional bootstrap performs well with one-way ECM.The block bootstrap performs well with temporal one-way ECM. The double resampling performs well with the general specification and all the sub-model specifications. The inference becomes more and more and accurate, when, the sample grows. The confidence interval appears conservative for the coefficient associated time varying coefficient when no temporal error term use. The same thing appear in cross-section dimension. The statistical point is the non-invertibility of Ω^{∞}_{β}. The choice of the block length is arbitrary in the simulation exercice. An optimal length choice for the double resampling is to be developed.

6 Conclusion

This paper considers the issue of bootstrap methods for panel data models. It is shown that resampling only in the cross-sectional dimension is not valid in presence of temporal heterogeneity. The block resampling only in the time dimension is not valid in presence of cross-sectional heterogeneity. The double resampling that combines the two previous resampling methods, is valid for panel data models with cross-sectional and/or temporal heterogeneity, with or without spatial dependence. This approach also avoids multiple asymptotics. The strength of the double resampling is to replicate the behavior of the main components of error terms, without having to separate them. All the properties demonstrated for the i.i.d. bootstrap or for the various block bootstrap methods are transferred to the appropriate error terms without restriction. The recent literature on bootstrap methods for time series can be accommodated to the double resampling if the effect on the idiosyncratic shock is well mastered.

		$(N;T)$	=	$(10;10)$	$(N;T)$	=	$(20;20)$
		Cros.	Bloc.	D-Res	Cros.	Bloc.	D-Res
	θ	**0.135**	0.634	**0.101**	**0.088**	0.680	**0.074**
Cross	τ	**0.135**	0.619	**0.054**	**0.081**	0.669	**0.033**
1-way	γ	**0.086**	0.122	**0.015**	**0.067**	0.063	**0.004**
ECM	ζ	**0.087**	0.090	**0.062**	**0.051**	0.054	**0.037**
Temp.	θ	0.685	**0.218**	**0.165**	0.731	**0.171**	**0.155**
1-way	τ	0.131	**0.139**	**0.038**	0.091	**0.099**	**0.034**
ECM	γ	0.573	**0.115**	**0.089**	0.668	**0.084**	**0.070**
	ζ	0.091	**0.087**	**0.066**	0.064	**0.073**	**0.063**
	θ	0.291	0.321	**0.131**	0.271	0.252	**0.104**
2-way	τ	0.120	0.628	**0.056**	0.074	0.682	**0.037**
ECM	γ	0.539	0.126	**0.098**	0.693	0.076	**0.066**
	ζ	0.088	0.087	**0.060**	0.070	0.057	**0.054**
2-way	θ	0.254	0.313	**0.097**	0.237	0.251	**0.089**
ECM	τ	0.124	0.509	**0.045**	0.069	0.587	**0.026**
with spatial	γ	0.466	0.110	**0.068**	0.603	0.082	**0.068**
dependence	ζ	0.071	0.100	**0.056**	0.078	0.081	**0.062**

Table 1 : Simulations results with percentile interval.

		(N;T)	=	(30;30)	(N;T)	=	(50;50)
		Cros.	Bloc.	D-Res	Cros.	Bloc.	D-Res
	θ	**0.076**	0.676	**0.065**	**0.064**	0.803	**0.058**
Cross	τ	**0.077**	0.774	**0.064**	**0.054**	0.797	**0.030**
1-way	γ	**0.059**	0.098	**0.015**	**0.054**	0.059	**0.014**
ECM	ζ	**0.053**	0.420	**0.059**	**0.065**	0.058	**0.069**
Temp.	θ	0.809	**0.121**	**0.110**	0.847	**0.070**	**0.065**
1-way	τ	0.066	**0.084**	**0.012**	0.045	**0.067**	**0.021**
ECM	γ	0.717	**0.072**	**0.066**	0.763	**0.059**	**0.056**
	ζ	0.337	**0.064**	**0.058**	0.042	**0.041**	**0.038**
	θ	0.248	0.226	**0.065**	0.261	0.216	**0.071**
2-way	τ	0.071	0.739	**0.052**	0.073	0.790	**0.038**
ECM	γ	0.744	0.069	**0.047**	0.784	0.072	**0.068**
	ζ	0.243	0.282	**0.055**	0.070	0.047	**0.051**
2-way	θ	0.231	0.226	**0.065**	0.228	0.205	**0.067**
ECM	τ	0.077	0.672	**0.035**	0.066	0.726	**0.036**
with spatial	γ	0.688	0.081	**0.068**	0.758	0.052	**0.049**
dependence	ζ	0.048	0.056	**0.042**	0.053	0.057	**0.051**

Table 2 : Simulations results with percentile interval.

Acknowledgements

For their comments and suggestions, the author is grateful to Benoit Perron and Silvia Gonçalves, the participants of the 2nd Granger Centre Conference (Nottingham, U.K., Sept. 2008), the participants of the Panel Session of 18th Meeting of the Midwest Econometrics Group (Lawrence, Kansas, USA, October, 2008) and some participants of the Econometric Theory Session of the Latin American Meeting of the Econometric Society (Rio de Janeiro, Brazil, November 2008).

References

[1] Anselin, L. (1988). Spatial Econometrics : Methods and Models, Kluwer Academic Publishers, Dordrecht.

[2] Bai, J. and Ng S. (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica 72*, 4, 1127-1177.

[3] Balestra, P. and Nerlove, M. (1966). Pooling Cross Section and Time Series data in the estimation of a Dynamic Model : The Demand for Natural Gas. *Econometrica 34*, 3, 585-612.

[4] Baltagi, B. (1995). Econometrics Analysis of Panel Data, John Wiley & Sons.

[5] Carlstein, E. (1986). The Use of Subseries Methods for Estimating the Variance of a General Statistic from a Stationary Time Series. *The Annals of Statistics 14*, 1171-1179.

[6] Carvajal, A. (2000). BC Bootstrap Confidence Intervals for Random Effects Panel Data Models, Working Paper Brown University, Department of Economics.

[7] Davidson, R. (2007). Bootstrapping Econometric Models, This paper has appeared, translated into Russian, in *Quantile*, 3, 13-36.

[8] Driscoll, J. C. and . Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics 80*, 4 , 549–560.

[9] Efron, B. (1979). Bootstrap Methods : Another Look at the Jackknife. *The Annals of Statistics 7,* 1, 1-26.

[10] Fuller, W. A. and Battese, G. E. (1974). Estimation of Linear Models with Crossed-error Structure. *Journal of Econometrics 2*, 1, 67-78.

[11] Gonçalves S. (2008) The Moving Blocks Bootstrap for Panel Regression Models with Individual Fixed Effects, Mimeo, Department of Economics, Université de Montréal

[12] Hounkannounon B. (2008) Bootstrap for Panel Data, Working Paper, Department of Economics, Université de Montréal.

[13] Hsiao, C. (2003). Analysis of Panel Data, 2nd Edition. Cambridge University Press.

[14] Ibragimov, I. A. (1962). Some Limit Theorems for Stationary Processes. *Theory of Probability and Its Applications 7*, 349–382.

[15] Kapetanios, G. (2008). A Bootstrap Procedure for Panel Datasets with Many Cross-sectional Units. *Econometrics Journal 11*, 2, 377-395.

[16] Künsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics 17*, 1217–1241.

[17] Lahiri, S. N. (2003). Resampling Methods for Dependent Data, Springer Series in Statistics, Springer-Verlag, New York.

[18] Liu, R. Y. and Singh, K. (1992). Moving Block Jackknife and Bootstrap Capture Weak Dependence. In Lepage and Billard (Eds.), *Exploring the Limits of Bootstrap*, John Wiley, NewYork.

[19] MacKinnon, J.G. and White, H. L. (1985). Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics 21*, 53-70.

[20] Moon, H. R. and Perron, B. (2004). Testing for Unit Root in Panels with Dynamic Factors. *Journal of Econometrics 122*, 81-126.

[21] Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica 46*, 1, 69-85.

[22] Phillips, P.C.B. and Sul, D. (2003). Dynamic Panel Estimation and Homogeneity Testing Under Cross-section Dependence. *Econometrics Journal 6*, 217-259.

[23] Phillips, P.C.B. and Moon, H.R. (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica 67*, 5,1057-1112.

[24] Politis, D. and Romano, J. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association 89*, 1303-1313.

[25] Politis, D. and Romano J. (1992). A Circular Block Resampling Procedure for Stationary Data. In Lepage and Billard (Eds.), *Exploring the Limits of Bootstrap*, John Wiley, NewYork.

[26] Pólya, G. (1920). Uber den Zentralen Grenzwertsatz der Wahrscheinlinchkeitsrechnug und Momenten Problem. *Math. Zeitschrift 8.*

[27] Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. New York: Wiley.

[28] Shao, J. and Tu, D. (1995). The Jackknife and Bootstrap, Sinpring-Verlag, New York.

7 APPENDIX

Proof of Proposition 1

$1-\frac{N}{T}\rightarrow\delta\in[0,\infty)$

$$\sqrt{N}\left(\widehat{\beta}-\beta\right) = \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\mu_i+\frac{\sqrt{N}}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_t\right] + \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\left(\lambda_iF_t+\varepsilon_{it}\right)\right)\right]$$

$$\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\left(\lambda_iF_t+\varepsilon_{it}\right)\right)\underset{N,T\rightarrow\infty}{\overset{m.s.}{\rightarrow}}0$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\mu_i\underset{N\rightarrow\infty}{\Longrightarrow}N\left(0,\sigma_\mu^2\overline{Q}\right)\quad\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_t\underset{T\rightarrow\infty}{\Longrightarrow}N\left(0,\Omega_{f\widetilde{Z}}^{\infty}\right)$$

$$\sqrt{N}\left(\widehat{\beta}-\beta\right)\underset{N,T\rightarrow\infty}{\Longrightarrow}N\left(0,\sigma_\mu^2\left(Q^{-1}\overline{Q}Q^{-1}\right)+\delta\left(Q^{-1}\Omega_{f\widetilde{Z}}^{\infty}Q^{-1}\right)\right)$$

This asymptotic distribution of $\underline{Z}_{(t)}^{/}f_t$ *is a multivariate case of Ibragimov(1962).*

$2-\ \frac{N}{T}\rightarrow\infty$

$$\sqrt{T}\left(\widehat{\beta}-\beta\right) = \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\frac{\sqrt{T}}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\mu_i\right)+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_t\right] + \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[+\frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\left(\lambda_iF_t+\varepsilon_{it}\right)\right)\right]$$

$$\frac{\sqrt{T}}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\mu_i\right)\underset{N,T\rightarrow\infty}{\overset{m.s.}{\rightarrow}}0$$

The result follows.

Proof of Proposition 2

$$\begin{aligned}\sqrt{NT}\left(\widehat{\beta}-\beta\right) &= \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\lambda_iF_t\right)+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\varepsilon_{it}\right]\\ &= \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\left(\frac{1}{\sqrt{N}}\sum_{i=1i}^{N}\overline{Z}_{(i)}^{/}\lambda_i\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}F_t\right)+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\varepsilon_{it}\right]\end{aligned}$$

The result follows by classical CLT.

Proof of Proposition 3

$$\underset{(NT,NT)}{Var^{**}}\left(\widetilde{\varepsilon}^{**}\right) = \Omega_v + \Omega_{cros} + \Omega_{temp}$$
$$\Omega_v = diag\left[Var^{**}\left(\varepsilon_{it}^{**}\right)\right]_{i,j=1...N}^{t,s=1...T}$$
$$\Omega_{cros} = \left[Cov^{**}\left(\varepsilon_{it}^{**}, \varepsilon_{is}^{**}\right) \times 1_{i=j}\right]_{i,j=1...N}^{t,s=1...T}$$
$$\Omega_{temp} = \left[Cov^{**}\left(\varepsilon_{it}^{**}, \varepsilon_{jt}^{**}\right) \times 1_{t=s}\right]_{i,j=1...N}^{t,s=1...T}$$

For CBB or i.i.d. resampling in the time dimension, ε_{it}^{**} *can take any of the* NT *values of elements of* $[\varepsilon]$ *with probability* $1/NT$ *then the expectation and the variance are identical to those obtained with i.i.d. bootstrap accommodated to panel data :* $E^{**}\left(\varepsilon_{it}^{**}\right) = E^{*}\left(\varepsilon_{it}^{*}\right), Var^{**}\left(\varepsilon_{it}^{**}\right) = Var^{*}\left(\varepsilon_{it}^{*}\right).$

$$Cov^{**}\left(\varepsilon_{it}^{**}, \varepsilon_{jt}^{**}\right) = \frac{1}{N^2T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{jt} - \left(\overline{\overline{\widehat{\varepsilon}}}\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\overline{\widehat{\varepsilon}}_{.t}\right)^2 - \left(\overline{\overline{\widehat{\varepsilon}}}\right)^2 = Var^{*}\left(\overline{\varepsilon}_{.t}^{*}\right) \geq 0$$

Similary $Cov^{**}\left(\varepsilon_{it}^{**}, \varepsilon_{is}^{**}\right) = Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right) \geq 0$, *thus*

$$\underset{(NT,NT)}{Var^{**}}\left(\widetilde{\varepsilon}^{**}\right) = \Omega_v + \Omega_{cros} + \Omega_{temp} \geq \underset{(NT,NT)}{Var^{*}}\left(\widetilde{\varepsilon}^{*}\right) = \Omega_v$$

Proof of Proposition 4

The fact that $\widehat{\beta}$ *and* $\widehat{\beta}^{*}$ *have the same asymptotic distribution, implies that* $\left|P^{*}(..) - P(..)\right|$ *converges to zero. Under continuity assumption, the uniform convergence is given by Pólya theorem (Pólya (1920) or Serfling (1980), p. 18)* endproof

Proof of Proposition 5

The presence of random effect oblige to use $\sqrt{N}$ or $\sqrt{T}$ as scaling factor of $\left(\widehat{\beta}^{**} - \widehat{\beta}\right)$. *A9 and A10 imply that the elements of* $Var^{**}\left(\widetilde{\varepsilon}^{**}\right)$ converges to zero when .$\sqrt{N}$ or $\sqrt{T}$ is used.

Proof of Proposition 6

$1-$ $\left([f]_{bl}^{*} - [f]\right)$ does not appear in $\left(\widehat{\beta}_{cros}^{*} - \widehat{\beta}\right) = \left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1}\widetilde{Z}^{/}\left[\widetilde{U}_{cross}^{*} - \widetilde{U}\right]$ but Σ_f^{∞} is present in the asymptotique distribution of $\widehat{\beta}$.

$$\sqrt{N}\left(\widehat{\beta} - \beta\right) \underset{\substack{N,T\to\infty \\ \frac{N}{T}\to\delta\in[0,\infty)}}{\Longrightarrow} N\left(0, \sigma_\mu^2\left(Q^{-1}\overline{Q}Q^{-1}\right) + \delta.\Sigma_f^{\infty}\right)$$

$2-$*When* $\frac{N}{T} \to 0$, Σ_f^{∞} vanishes also inthe classical asymptotic distribution thus : consistency.

Proof of Proposition 7

The proof is the symmetric of the previous.

Proof of Proposition 8

$$\sqrt{N}\left(\widehat{\beta}^{**}-\widehat{\beta}\right)=\left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\begin{array}{c}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\left(\mu_i^{*}-\widehat{\mu}_i\right)+\frac{\sqrt{N}}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}\left(f_{bl,t}^{*}-\widehat{f}_t\right)\\+\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\left(\left(\lambda_i^{*}F_{bl,t}^{*}-\widehat{\lambda}_i\widehat{F}_t\right)+\left(\varepsilon_{it}^{**}-\widehat{\varepsilon}_{it}\right)\right)\right)\end{array}\right]$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\left(\mu_i^{*}-\widehat{\mu}_i\right)\underset{N\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\sigma_\mu^2Q^{-1}\right)\quad\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}\left(f_{bl,t}^{*}-\widehat{f}_t\right)\underset{T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\Omega_{f\widetilde{Z}}^{\infty}\right)$$

This results hold using usual bootstrap asymptotic properties in one dimension (cross-section or temporal)

$$\sqrt{N}\left(\widehat{\beta}^{**}-\widehat{\beta}\right)\underset{\substack{N,T\to\infty\\ \frac{N}{T}\to\delta\in[0,\infty)}}{\overset{*}{\Longrightarrow}}N\left(0,\sigma_\mu^2\left(Q^{-1}\overline{Q}Q^{-1}\right)+\delta\Sigma_f^{\infty}\right)\quad\sqrt{T}\left(\widehat{\beta}^{**}-\widehat{\beta}\right)\underset{\substack{N,T\to\infty\\ \frac{N}{T}\to\infty}}{\overset{*}{\Longrightarrow}}N\left(0,\Sigma_f^{\infty}\right)$$

Covariance Matrix Analysis

The asymptotic variance of $\widehat{\beta}$ is $V_{\widehat{\beta}}^{\infty}=Q^{-1}\left(\sigma_\mu^2\overline{Q}+\delta.\Omega_{fZ}^{\infty}\right)Q^{-1}$. The invertibility of the asymptotic variance of $\widehat{\beta}$ depends on the choice of the regressors and the composition the structure of the error.

$$\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}=\frac{1}{NT}\begin{bmatrix}\underset{(1,NT)}{1^{/}}\\\underset{(K_1,NT)}{\widetilde{V}^{/}}\\\underset{(K_2,NT)}{\widetilde{W}^{/}}\\\underset{(K_3,NT)}{\widetilde{X}^{/}}\end{bmatrix}\begin{bmatrix}\underset{(NT,1)}{1}&\underset{(NT,K_1)}{\widetilde{V}}&\underset{(NT,K_2)}{\widetilde{W}}&\underset{(NT,K_3)}{\widetilde{X}}\end{bmatrix}$$

$$=\begin{bmatrix}1&\overline{V}&\overline{W}&\overline{\overline{X}}\\\overline{V}^{/}&\frac{\widetilde{V}^{/}\widetilde{V}}{NT}&\frac{\widetilde{U}^{/}\widetilde{W}}{NT}&\frac{\widetilde{V}^{/}\widetilde{X}}{NT}\\\overline{W}^{/}&\frac{\widetilde{W}^{/}\widetilde{V}}{NT}&\frac{\widetilde{W}^{/}\widetilde{W}}{NT}&\frac{\widetilde{W}^{/}\widetilde{X}}{NT}\\\overline{\overline{X}}^{/}&\frac{\widetilde{X}^{/}\widetilde{V}}{NT}&\frac{\widetilde{X}^{/}\widetilde{W}}{NT}&\frac{\widetilde{X}^{/}\widetilde{X}}{NT}\end{bmatrix}=\begin{bmatrix}1&0&0&0\\0&\frac{V^{/}V}{N}&0&\frac{\widetilde{V}^{/}\widetilde{X}}{NT}\\0&0&\frac{W^{/}W}{T}&\frac{\widetilde{W}^{/}\widetilde{X}}{NT}\\0&\frac{\widetilde{X}^{/}\widetilde{V}}{NT}&\frac{\widetilde{X}^{/}\widetilde{W}}{NT}&\frac{X^{/}X}{NT}\end{bmatrix}\to Q$$

$$\underset{(1,K)}{\overline{Z}_{(i)}}=\frac{1}{T}\sum_{t=1}^{T}Z_{(it)}\quad,\quad\underset{(1,K)}{\underline{Z}_{(t)}}=\frac{1}{N}\sum_{i=1}^{N}Z_{(it)}$$

$$\overline{Z} = \begin{bmatrix} \underset{(N,1)}{1} & \underset{(N,K_1)}{V} & \underset{(N,K_2)}{\overline{W}} & \underset{(N,K_3)}{\overline{X}} \end{bmatrix} = \begin{bmatrix} \underset{(N,1)}{1} & \underset{(N,K_1)}{V} & \underset{(N,K_2)}{0} & \underset{(N,K_3)}{\overline{X}} \end{bmatrix}$$

$$\frac{\overline{Z}^{/}\overline{Z}}{N} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{V^{/}V}{N} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\overline{X}^{/}\overline{X}}{N} \end{bmatrix} \to \overline{Q}$$

With cross-section heterogeneity only, $V_{\widehat{\beta}}^{\infty} = \sigma_{\mu}^{2}\left(Q^{-1}\overline{Q}Q^{-1}\right)$ is not invertible if Z contains regressors varying in time dimension only.

$$\underline{Z} = \begin{bmatrix} \underset{(T,1)}{1} & \underset{(T,K_1)}{\underline{V}} & \underset{(T,K_2)}{W} & \underset{(N,K_3)}{\underline{X}} \end{bmatrix} = \begin{bmatrix} \underset{(T,1)}{1} & \underset{(T,K_1)}{0} & \underset{(T,K_2)}{W} & \underset{(T,K_3)}{\underline{X}} \end{bmatrix}$$

$$\frac{\underline{Z}^{/}\underline{Z}}{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{W^{/}W}{T} & 0 \\ 0 & 0 & 0 & \frac{\underline{X}^{/}\underline{X}}{T} \end{bmatrix} \to \underline{Q}$$

With temporal heterogeneity only $V_{\widehat{\beta}}^{\infty} = Q^{-1}\left(\Omega_{fZ}^{\infty}\right)Q^{-1}$ with $\Omega_{fZ}^{\infty} = Lim\ Var\left(\underline{Z}^{/}f\right) = Lim\left(\underline{Z}^{/}\Omega_f\underline{Z}^{/}\right)$. $\Omega_f = Var(f)$ is symmetric definite positive thus by Cholesky decomposition $\Omega_f = T^{/}T$ with T, a triangular matrix. $\underset{(K,T)(T,T)(T,K)}{\underline{Z}^{/}\Omega_f\underline{Z}}{}^{/} = \underline{Z}^{/}T^{/}T\underline{Z} = \underset{(K,K)}{\|T\underline{Z}\|}$. $V_{\widehat{\beta}}^{\infty}$ is not invertible if Z contains regressors varying in cross-section only.