Université de Montréal

Bootstrap for panel data models with an application to the evaluation of public policies

par

Bertrand G. B. Hounkannounon

Département de sciences économiques

Faculté des arts et des sciences

Thèse présentée à la Faculté arts et des sciences
en vue de l'obtention du grade de
Philosophiae Doctor (Ph.D.)
en sciences économiques

Août 2011

Université de Montréal
Faculté des arts et des sciences

Cette thèse intitulée :

Bootstrap for panel data models with an application to the evaluation of public policies

présentée par

Bertrand G. B. Hounkannounon

a été évaluée par un jury composé des personnes suivantes :

..........................Marine Carrasco............................

président-rapporteur

..........................Benoit Perron............................

directeur de recherche

............................Silvia Gonçalves................................

membre du jury

....................Nikolay Gospodinov.....................

examinateur externe

...........................Martin Bilodeau.........................

représentant du doyen de la FAS

Résumé

Le but de cette thèse est d'étendre la théorie du bootstrap aux modèles de données de panel. Les données de panel s'obtiennent en observant plusieurs unités statistiques sur plusieurs périodes de temps. Leur double dimension individuelle et temporelle permet de contrôler l'hétérogénéité non observable entre individus et entre les périodes de temps et donc de faire des études plus riches que les séries chronologiques ou les données en coupe instantanée. L'avantage du bootstrap est de permettre d'obtenir une inférence plus précise que celle avec la théorie asymptotique classique ou une inférence impossible en cas de paramètre de nuisance. La méthode consiste à tirer des échantillons aléatoires qui ressemblent le plus possible à l'échantillon d'analyse. L'objet statitstique d'intérêt est estimé sur chacun de ses échantillons aléatoires et on utilise l'ensemble des valeurs estimées pour faire de l'inférence. Il existe dans la littérature certaines application du bootstrap aux données de panels sans justification théorique rigoureuse ou sous de fortes hypothèses. Cette thèse propose une méthode de bootstrap plus appropriée aux données de panels. Les trois chapitres analysent sa validité et son application.

Le premier chapitre postule un modèle simple avec un seul paramètre et s'attaque aux propriétés théoriques de l'estimateur de la moyenne. Nous montrons que le double rééchantillonnage que nous proposons et qui tient compte à la fois de la dimension individuelle et la dimension temporelle est valide avec ces modèles. Le rééchantillonnage seulement dans la dimension individuelle n'est pas valide en présence d'hétérogénéité temporelle. Le rééchantillonnage dans la dimension temporelle n'est pas valide en présence d'hétérogénéité individuelle.

Le deuxième chapitre étend le précédent au modèle panel de régression

linéaire. Trois types de régresseurs sont considérés : les caractéristiques individuelles, les caractéristiques temporelles et les régresseurs qui évoluent dans le temps et par individu. En utilisant un modèle à erreurs composées doubles, l'estimateur des moindres carrés ordinaires et la méthode de bootstrap des résidus, on montre que le rééchantillonnage dans la seule dimension individuelle est valide pour l'inférence sur les coefficients associés aux régresseurs qui changent uniquement par individu. Le rééchantillonnage dans la dimension temporelle est valide seulement pour le sous vecteur des paramètres associés aux régresseurs qui évoluent uniquement dans le temps. Le double rééchantillonnage est quand à lui est valide pour faire de l'inférence pour tout le vecteur des paramètres.

Le troisième chapitre re-examine l'exercice de l'estimateur de différence en différence de Bertrand, Duflo et Mullainathan (2004). Cet estimateur est couramment utilisé dans la littérature pour évaluer l'impact de certaines politiques publiques. L'exercice empirique utilise des données de panel provenant du Current Population Survey sur le salaire des femmes dans les 50 états des Etats-Unis d'Amérique de 1979 à 1999. Des variables de pseudo-interventions publiques au niveau des états sont générées et on s'attend à ce que les tests arrivent à la conclusion qu'il n'y a pas d'effet de ces politiques placebos sur le salaire des femmes. Bertrand, Duflo et Mullainathan (2004) montre que la non-prise en compte de l'hétérogénéité et de la dépendance temporelle entraîne d'importantes distorsions de niveau de test lorsqu'on évalue l'impact de politiques publiques en utilisant des données de panel. Une des solutions préconisées est d'utiliser la méthode de bootstrap. La méthode de double rééchantillonnage développée dans cette thèse permet de corriger le problème de niveau de test et donc d'évaluer correctement l'impact des politiques publiques.

Mots clés : Modèles de données de panel, Bootstrap, Evaluation de Politiques Publiques.

Abstract

The purpose of this thesis is to develop bootstrap methods for panel data models and to prove their validity. Panel data refers to data sets where observations on individual units (such as households, firms or countries) are available over several time periods. The availability of two dimensions (cross section and time series) allows for the identification of effects that could not be accounted for otherwise. In this thesis, we explore the use of the bootstrap to obtain estimates of the distribution of statistics that are more accurate than the usual asymptotic theory. The method consists in drawing many random samples that resembles the sample as much as possible and estimating the distribution of the object of interest over these random samples. It has been shown, both theoretically and in simulations, that in many instances, this approach improves on asymptotic approximations. In other words, the resulting tests have a rejection rate close to the nominal size under the null hypothesis and the resulting confidence intervals have a probability of including the true value of the parameter that is close to the desired level.

In the literature, there are many applications of the bootstrap with panel data, but these methods are carried out without rigorous theoretical justification. This thesis suggests a bootstrap method that is suited to panel data (which we call double resampling), analyzes its validity, and implements it in the analysis of treatment effects. The aim is to provide a method that will provide reliable inference without having to make strong assumptions on the underlying data-generating process.

The first chapter considers a model with a single parameter (the overall expectation) with the sample mean as estimator. We show that our double resampling is valid for panel data models with some cross section and/or

temporal heterogeneity. The assumptions made include one-way and two-way error component models as well as factor models that have become popular with large panels. On the other hand, alternative methods such as bootstrapping cross-sections or blocks in the time dimensions are only valid under some of these models.

The second chapter extends the previous one to the panel linear regression model. Three kinds of regressors are considered : individual characteristics, temporal characteristics and regressors varying across periods and cross-sectional units. We show that our double resampling is valid for inference about all the coefficients in the model estimated by ordinary least squares under general types of time-series and cross-sectional dependence. Again, we show that other bootstrap methods are only valid under more restrictive conditions.

Finally, the third chapter re-examines the analysis of differences-in-differences estimators by Bertrand, Duflo and Mullainathan (2004). Their empirical application uses panel data from the Current Population Survey on wages of women in the 50 states. Placebo laws are generated at the state level, and the authors measure their impact on wages. By construction, no impact should be found. Bertrand, Duflo and Mullainathan (2004) show that neglected heterogeneity and temporal correlation lead to spurious findings of an effect of the Placebo laws. The double resampling method developed in this thesis corrects these size distortions very well and gives more reliable evaluation of public policies.

Key words : Panel data models, Bootstrap, Public Policy Evaluation.

Table des matières

Résumé i
Abstract iv
Liste des figures viii
Dédicace ix
Remerciements x

Introduction Générale **1**

1 Double resampling bootstrap for the mean of a panel **5**
1.1 Introduction 7
1.2 Panel Data Models and Assumptions 8
1.3 Asymptotic Theory 14
1.4 Resampling Methods 17
1.5 Bootstrap Validity 22
1.6 Bootstrap Confidence Interval 29
1.7 Simulations 33
1.8 Conclusion 35

2 Bootstrap for panel regression models with random effects **57**
2.1 Introduction 59
2.2 Panel Data Models 61

2.3 Bootstrap Methods 67

2.4 Theoretical Results 73

2.5 Simulations 78

2.6 Conclusion 83

3 Bootstrapping differences-in-differences estimator 94

3.1 Introduction 96

3.2 Differences-in-differences Estimation 97

3.3 Bootstrap Method 99

3.3.1 Residual-based Bootstrap 99

3.3.2 Pair bootstrap 100

3.3.3 Bootstrap Confidence Intervals 101

3.3.4 Panel Resampling Methods 101

3.4 Empirical Application 106

3.4.1 Specification 106

3.4.2 Placebo Laws 108

3.4.3 Simulation Results 109

3.5 Conclusion 111

Conclusion Générale 113

Bibliographie 115

Table des figures

3.1 Time Evolution of Wage by State 108

A la mémoire du Professeur Henri Michel Hounkannounon

Remerciements

Je tiens à rendre hommage à mon directeur de recherche Benoit Perron pour sa disponibilité, sa patience, sa contribution à la réalisation de ce travail. Je remercie les professeurs Silvia Gonçalves et Marine Carrasco pour les discussions et les commentaires constructifs.

Mes remerciements vont aussi au Département de sciences économiques de l'Université de Montréal et au Centre Interuniversitaire de Recherche en Economie Quantitative (CIREQ) pour le soutien financier et la bonne ambiance de recherche.

Un remerciement tout spécial à ma famille et mes ami(e)s pour l'encouragement et le soutien moral. Que tous ceux qui de près ou de loin ont contribué à ce travail, reçoivent l'expression de ma sincère gratitude.

Introduction générale

Les données de panel s'obtiennent en observant plusieurs unités statistiques sur plusieurs périodes de temps. Leur double dimension individuelle et temporelle permet de contrôler l'hétérogénéité non observable entre individus et entre les périodes de temps. Ceci permet de faire des analyses difficilement faisables avec juste des séries temporelles ou des coupes transversales de données. L'inférence avec les modèles de panel, comme dans tout autre modèle statistique nécessite le recours à des statistiques de test. En pratique, la véritable distribution de probabilité d'une statistique de test est rarement connue. En général, nous utilisons la loi asymptotique comme approximation de la vraie loi. Si la taille de l'échantillon n'est pas assez grande, le comportement asymptotique de la statistique pourrait être une mauvaise approximation de la réalité.

Un avantage important de la technique de rééchantillonnage bootstrap est de permettre d'obtenir une approximation de la distribution d'une statistique de test plus précise que l'approximation asymptotique lorsque la taille de l'échantillon est faible. Cette technique a été proposée originalement pour l'analyse statistique des observations indépendantes et identiquement distribuées. Des extensions ont été faites pour l'adapter à l'analyse de données avec de la dépendance entre les observations comme les séries temporelles. La littérature sur l'utilisation du bootstrap avec des données de panel est assez restreinte. On note des exemples d'utilisation sans justifications théoriques ou quand ces justifications existent c'est pour des cas très particuliers. Comme contribution récente à la littérature théorique, nous pouvons citer Kapetanios (2008) et Gonçalves (2010).

La double dimension des modèles de panels pose néanmoins quelques défis en pratique : les théories asymptotiques multiples. La façon dont on suppose que le nombre d'unités statistiques (N) et/ou le nombre de périodes temporelles (T) tend vers l'infini, n'est pas sans consequence sur la distribution asymptotique obtenue. En pratique, face à un échantillon particulier, il n'y a pas de méthode pour choisir laquelle des distributions il faut utiliser. Le recours à une méthodologie qui ne diffère pas d'une distribution asymptotique à l'autre permettrait de contourner ce genre de problème.

L'évaluation de politiques publiques amène à considerer deux groupes d'individus : ceux qui sont par affectés par la politique (groupe de traitement) et ceux qui ne sont pas affectés (groupe de contrôle). Le second groupe sert de groupe de témoins et permet de contrôler des effets temporels qui seraient produits en l'absence de la politique et permet d'apprécier à sa juste valeur, l'impact de de la politique publique (ou traitement). Dans l'approche la plus simpliste, on considère deux périodes de temps : une période avant la mise en place de la politique et une période après la mise en place. L'impact de la politique est mésuré en prenant la variation de la variable cible dans le groupe de traitement auquel on soustrait la variation dans le groupe de contrôle. Cette technique s'appelle la methode des différencesen différences (ou méthode des doubles différences). Elle serait tout à fait justifiée si les individus étaient affectés arbitrairement dans chacun des groupes. En réalité, la mise en place d'une intervention du pouvoir publique est justifiée par des besoins d'objectif à atteindre. Une localité va bénificier d'un projet particulier parce qu'on veut y réduire par exemple le taux de décrochage scolaire qui y est plus élevé que dans d'autres zone scolaires. Le choix des individus du groupe de contrôle et ceux qui sont dans le groupe de traitement est donc loin d'être arbitraire. L'appréciation du gain de la politique peut donc être est biaisé par un effet de sélection. Pour tenir compte du fait que l'appartenance à l'un des deux groupes peut dépendre des caractéristiques individuelles, il faut les inclure dans l'évaluation d'impact. L'approche générale est de postuler un modèle linéaire ou la variable d'intérêt y est fonction des caractériques individuelles et d'une variable indicatrice qui prend la valeur 1 quand l'individu est affectée par la politique en second période et la valeur 0 sinon. Le coefficient associé à cette variable indictarice mesure l'impact de la politique étudiée.

Dans une approche plus générale, on considère les deux mêmes groupes mais cette fois-ci, pendant plusieurs périodes de temps. Cette approche permet de mieux tenir compte de la dynamique temporelle et on à ce moment des données de panel. Une difficulté pratique dans l'évaluation des politiques publiques est la limitation du nombre d'observation. En effet, avant la mise en place à une plus grande échelle, une politique peut d'abord est testée sur un échantillon, le nombre d'unités statistiques impliquées est alors modéré. La nécessité d'avoir les premiers résultats d'un programme dans un laps de temps raisonnable limite le nombre de périodes de notre panel. Cette double restriction fragilise la qualité de l'inférence que l'on a recours à l'asymptotique. Malgré ces difficultés, le chercheur doit faire de son mieux pour tirer la meilleure information de l'échantillon dont il dispose. L'utilisation des méthodes de bootstrap peut accroître la qualité de l'inférence.

La présente thèse examine le développement de méthodes bootstrap appropriés aux modèles de panel, leurs justifications théoriques et applications. Le premier chapitre postule un modèle simple avec un seul paramètre et s'attaque aux propriétés théoriques de l'estimateur de la moyenne[1].

Le deuxième chapitre étend le précédent au modèle panel de régression linéaire. Trois types de régresseurs sont considérés : les caractéristiques individuelles, les caractéristiques temporelles et les régresseurs qui évoluent dans

[1] Il est commun de démontrer la validité d'une méthode de rééchantillonnage pour la moyenne avant de s'attaquer à sa validité pour des statistiques plus complexes.

le temps et par individu. En utilisant un modèle à erreurs composées doubles, l'estimateur des moindres carrés ordinaires et la méthode de bootstrap des résidus, on montre que le rééchantillonnage dans la seule dimension individuelle est valide pour l'inférence sur les coefficients associés aux régresseurs qui changent uniquement par individu. Le rééchantillonnage dans la dimension temporelle est valide seulement pour le sous vecteur des paramètres associé aux régresseurs qui évoluent uniquement dans le temps. Le double rééchantillonnage est quand à lui est valide pour faire de l'inférence pour tout le vecteur des paramètres.

Le troisième chapitre re-examine l'exercice de l'estimateur des doubles différences de Bertrand, Duflo et Mullainathan (2004). L'exercice empirique utilise des données de panel provenant du Current Population Survey sur le salaire des femmes dans les 50 états des Etats-Unis d'Amérique de 1979 à 1999. Des variables de pseudo-interventions publiques au niveau des états sont générées et on s'attend à ce que les tests arrivent à la conclusion qu'il n'y a pas d'effet de ces politiques placebos sur le salaire des femmes. Bertrand, Duflo et Mullainathan (2004) montre que la non-prise en compte de la dépendance temporelle entraîne d'importantes distorsions de niveau de test lorsqu'on évalue l'impact de politiques publiques en utilisant des données de panel. La méthode de double rééchantillonnage développée dans cette thèse permet de corriger le problème de niveau de test et donc d'évaluer correctement l'impact des politiques publiques.

Chapitre 1

Double resampling bootstrap for the mean of a panel

Abstract

This paper considers bootstrap methods for the sample mean in panel data. It is shown that double resampling that combines cross-sectional and temporal resampling is valid under general conditions on cross-sectional and temporal heterogeneity as well as cross-sectional dependence. On the other hand, resampling only in the cross section dimension is not valid in the presence of temporal heterogeneity, while block resampling only in the time series dimension is not valid in the presence of cross section heterogeneity. The bootstrap does not require the researcher to choose one of several asymptotic approximations available for panel models. Simulations confirm these theoretical results.

JEL Classification : C15, C23.

Keywords : Bootstrap, Panel Data Models.

1.1 Introduction

This paper analyzes properties of bootstrap methods in carrying out inference on the mean of panel data. The goal is to try to construct confidence intervals and conduct hypothesis tests without having to make strong assumptions regarding either the serial correlation or cross-sectional dependence of the data.

While there is an abundant literature on asymptotic theory for panel data models, there is much less on the bootstrap. There are some simulation results suggesting that some resampling methods work well in practice but theoretical results are rather limited or require strong assumptions. For example, Kapetanios (2008) recently presented theoretical results in a linear panel regression model when the cross-sectional dimension goes to infinity, under the assumption that cross-sectional vectors of regressors and errors terms are i.i.d.. This assumption is quite restrictive and does not allow time-varying regressors or temporal aggregate shocks in errors terms. Gonçalves (2010) explores the moving blocks bootstrap in a linear regression model as well, and Palm, Smeekes and Urbain (2011) develop the bootstrap for nonstationary panel models.

Asymptotic analysis in panel models is complicated by the fact we have cross-sectional and time series dimensions. Thus, several asymptotic approximations can be developed, depending on the assumptions one is willing to make on the size of these two dimensions. Typically, the resulting approximations will be different, forcing an applied researcher to choose among these various approximations in order to obtain a critical value for a hypothesis test. One of the main advantages of the bootstrap in the context of panel models is that it is not necessary to make such a choice. The bootstrap will provide valid critical values for various asymptotic scenarios under appropriate conditions.

The paper is organized as follows. In the second section, different panel data models are presented. Section 3 presents the asymptotic theory. Section 4 presents three bootstrap resampling methods for panel data. The fifth section presents theoretical results, analyzing the validity of each resampling method. The seventh section bootstrap confidence interval and analyzes their validity. In section 7, simulation results are presented and confirm the theoretical results. The eighth section concludes. Proofs of propositions are given in the appendix.

1.2 Panel Data Models and Assumptions

We suppose that we observe panel data y_{it} for cross-sectional unit i at time t. There are N cross-sectional units (typically households, firms or countries) and T time periods. One could consider unbalanced panels where the number of observations for each unit would differ, but for simplicity, we do not consider this case.

In this chapter, we consider a panel model without regressors.

$$y_{it} = \theta + \nu_{it} \tag{2.1}$$

where θ is an unknown parameter of interest and ν_{it} is random. The goal is to carry out inference on the parameter θ using the sample mean as estimator (which is the OLS estimator) without making strong assumptions on ν_{it}. Chapter 2 will consider the more general case where regressors varying over i and t will be included in the model. It is common to first analyze the properties of a bootstrap method for the sample mean before investigating more complicated statistics.

It will prove convenient to represent our panel data as a matrix. We will do so by putting into rows the observations for each cross-sectional unit and then stacking these rows. The resulting matrix, which we denote by Y, is of dimension $N \times T$:

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

We will analyze the properties of bootstrap methods in *(2.1)* that do not require making strong assumptions during implementation. We will do so under various scenarios on the properties of v_{it}. We conjecture that our results will extend to even more general structures, and this will be the subject of future research.

We decompose ν_{it} into four components :

$$\nu_{it} = \mu_i + f_t + \lambda_i F_t + \varepsilon_{it}. \tag{2.2}$$

It is customary to call μ_i the individual effect and f_t the time effect. The term $\lambda_i F_t$ represents the contribution from a factor model. In that model, each unit i is allowed to respond heterogeneously to a set of common factors F_t. Finally, the last term is the remainder and will be called the idiosyncratic component.

Assumption A (individual effects)

The individual effects μ_i are drawn independently across i from some distribution with mean 0 and variance σ_μ^2 (where $0 < \sigma_\mu^2 < \infty$.) and is independent of the cross-sectional and/or temporal heterogeneity.

Assumption A requires the individual heterogeneities to be independent and identically distributed with finite variance. The assumption of a zero mean is an identification assumption as any non-zero mean could be subsumed into the overall mean θ. The *i.i.d.* assumption, strong for classical asymptotic distribution is however important for bootstrap validity because i.i.d. bootstrap will be used in the cross-sectional dimension.

Assumption B (time effects)

$\{f_t\}$ is a stationary and α-mixing process with mixing coefficients $\alpha(j)$; $E(f_t) = 0$ and $\{f_t\}$ verifies Ibragimov's assumptions, that is $\exists\ \delta \in (0,\infty)$ such that $E|f_t|^{2+\delta} < \infty$ and $\sum_{j=1}^{\infty} \alpha(j)^{\delta/(2+\delta)} < \infty$ with finite long-run variance $V_f^\infty = \sum_{h=-\infty}^{\infty} Cov(f_t, f_{t+h}) \in (0,\infty)$.

Assumption B imposes some conditions on the time-series heterogeneity of our panel data. In particular, it requires it to be generated from a stationary process and that the dependence between f_t and f_k vanishes sufficiently fast as the distance between them increases.

Assumptions C (idiosyncratic error)

C : The idiosyncratic error ε_{it} is drawn independently across i and over t from some distribution with mean 0 and variance σ_ε^2 where $0 < \sigma_\varepsilon^2 < \infty$.

C' : The idiosyncratic error ε_{it} satisfies the following condition : the scaled sample mean $\sqrt{M}\bar{\bar{\varepsilon}}$ (with $M \in \{N,T\}$) converges to zero in probability.

Assumption C requires that the idiosyncratic error is to *i.i.d.* in both dimensions. The assumption is strong and will give us asymptotic distribution when only one dimension goes to infinity. When N and T go to infinity, we will use a weaker version of assumption C, assumption C' that allows for weak dependence such as spatial, etc...

Assumption D (independence)

The two processes $(\mu_1, .., \mu_N)$ and $(f_1, ..., f_T)$ are independent.

Assumption D imposes independence between the vector of individual heterogeneities and the vector of temporal heterogeneities. It is essential that there is no dependence between the two types of heterogeneity because the double resampling bootstrap method we will present later would destroy any dependence between the two dimensions.

Assumptions E (factor)

E1 : The factor loadings λ_i are drawn independently across i from some distribution with mean 0 and variance σ_λ^2 where $0 < \sigma_\lambda^2 < \infty$.

E2 : The factors (F_t) are a stationary and α-mixing process with mean 0 satisfying Ibragimov's assumptions.

E3 : The two processes $(\lambda_1, .., \lambda_N)$ and $(F_1, ..., F_T)$ are independent.

Assumptions E are about a factor model. Assumption E1 requires the loadings in a factor model to be independent and identically distributed with finite variance. Assumption E2 is similar to assumption B, but applied to the factors in a factor model. Assumption E3 imposes independence between the vector of loadings and the vector of factors in an factor model. The reason is similar to B.

This general decomposition nests most popular panel data models. Making assumptions on the properties of each of these components defines particular panel data models : the cross-sectional one-way error component model

(ECM), the temporal one-way ECM, the two-way ECM and the Factor Model.

Cross-sectional one-way ECM

$$y_{it} = \theta + \mu_i + \varepsilon_{it} \qquad (2.3)$$

under the assumptions A and C (C'). This model captures a single source of heterogeneity, that is systematic differences across units that results in a parallel shift. It is important to emphasize that we teat this heterogeneity as nuisance and not as parameters to be estimated. The parameter of interest is θ. To consider the properties of our bootstrap schemes under this model, we will assume a random parameter model. In other words, the individual effects μ_i will be assumed to be drawn from some distribution.

Temporal one-way ECM

A second special case of our general model (*2.1*) is the temporal one-way ECM :

$$y_{it} = \theta + f_t + \varepsilon_{it}. \qquad (2.4)$$

under the assumptions B and C (C'). In contrast to the cross-sectional ECM model discussed above, the only heterogeneity considered is with respect to the time periods. This model is obviously much less common than the cross-section ECM, but we present it for completeness. Assumption B is somewhat different from Assumption A because we want to allow for some serial correlation in the time-specific effects f_t.

Two-way ECM

The two-way error component model allows to control for both cross-sectional and temporal heterogeneity. It is a combination of both one-way ECMs discussed above :

$$y_{it} = \theta + \mu_i + f_t + \varepsilon_{it} \qquad (2.5)$$

As in both one-way ECMs, cross section and temporal heterogeneities will be treated as nuisance random variables. Since it is a combination of the preceding two models. (2.5) is defined under the assumptions A, B, C' and D.

Factor Model

While the two-way ECM assumes that all cross-sectional units respond homogeneously to time variation, the factor model allows this response to be heterogeneous across units. These factor models have become highly popular in panel data either to summarize a large amount of information that can be used later (for example for forecasting, see Stock and Watson, 2002) or to model cross-sectional dependence in large panels (for example in finance or for panel unit root tests as in Bai and Ng (2004), Moon and Perron (2004) or Phillips and Sul (2003)).

The factor model we will study is :

$$y_{it} = \theta + \mu_i + \lambda_i F_t + \varepsilon_{it} \qquad (2.6)$$

The model is a single-factor model because only one factor process F_t is involved in the specification. The parameters $\lambda_1, ..., \lambda_N$ are called the factor loadings and represent the sensitivity of unit i to changes in the factor. The Model (2.6) is under the assumptions A, C'and E.

1.3 Asymptotic Theory

This section presents theoretical results on the asymptotic distribution of the sample mean.

One difficulty with asymptotic theory for panel data is the assumption made on the size of N and T. Traditionally, because panel data was mostly used in microeconometrics with large cross-sectional dimension but short time dimension, the assumption was made that N was large (approaching infinity) but that T remained finite. Conversely, in multiple time series models, the asymptotic analysis typically assumes that the number of series N is small while the number of time series observations T is large. Of course, these two asymptotic scenarios lead to different approximations and one is left to wonder which one is most appropriate for a given application at hand.

Recently. the analysis of large macro-type panels where both dimensions are reasonably large has allowed both dimensions to diverge. Phillips and Moon (1999) have provided underpinnings for these asymptotic analyses and have defined different frameworks. A sequential limit is obtained when an index is fixed at first, and the other goes to infinity, to have intermediate result. Next, the final result is obtained by allowing the fixed index to go to infinity. On the other hand, in a diagonal path limit*,* N and T approach to infinity along a specific path, for example $T = T(N)$ and $N \to \infty$. Finally, in a joint limit*,* N and T pass to infinity simultaneously. Sometimes, it is necessary to control the relative expansion rate of N and T. For equivalence conditions between sequential and joint limits, see Phillips and Moon (1999).

Again, in practice, when faced with a particular application, it is not always obvious how to choose among these multiple asymptotic distributions, which may very be different. One of the advantages of the bootstrap approach we are analyzing is that it avoids having to choose between these competing approximations.

In order to prove the validity of the bootstrap for inference about θ,we need to show that it reproduces the asymptotic distribution of the estimator $\overline{\overline{y}}$. The purpose of this section is to develop the asymptotic distribution of $\overline{\overline{y}}$ in the various panel data models described in section 2 and under various scenarios on N and T. Then, the next section will show that the bootstrap will (or will not) reproduce these asymptotic distributions.

The asymptotic analysis is carried out by noting that, using (*2.1*) and (*2.2*), the sample mean can be written as :

$$\begin{aligned}
\overline{\overline{y}} &= \theta + \frac{1}{N}\sum_{i=1}^{N}\mu_i + \frac{1}{T}\sum_{t=1}^{T} f_t + \left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i\right)\left(\frac{1}{T}\sum_{t=1}^{T}F_t\right) && (3.1)\\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it} \\
&= \theta + \overline{\mu} + \overline{f} + \overline{\lambda}.\overline{F} + \overline{\overline{\varepsilon}} && (3.2)
\end{aligned}$$

The asymptotic behavior of the sample mean will thus depend on the behavior of these 4 sample means. It is important to mention that they do not converge at the same rate. For example, $\overline{\mu}$ and $\overline{\lambda}$ $\left(\text{resp. } \overline{f} \text{ and } \overline{F}\right)$ are averages of N (resp. T) elements when $\overline{\overline{\varepsilon}}$ is the average of NT elements. This difference of convergence rates among elements implies that some elements become negligible more rapidly when the sample size increases than others.

Two asymptotic theories are available for the cross-sectional and temporal one-way ECM. In the case of the two-way ECM, N and T must go to infinity. The relative convergence rate between the two indexes, δ defines a continuum of asymptotic distributions. The factor model has a unique asymptotic distribution when the two dimensions go to infinity. The spatial dependence possibly contained in $\{\varepsilon_{it}\}$ for all the specifications, or in $\{\lambda_i F_t\}$ for factor model, vanishes when N and T go to infinity.

Table 1. Asymptotic distributions of the sample mean

<table>
<tr><th>Model</th><th>Assum.</th><th>Asymptotic distribution</th><th>ω</th></tr>
<tr><td>$Cross-sect.$</td><td>$A,\, C$</td><td>$\sqrt{N}\left(\overline{\overline{y}}-\theta\right) \underset{N\to\infty}{\Longrightarrow} N\left(0,\omega\right)$</td><td>$\sigma_\mu^2+\frac{\sigma_\varepsilon^2}{T}$</td></tr>
<tr><td>$One-way$
ECM</td><td>A, C'</td><td>$\sqrt{N}\left(\overline{\overline{y}}-\theta\right) \underset{N,T\to\infty}{\Longrightarrow} N\left(0,\omega\right)$</td><td>$\sigma_\mu^2$</td></tr>
<tr><td>$Temporal$</td><td>$B,\, C$</td><td>$\sqrt{T}\left(\overline{\overline{y}}-\theta\right) \underset{T\to\infty}{\Longrightarrow} N\left(0,\omega\right)$</td><td>$V_f^\infty+\frac{\sigma_\varepsilon^2}{N}$</td></tr>
<tr><td>$One-way$
ECM</td><td>B, C'</td><td>$\sqrt{T}\left(\overline{\overline{y}}-\theta\right) \underset{N,T\to\infty}{\Longrightarrow} N\left(0,\omega\right)$</td><td>$V_f^\infty$</td></tr>
<tr><td rowspan="2">$Two-way$
ECM</td><td rowspan="2">A, B, C', D</td><td>$\sqrt{N}\left(\overline{\overline{y}}-\theta\right) \underset{\substack{N,T\to\infty \\ \frac{N}{T}\to\delta\in[0,\infty)}}{\Longrightarrow} N\left(0,\omega\right)$</td><td>$\sigma_\mu^2+\delta.V_f^\infty$</td></tr>
<tr><td>$\sqrt{T}\left(\overline{\overline{y}}-\theta\right) \underset{\substack{N,T\to\infty \\ \frac{N}{T}\to\infty}}{\Longrightarrow} N\left(0,\omega\right)$</td><td>$V_f^\infty$</td></tr>
<tr><td>$Factor$
$model$</td><td>A, C', E</td><td>$\sqrt{N}\left(\overline{\overline{y}}-\theta\right) \underset{N,T\to\infty}{\Longrightarrow} N\left(0,\omega\right)$</td><td>$\sigma_\mu^2$</td></tr>
</table>

1.4 Resampling Methods

In this section, we present the methods we use to resample panel data. These methods have in common that they resample the observed data y_{it}. In other words, their implementation do not depend on the choice of a particular structure for the data. However, of course the validity of each method will depend on the properties of the underlying data-generating process. In other words, we want to make inference that is robust to the panel data models described in the previous section without having to impose that model in resampling.

From our initial $N \times T$ data matrix Y, bootstrapping will create a new matrix Y^* by resampling with replacement elements of Y. Statistics are computed on this pseudo-sample, and we repeat this operation B times. We use the sequence of B statistics generated by the bootstrap to make inference about the parameter θ.

A word on notation before presenting the resampling methods. Bootstrap quantities will be denoted by an asterisk. The probability measure induced by the resampling method conditional on Y is noted P^*. $E^*()$ and $Var^*()$ are respectively the expectation and the variance associated with P^*.

Cross-sectional Resampling Bootstrap

For a $N \times T$ matrix Y, *cross-sectional resampling* constructs a new $N \times T$ matrix Y^* with rows obtained by resampling with replacement the rows of Y. In other words, we resample the vectors of T observations for each individual. As a consequence, conditionally on Y, the rows of Y^* are independent and identically distributed. y_{it}^* can only take one of the N values $\{y_{it}\}_{i=1,\ldots N}$, those that were observed for some individuals at time t . Y^* takes the following form :

$$\underset{(N,T)}{Y^*} = \begin{pmatrix} y^*_{11} = y_{i_1 1} & y^*_{12} = y_{i_1 2} & ... & y^*_{1T} = y_{i_1 T} \\ y^*_{21} = y_{i_2 1} & y^*_{22} = y_{i_2 2} & ... & y^*_{2T} = y_{i_2 T} \\ ... & ... & .. & ... \\ y^*_{N1} = y_{i_N 1} & y^*_{N2} = y_{i_N 2} & ... & y^*_{NT} = y_{i_N T} \end{pmatrix} \qquad (4.1)$$

where each of the indices $(i_1, i_2, ..., i_N)$ is obtained by *i.i.d.* drawing with replacement from $(1, 2, ..., N)$. The mean of Y^* obtained by the cross-sectional bootstrap is denoted by $\overline{\overline{y}}^*_{cros}$.

Block Resampling Bootstrap

This method is a direct generalization 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 the time dimension blocks of consecutive periods in order to capture temporal dependence. All the observations at each time period are kept together in the hope of preserving their dependence.

The block bootstrap resampling constructs a new $N \times T$ matrix Y^* with columns obtained by resampling with replacement blocks of columns of Y. As a consequence, in this method, y^*_{it} can only take one of the T values

$\{y_{it}\}_{t=1,...T}$, those that were observed for individual i at some time t. The mean of Y^* obtained by block bootstrap method is noted $\overline{\overline{y}}^*_{bl}$. Y^* takes the following form :

$$\underset{(N,T)}{Y^*} = \begin{pmatrix} y^*_{11} = y_{1t_1} & y^*_{12} = y_{1t_2} & ... & y^*_{1T} = y_{1t_T} \\ y^*_{21} = y_{2t_1} & y^*_{22} = y_{2t_2} & ... & y^*_{2T} = y_{2t_T} \\ ... & ... & .. & ... \\ y^*_{N1} = y_{Nt_1} & y^*_{N2} = y_{Nt_2} & ... & y^*_{NT} = y_{Nt_T} \end{pmatrix} \qquad (4.2)$$

The choice of $(t_1, t_2, .., t_T)$ depends on the which block bootstrap method is used in the time dimension. With the CBB bootstrap resampling, we have $(t_1, t_2, ., t_T)$ taking the form

$$\underbrace{\tau_1, \tau_1+1, .., \tau_1+l-1}_{block\ 1}\ \underbrace{\tau_2, \tau_2+1, .., \tau_2+l-1}_{block\ 2}, ..., \underbrace{\tau_{[T/l]}, \tau_{[T/l]}+1, .., \tau_{[T/l]}+l-1}_{block\ [T/l]}$$

where the vector of indices $\left(\tau_1, \tau_2, ..., \tau_{[T/l]}\right)$ is obtained by $i.i.d.$ drawing with replacement from $(1, 2, ..., T)$, l denoting the block length[1]. Conditionally to Y, the blocks are $i.i.d.$ and the properties of the original i.i.d. bootstrap are transferred to the blocks as statistical units. With the CBB there are T possible overlapping blocks of length beginning with each periods from $t = 1....$ to $t = T$

$$\underbrace{1, 2, .., l}_{block\ 1}\ \underbrace{2, 3, .., l+1}_{block\ 2}\ \underbrace{3, 4, .., l+2}_{block\ 2}, ...\underbrace{k, k+1, .., k+l-1}_{block\ k}, ..., \underbrace{T, 1, .., l-1}_{block\ T}.$$

The CBB resampling on matrix Y is $i.i.d.$ drawing with replacement of $K = [T/l]$ blocks from the T possible blocks. Let 's define a new matrix Z, a transformation of the sample Y.

[1] The name Circular come from the fact that when $\tau_t > T - l$, the index of some observations exceed T and are replace using the rule : $T + t \longleftrightarrow t$, as if the original data are around a circle and after T we continue with the first observation $t = 1$.

$$\underset{(N,T)}{Z} = \begin{pmatrix} z_{11} & z_{12} & \dots & \dots & \dots & z_{1T} \\ \dots & \dots & .. & \dots & \dots & \dots \\ \dots & \dots & .. & z_{ik} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{N1} & z_{N2} & \dots & \dots & \dots & z_{NT} \end{pmatrix} \qquad (4.3)$$

where $z_{ik} = \frac{1}{l} \sum_{t \in \ block \ k} y_{it}$ is for given unit i, the average of the observations of block k. The matrix Z will be useful to derive some theoretical result in the next section.

Others block bootstrap methods can also be accommodated in the time dimension to panel data. In this chapter the theoretical results will be given for the CBB. They remain valid for the NMB and the MMB because the three methods are asymptotically equivalent.

Double Resampling Bootstrap

This method is a combination of the two previous resampling methods. 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 Y, an intermediate matrix Y^* is obtained either by cross-sectional resampling or block resampling. It turns out that the resampling is symmetric so it does not matter which dimension is resampled first. Then, another resampling is made in the second dimension : from Y^* the final matrix Y^{**} is obtained. If we resampled in the cross-sectional dimension in the first step, then we resampled columns of the intermediate matrix in order to get our resampled matrix Y^{**}.[2] The mean of Y^{**} is noted $\overline{\overline{y}}^{**}$.

Carvajal (2000) and Kapetanios (2008) have both suggested this double resampling in the special case where the block length is 1. They also analyze this resampling method by Monte Carlo simulations but give no theoretical support. The idea is that by drawing in one dimension, we preserve the dependence in that dimension in the first step. In the second step, we reproduce the properties in the other dimension by preserving the vectors drawn in the first step. Y^{**} takes the following form :

$$\underset{(N,T)}{Y^{**}} = \begin{pmatrix} y_{11}^{**} = y_{i_1 t_1} & y_{12}^{**} = y_{i_1 t_2} & \cdots & y_{1T}^{**} = y_{i_1 t_T} \\ y_{21}^{**} = y_{i_2 t_1} & y_{22}^{**} = y_{i_2 t_2} & \cdots & y_{2T}^{**} = y_{i_2 t_T} \\ \cdots & \cdots & .. & \cdots \\ y_{N1}^{**} = y_{i_N t_1} & y_{N2}^{**} = y_{i_N t_2} & \cdots & y_{NT}^{**} = y_{i_N t_T} \end{pmatrix} \qquad (4.4)$$

where the indices $(i_1, i_2, ..., i_N)$ and $(t_1, t_2, ., t_T)$ are chosen as described in the the two previous sub-sections. One important aspect of our analysis of double resampling is the properties of y_{it}^{**}. Conditionally on the matrix $[Y]$, the elements of $[Y^{**}]$ have a particular dependence structure. In fact each element y_{it}^{**} depends on the elements in its column and on its row. This link exists because elements on the same line belong to the same unit i and elements in the same column refer to the same period t. This structure of dependence and the validity of the bootstrap methods will be analyzed in the next section.

[2] We will use double asterisks** denote the quantities induced by double resampling.

1.5 Bootstrap Validity

This section analyzes the properties of the bootstrap methods described in the previous section in the panel models described in section 2. A bootstrap method is *consistent* for the sample mean if the distance between the bootstrap distribution function and the sampling distribution of the statistic converges to 0 asymptotically. Since we have different (three) modes of convergence, we have three definitions of consistency. In order to avoid overburdening the text, we will denote by $\underset{NT\to\infty}{\overset{P}{\to}}$ the convergence in probability under either case of asymptotic analysis : N fixed with T going to infinity, T fixed with N going to infinity, and finally N and T going to infinity simultaneously. With this notation, we will say that the bootstrap is consistent if :

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\sqrt{M}\left(\overline{\overline{y}}^* - \overline{\overline{y}}\right) \leq x\right) - P\left(\sqrt{M}\left(\overline{\overline{y}} - \theta\right) \leq x\right)\right| \underset{NT\to\infty}{\overset{P}{\to}} 0 \qquad (5.1)$$

with $M \in \{N, T, NT\}$.

M is the scaling factor and depends on the panel model specification. In the special case where the sample mean asymptotic distribution is available, consistency can be established by showing that the bootstrap sample mean has the same distribution. The next proposition expresses this idea.

Proposition 1 *: Assume that* $\sqrt{M}\left(\overline{\overline{y}} - \theta\right) \Longrightarrow L$ *and* $\sqrt{M}\left(\overline{\overline{y}}^* - \overline{\overline{y}}\right) \overset{*}{\Longrightarrow} L^*$*. If* L^* *and* L *are identical and continuous, then*

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\sqrt{M}\left(\overline{\overline{y}}^* - \overline{\overline{y}}\right) \leq x\right) - P\left(\sqrt{M}\left(\overline{\overline{y}} - \theta\right) \leq x\right)\right| \underset{NT\to\infty}{\overset{P}{\to}} 0$$

where "$\overset{*}{\Longrightarrow}$*" means "converge in distribution conditionally on* Y*".*

To understand the behavior of the resampling schemes, it is convenient to decompose the error term. Using the matrix notation developed above, we can rewrite the data matrix Y as

$$
\begin{aligned}
Y &= \underbrace{\theta \iota_N \iota_T'}_{[\theta]} + \underbrace{\begin{pmatrix} \mu_1 & ... & \mu_1 \\ \mu_2 & ... & \mu_2 \\ ... & .. & ... \\ \mu_N & ... & \mu_N \end{pmatrix}}_{[\mu]} + \underbrace{\begin{pmatrix} f_1 & ... & f_T \\ f_1 & ... & f_T \\ ... & ... & ... \\ f_1 & ... & f_T \end{pmatrix}}_{[f]} + \underbrace{\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ ... \\ \lambda_N \end{pmatrix}}_{[\lambda]} \underbrace{\begin{pmatrix} F_1 & ... & F_T \end{pmatrix}}_{[F]} \\
&\quad + \underbrace{\begin{pmatrix} \varepsilon_{11} & ... & \varepsilon_{1T} \\ \varepsilon_{21} & ... & \varepsilon_{2T} \\ ... & .. & ... \\ \varepsilon_{N1} & ... & \varepsilon_{NT} \end{pmatrix}}_{[\varepsilon]} \\
&= [\theta] + [\mu] + [f] + [\lambda]\,[F] + [\varepsilon]
\end{aligned} \tag{5.2}
$$

Thus, each line of the matrix $[\mu]$ contains T times the same value. Thus, if one were to resample $[\mu]$ in the cross-sectional dimension (i.e. drawing rows) and take the overall average would be equivalent to an i.i.d. resampling of μ_i. Similarly, cross-sectional resampling is also equivalent to i.i.d. resampling of the factor loadings λ_i.

On the other hand, the rows of the matrices $[f]$ and $[F]$ are identical. This means that cross-sectional resampling does not do anything and returns the original matrices $[f]$ and $[F]$. Effectively, it treats $(f_1, ..., f_T)$ and $(F_1, ..., F_T)$ as constants. In other words, when doing cross-sectional resampling, each bootstrap observation can be decomposed as :

$$
y^*_{it,cros} = \theta + \mu^*_i + f_t + \lambda^*_i F_t + \varepsilon^*_{it,cros} \tag{5.3}
$$

where μ_i^*, λ_i^* are i.i.d. draws.

The analysis for temporal block resampling is symmetrical. It is equivalent to block resampling on the time effects $f_1, .., f_T$ and the factors $F_1, ..., F_T$. However, it does not resample the individual effects and factor loadings and teats them as constants. Hence, each bootstrap observation can be written as

$$y_{it,bl}^* = \theta + \mu_i + f_{t,bl}^* + \lambda_i F_{t,bl}^* + \varepsilon_{it,bl}^*. \tag{5.4}$$

Finally, double resampling is the combination of the two previous methods. It is equivalent to the combination of i.i.d. resampling on the individual effects $(\mu_1, ..., \mu_N)$ and factor loadings $(\lambda_1, ..., \lambda_N)$ and block resampling on the time effects $(f_1, ..., f_T)$ and factors $(F_1, ..., F_T)$.

$$y_{it}^{**} = \theta + \mu_i^* + f_{t,bl}^* + \lambda_i^* F_{t,bl}^* + \varepsilon_{it}^{**}. \tag{5.5}$$

Using the above expression, we can express the bootstrap means as :

$$\begin{aligned}
\left(\overline{\overline{y}}_{cros}^* - \overline{\overline{y}}\right) &= \left(\overline{\mu}^* - \overline{\mu}\right) + \left(\overline{\lambda}^* \overline{F} - \overline{\lambda F}\right) + \left([\overline{\varepsilon}_{inter}]^* - \overline{\overline{\varepsilon}}\right) && (5.6)\\
\left(\overline{\overline{y}}_{bl}^* - \overline{\overline{y}}\right) &= \left(\overline{f}_{bl}^* - \overline{f}\right) + \left(\overline{\lambda F}_{bl}^* - \overline{\lambda F}\right) + \left([\overline{\varepsilon}_{inter}]_{bl}^* - \overline{\overline{\varepsilon}}\right) && (5.7)\\
\left(\overline{\overline{y}}^{**} - \overline{\overline{y}}\right) &= \left(\overline{\mu}^* - \overline{\mu}\right) + \left(\overline{f}_{bl}^* - \overline{f}\right) + \left(\overline{\lambda}^* \overline{F}_{bl}^* - \overline{\lambda F}\right) + \left(\overline{\overline{\varepsilon}}^{**} - \overline{\overline{\varepsilon}}\right) && (5.8)
\end{aligned}$$

It must be noted that the centering eliminates f_t in the case of cross-sectional resampling and μ_i in the case of temporal block resampling. It follows immediately that cross-sectional resampling is inconsistent in the presence of temporal heterogeneity as it cannot reproduce it. Similarly, the temporal block resampling is inconsistent in presence of cross-sectional heterogeneity.

The particular dependence structure in Y^{**}, induces a particular form for the bootstrap variance of $\overline{\overline{y}}^{**}$ as expressed by the following proposition.

Proposition 2 *: ∀ N, T, the double resampling bootstrap-variance is :*

$$
\begin{aligned}
Var^{**}\left(\overline{\overline{y}}^{**}\right) &= Var^{*}\left(\overline{\overline{z}}^{*}\right) + \left(1-\frac{1}{K}\right) Var^{*}\left(\overline{\overline{y}}^{*}_{cros}\right) + \left(1-\frac{1}{N}\right) Var^{*}\left(\overline{\overline{y}}^{*}_{bl}\right) \\
Var^{**}\left(\overline{\overline{y}}^{**}\right) &> \left(1-\frac{1}{K}\right) Var^{*}\left(\overline{\overline{y}}^{*}_{cros}\right) \qquad (5.9) \\
Var^{**}\left(\overline{\overline{y}}^{**}\right) &> \left(1-\frac{1}{N}\right) Var^{*}\left(\overline{\overline{y}}^{*}_{bl}\right)
\end{aligned}
$$

(5.9) gives the expression of the double resampling bootstrap mean variance. It is important to mention that these results are finite sample properties, holding without any assumption about y_{it}. The first term $Var^{*}\left(\overline{\overline{z}}^{*}\right)$ is the i.i.d. bootstrap mean variance for transformed data z_{ik} where in the time dimension, we make the average of observations by block as described in *(4.5)*. The second component of $Var^{**}\left(\overline{\overline{y}}^{**}\right)$ is the cross-sectional resampling bootstrap variance times $\left(1-\frac{1}{K}\right)$ and the third component is the block resampling bootstrap variance times $\left(1-\frac{1}{N}\right)$. The two inequalities mean that the double resampling bootstrap induces a greater variance than the cross-sectional resampling bootstrap and the block resampling bootstrap. That implies that in some cases the cross-sectional resampling bootstrap or the block resampling bootstrap could reject the null hypothesis while the double resampling bootstrap does not reject it. Inversely, if the double resampling bootstrap rejects the null hypothesis, there is no chance that a bootstrap method in one dimension does not reject it.

Another implication of *(5.9)* what happens in the particular case when the block length $l = 1$. (5.9) becomes :

$$
\begin{aligned}
Var^{**}\left(\overline{\overline{y}}^{**}\right) &= Var^{*}\left(\overline{\overline{y}}^{*}\right) + \left(1-\frac{1}{T}\right) \frac{Var^{*}\left(\overline{y}^{*}_{i.}\right)}{N} + \left(1-\frac{1}{N}\right) \frac{Var^{*}\left(\overline{y}^{*}_{.t}\right)}{T} \\
Var^{**}\left(\overline{\overline{y}}^{**}\right) &\geq Var^{*}\left(\overline{\overline{y}}^{*}\right) \qquad (5.10)
\end{aligned}
$$

It is important to mention two things about inequality *(5.10).* First, the equality $Var^{**}\left(\overline{\overline{y}}^{**}\right) = Var^{*}\left(\overline{\overline{y}}^{*}\right)$ holds in *(5.10)* when $T = 1$ (cross-section data) or $N = 1$ (time series). Second, *(5.10)* means that in finite sample, the double resampling bootstrap induces a greater variance than the i.i.d. bootstrap. In particular the next proposition expresses what happens asymptotically when the double resampling bootstrap is applied to *i.i.d.* error term ε_{it}.

Proposition 3 *: Under Assumption C, using the double resampling bootstrap with block length l=1 we have :*

$$Var^{**}\left(\sqrt{NT}\;\overline{\overline{\varepsilon}}^{**}\right) \underset{N,T\rightarrow\infty}{\overset{P}{\rightarrow}} 3.\sigma_{\varepsilon}^{2} \qquad (5.11)$$

In the absence of random heterogeneities, the double resampling induces a bootstrap-variance three times larger than *i.i.d.* bootstrap inducing a conservative confidence interval.

Let's introduce new assumptions about the error term ε_{it}.

Assumptions C" (idiosyncratic error)

C"1 **:** the scaled sample mean $\sqrt{M}\overline{\overline{\varepsilon}}$ (with $M \in \{N, T\}$) converges in probability to zero.

C"2 : the empirical mean of squares of cross-section averages $\frac{1}{N}\sum_{i}\left(\overline{\varepsilon}_{i.}\right)^{2}$ converges in probability to zero.

C"3 : the empirical mean of squares of temporal block averages $\frac{1}{[T/l]}\sum_{k}\left(\overline{e}_{.k}\right)^{2}$ converges in probability to zero.

Assumption C" is a weaker version of assumption C. The first assumption ensures that the error term ε_{it} is asymptotically negligible : it is the assumption C'). Assumption C"2 ensures that there is no cross-sectional heterogeneity remaining in ε_{it} . Assumption C"3 excludes temporal heterogeneity in ε_{it}. $[T/l]$ denotes the number of blocks in the time dimension and $\overline{e}_{k.}$ the average of the term ε_{it} in the block k for all the individuals as exposed in *(4.3).*

Under assumption *C"* about the error term ε_{it} the next proposition analyzes the asymptotic behavior of $\overline{\overline{\varepsilon}}^{**}$when the scaling factor is $\sqrt{N}$ or $\sqrt{T}$.

Proposition 4 *Under assumption C", using the double resampling bootstrap we have :*

$$Var^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right) \underset{N,T\to\infty}{\overset{P}{\to}} 0 \text{ and } Var^{**}\left(\sqrt{T}\overline{\overline{\varepsilon}}^{**}\right) \underset{N,T\to\infty}{\overset{P}{\to}} 0 \qquad (5.12)$$

The implication of Proposition 4 is that when the scaling factor of $\overline{\overline{y}}^{**}$ is $\sqrt{N}$ or $\sqrt{T}$ (presence of heterogeneities) the distribution of $\overline{\overline{\varepsilon}}^{**}$ does not appear in the asymptotic distribution of $\overline{\overline{y}}^{**}$. That means that the double resampling bootstrap method is valid under general spatial dependence. Thus validity of the bootstrap method will be focused of the components $[\mu]$, $[f]$, $[\lambda]$ and $[F]$.

Our validity proofs will imitate the procedure in *Proposition 1.* For each bootstrap method, by deducing the asymptotic distribution of the components of $\overline{\overline{y}}^{*} - \overline{\overline{y}}$, using the appropriate scaling factor and comparing with the asymptotic distributions in Table 1, one can identify consistent and inconsistent bootstrap for the different panel model specifications. The results are in the following proposition.

Proposition 5 *: Consistency.*

1 - In the presence of temporal heterogeneity, the cross-sectional bootstrap is inconsistent.

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\sqrt{M}\left(\overline{\overline{y}}^*_{cros}-\overline{\overline{y}}\right)\leq x\right)-P\left(\sqrt{M}\left(\overline{\overline{y}}-\theta\right)\leq x\right)\right|\underset{NT\to\infty}{\overset{P}{\nrightarrow}}0$$

with $M\in\{N,T,NT\}$.

2 - In the presence of cross-sectional heterogeneity, the block bootstrap methods are inconsistent.

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\sqrt{M}\left(\overline{\overline{y}}^*_{bl}-\overline{\overline{y}}\right)\leq x\right)-P\left(\sqrt{M}\left(\overline{\overline{y}}-\theta\right)\leq x\right)\right|\underset{NT\to\infty}{\overset{P}{\nrightarrow}}0$$

with $M\in\{N,T,NT\}$.

3 - In the presence of cross-sectional and/or temporal heterogeneity, under the assumption that $l^{-1}+lT^{-1}=o(1)$ *as* $T\to\infty$*, the double resampling bootstrap is consistent when* N *and* T *go to infinity*

$$\sup_{x\in\mathbb{R}}\left|P^{**}\left(\sqrt{M}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right)\leq x\right)-P\left(\sqrt{M}\left(\overline{\overline{y}}-\theta\right)\leq x\right)\right|\underset{NT\to\infty}{\overset{P}{\to}}0$$

with $M\in\{N,T\}$.

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. The strength of the double resampling is to replicate the behavior of the main components of errors terms, without having to separate them. It is thus robust to the presence of these two types of heterogeneity and will allow for valid inference without having to make parametric assumptions. The consistency of the bootstrap methods for the different panel model specifications are presented in Table 2.

Table 2 : Summary of bootstrap consistency

	Cross-sect. Resampling	Block Resampling	Double Resampling
Cross. one-way ECM	Consistent		Consistent
Temp. one-way ECM		Consistent	Consistent
Two-way ECM			Consistent
Factor model	Consistent		Consistent

1.6 Bootstrap Confidence Interval

Once we have used the bootstrap to generate B pseudo samples, we can construct confidence intervals for θ. In the literature, there are several bootstrap confidence intervals. The percentile confidence interval and the percentile-t confidence intervals are the commonly used.

Bootstrap Percentile Interval

The first type of interval is based on the distribution of the bootstrap mean. For each pseudo-sample Y_b^*, we compute the bootstrap-sample mean : $\overline{\overline{y}}_b^*$ and the centered statistic $r_b^* = \overline{\overline{y}}_b^* - \overline{\overline{y}}$. The empirical distribution of these B realizations is :

$$\widehat{R}^*(x) = \frac{1}{B}\sum_{b=1}^{B} I\left(r_b^* \leq x\right) \tag{6.1}$$

$\widehat{R}^*$is an approximation of the cumulative distribution function of the bootstrap-mean . The *percentile* confidence interval of level $(1-\alpha)$ for the parameter θ is then constructed as

$$CI_{1-\alpha}^* = \left[\overline{\overline{y}} - r_{1-\alpha/2}^*; \overline{\overline{y}} - r_{\alpha/2}^*\right] \tag{6.2}$$

where $r_{\alpha/2}^*$ and $r_{1-\alpha/2}^*$ are respectively the $\alpha/2$-percentile and $(1-\alpha/2)$-percentile of $\widehat{R}^*$. B should be chosen so that $\alpha(B+1)$ is an integer. When $\widehat{R}^*(x)$ is symmetric, $r_{\alpha/2}^* = -r_{1-\alpha/2}^*$ and a symmetric percentile interval is :

$$CI_{1-\alpha}^* = \left[C_{\alpha/2}^*; C_{1-\alpha/2}^*\right] \tag{6.3}$$

where $C_{\alpha/2}^*$ and $C_{1-\alpha/2}^*$ are respectively the $\alpha/2$-percentile and the $(1-\alpha/2)$-percentile of the empirical distribution function of $\left\{\overline{\overline{y}}_b^*\right\}_{b=1..B+1}$. This is a simple way of constructing a non-parametric confidence interval.

Bootstrap Percentile-t Interval

Alternatively, one could build a percentile-t interval. These are often preferred because they involve bootstrapping pivotal statistics (statistics that do not depend on nuisance parameters) and sometimes allow proving asymptotic refinements (though we will not prove any such refinement in this thesis).

To construct this type of intervals, we compute the t statistic on each pseudo-sample Y_b^* :

$$t_b^* = \frac{\overline{\overline{y}}_b^* - \overline{\overline{y}}}{\sqrt{\widehat{V}ar^*\left(\overline{\overline{y}}^*\right)}} \tag{6.4}$$

The empirical distribution of these B realizations is

$$\widehat{G}^{*}(x)=\frac{1}{B}\sum_{b=1}^{B}I\left(t_{b}^{*}\leq x\right) \tag{6.5}$$

A percentile-t confidence interval of level $(1-\alpha)$ is

$$CI_{1-\alpha}^{*}=\left[\overline{\overline{y}}-\sqrt{\widehat{V}ar\left(\overline{\overline{y}}\right)}.t_{1-\frac{\alpha}{2}}^{*};\overline{\overline{y}}-\sqrt{\widehat{V}ar\left(\overline{\overline{y}}\right)}.t_{\frac{\alpha}{2}}^{*}\right] \tag{6.6}$$

where $t_{\alpha/2}^{*}$ and $t_{1-\alpha/2}^{*}$ are respectively the $\alpha/2$-percentile and $(1-\alpha/2)$-percentile of $\widehat{G}^{*}$. The construction of these intervals resembles standard Wald-type statistics where one adds and subtracts a given quantile from the normal distribution (for example 1.96 for a 95% interval). The bootstrap is only used to compute the appropriate multiple of the standard error to add and subtract to the point estimate.

Bootstrap Interval Validity

The consistency of a bootstrap method implies the validity of the associated percentile confidence interval. If the asymptotic law is continuous, strictly increasing and symmetric, confidence interval using directly the percentile of $\left\{\overline{\overline{y}}_{b}^{*}\right\}$ is also valid [3].

For the consistency of percentile-t confidence interval, we need to show the consistency as expressed in *(5.1)* but applied to studentized statistics. The next proposition analyzes the case of the double resampling bootstrap.

[3] See Theorem 4.1 of Shao and Tu (1995) for technical proof.

Proposition 6 *In the presence of cross-sectional and/or temporal heterogeneity, under assumptions A - E and the assumption that* $l^{-1} + lT^{-1} = o(1)$ *as* $T \to \infty$*, we have :*

$$\sup_{x\in\mathbb{R}} \left| P^{**}\left(\frac{\overline{\overline{y}}^{**} - \overline{\overline{y}}}{\sqrt{\widehat{V}ar^{**}\left(\overline{\overline{y}}^{**}\right)}} \leq x \right) - P\left(\frac{\overline{\overline{y}} - \theta}{\sqrt{\widehat{V}ar\left(\overline{\overline{y}}\right)}} \leq x \right) \right| \underset{N,T\to\infty}{\overset{P}{\to}} 0 \quad (6.7)$$

where $\widehat{V}ar\left(\overline{\overline{y}}\right) = Var^{**}\left(\overline{\overline{y}}^{**}\right)$ *and* $\widehat{V}ar^{**}\left(\overline{\overline{y}}^{**}\right)$ *is the analog of* $Var^{**}\left(\overline{\overline{y}}^{**}\right)$ *on the pseudo-sample* Y^{**}*.*

The intuition is that with the consistency as defined in *(5.1)*, $Var^{**}\left(\overline{\overline{y}}^{**}\right)$ is asymptotically equivalent to $Var\left(\overline{\overline{y}}\right)$ thus it is a consistent estimator. In the bootstrap world the analog of $Var^{**}\left(\overline{\overline{y}}^{**}\right)$ is a good choice to studentize $\left(\overline{\overline{y}}^{**} - \overline{\overline{y}}\right)$. Like this the consistency with $t_b^{**}$is also given, justifying the use of percentile-t confidence interval. Similar results are given with the cross-sectional resampling bootstrap and the block resampling bootstrap using respectively $\widehat{V}ar\left(\overline{\overline{y}}\right) = Var^{*}\left(\overline{\overline{y}}^{*}_{cros}\right)$, $\widehat{V}ar\left(\overline{\overline{y}}\right) = Var^{*}\left(\overline{\overline{y}}^{*}_{bl}\right)$ and their analogs in the bootstrap world.

The consistency of the double resampling bootstrap percentile-t confidence interval, as defined in *(5.1)*, has been provided when N and T go to infinity. A question arises : the validity of the double resampling bootstrap method for inference when only one dimension goes to infinity. The next proposition compare the percentile-t confidence intervals.

Proposition 7 *For N and T large enough*

$$CI^{*cros}_{1-\alpha} \in CI^{**}_{1-\alpha} \quad (6.8)$$

$$CI^{*bl}_{1-\alpha} \in CI^{**}_{1-\alpha} \quad (6.9)$$

where

$$CI^{*cros}_{1-\alpha} = \left[\overline{\overline{y}} - \sqrt{Var^{*}\left(\overline{\overline{y}}^{*}_{cros}\right)}.t^{*cros}_{1-\frac{\alpha}{2}} ; \overline{\overline{y}} - \sqrt{Var^{*}\left(\overline{\overline{y}}^{*}_{cros}\right)}.t^{*cros}_{\frac{\alpha}{2}}\right]$$

$$CI^{*bl}_{1-\alpha} = \left[\overline{\overline{y}} - \sqrt{Var^{*}\left(\overline{\overline{y}}^{*}_{bl}\right)}.t^{*bl}_{1-\frac{\alpha}{2}} ; \overline{\overline{y}} - \sqrt{Var^{*}\left(\overline{\overline{y}}^{*}_{bl}\right)}.t^{*bl}_{\frac{\alpha}{2}}\right]$$

$$CI^{**}_{1-\alpha} = \left[\overline{\overline{y}} - \sqrt{Var^{**}\left(\overline{\overline{y}}^{**}\right)}.t^{**}_{1-\frac{\alpha}{2}} ; \overline{\overline{y}} - \sqrt{Var^{**}\left(\overline{\overline{y}}^{**}\right)}.t^{**}_{\frac{\alpha}{2}}\right]$$

For the validity of confidence interval associated to cross-sectional (resp. block) resampling bootstrap we need N (resp. T) to go to infinity. When the other dimension is large enough, the Proposition 7 ensures that the valid percentile-t confidence interval associated belong to the double resampling bootstrap percentile-t confidence interval that is valid even if the second dimension is fixed, in the sense that the level is controlled.

$$\Pr\left(\theta \in CI^{**}_{1-\alpha}\right) \geq 1 - \alpha \quad (6.10)$$

With all the theoretical results in hand, in the next section we will see the behavior of the bootstrap methods in finite sample, using simulations.

1.7 Simulations

This section presents results from a small simulation experiment to illustrate our theoretical results. The data generating process is (*2.1*) and (*2.2*).

The individual effects are standard normal and independent across units :

$$\mu_i \sim i.i.d.N\left(0,1\right),$$

while both the time effect and common factor are $AR\left(1\right)$ process with parameter 0.5 :

$$\begin{aligned} f_t &= \rho f_{t-1} + \varpi_t \\ F_t &= \rho F_{t-1} + \eta_t \\ \eta_t, \varpi_t &\sim i.i.d.N\left(0, \left(1-\rho^2\right)\right) \\ \rho &= 0.5 \end{aligned}$$

The factor loadings are standard normal :

$$\lambda_i \sim i.i.d.N\left(0,1\right)$$

and the idiosyncratic errors are also standard normal :

$$\varepsilon_{it} \sim i.i.d.N\left(0,1\right).$$

Six panel dimensions are considered : $(N,T) = (10,10)\,,(30,30)\,,(60,60)\,,$ $(10,6)$ and $(6,10)$. Temporal resampling is carried out with the Circular Block Bootstrap (CBB) with block length $l = 2,2,3,$ and 4 respectively for $T = 6$, 10, 30 and 60. For each bootstrap resampling scheme, B is equal to 999 and the number of simulations is 1000.

Tables 3 gives rejection rates for a two-tailed test for the null hypothesis that $\theta = 0$ at nominal level 5%. The rejection rates close to 5% are presented in bold.

The simulations confirm the theoretical results. In particular, we see that the double resampling performs well for all models considered : the cross-sectional and temporal one-way ECM, the two-way ECM and the factor model. The other bootstrap schemes fail for at least one model. The cross-sectional bootstrap performs well with one-way ECM and factor model, but cannot reproduce temporal heterogeneity. Similarly, the block bootstrap performs well with temporal one-way ECM, but it cannot provide reliable inference in the cross-sectional one-way or two-way ECM or the factor model. The implication of *Proposition 2* is visible in Table 3 : for any sample size, the double resampling bootstrap induces a rejection rate smaller than the block resampling bootstrap and the cross-sectional resampling bootstrap rejection rates.

1.8 Conclusion

This chapter considers bootstrap resampling for panel data..It is shown that double resampling that combines cross-sectional and block resampling is valid for panel data models with cross-sectional and/or temporal heterogeneity. Some weak forms of spatial and serial dependence in the idiosyncratic errors can even be allowed for if both the cross-sectional and time dimensions are large. On the other hand, resampling only in the cross-sectional dimension is not valid in presence of temporal heterogeneity, and block resampling in the time dimension only is not valid in the presence of cross-sectional heterogeneity.

There are two important advantages of the methods proposed in this paper. The first one is that double resampling is able to replicate the behavior of the error term, without having to separate it into components (which would require making strong parametric assumptions). Secondly, the bootstrap has the nice advantage of avoiding having to choose among multiple asymptotic approximations.

There are several directions in which the current work can be extended to be made more realistic. One would be to relax some of the strong assumptions that were made on the individual effects. Also, one would like to introduce regressors in the model. This will be the subject of the next chapter.

Table 3 : Simulation results with percentile-t

Models	(N ;T)	Cross.	Block	D-Res
	(10 ;10)	**5.5**	60.2	**4.2**
Cross − sectional	(30 ;30)	**4.9**	73.1	**4.4**
	(60 ;60)	**5.2**	79.8	**5.1**
One − way	(10 ;06)	**6.9**	57.9	**5.1**
ECM	(06 ;10)	10.8	62.6	**6.8**
	(10 ;10)	58.8	11.8	**6.6**
Temporal	(30 ;30)	73.1	**6.4**	**5.7**
	(60 ;60)	81.1	**6.3**	**5.5**
One − way	(10 ;06)	59.1	17.7	10.7
ECM	(06 ;10)	57.1	11.4	**5.0**
	(10 ;10)	20.5	23.5	**5.5**
Two − way	(30 ;30)	18.6	20.3	**5.2**
	(60 ;60)	17.5	18.6	**5.3**
ECM	(06 ;10)	19.2	28.4	**5.6**
	(10 ;06)	19.5	28.1	**6.5**
	(10 ;10)	8.3	52.9	**4.1**
	(30 ;30)	**6.4**	65.1	**5.1**
Factor	(60 ;60)	**4.6**	75.3	**4.1**
model	(06 ;10)	10.7	48.7	**5.1**
	(10 ;06)	9.5	51.1	**4.5**

The rejection rates close to 5% are presented in bold

APPENDIX

Proposition 8 *: Assume that A2 holds, assume also that* $l^{-1}+lT^{-1}=o\left(1\right)$ *as* $T\rightarrow\infty$*,, using NMB, MBB or CBB, we have*

$$\sup_{x\in\mathbb{R}}\left|P^{*}\left(\sqrt{T}\left(\overline{f}_{bl}^{*}-\overline{f}\right)\leq x\right)-P\left(\sqrt{T}\left(\overline{f}-0\right)\leq x\right)\right|\underset{T\rightarrow\infty}{\overset{P}{\rightarrow}}0$$

$$\sqrt{T}\left(\overline{f}_{bl}^{*}-\overline{f}\right)\underset{T\rightarrow\infty}{\overset{*}{\Longrightarrow}}N\left(0,V_{f}^{\infty}\right)$$

Proof. [Proof of Proposition 8] *Under the assumptions and the convergence rate imposed to l, a demonstration of the consistency of MMB, NMB and CBB for time series, can be seen for example in Lahiri (2003), p. 55 .* ■

Classical Asymptotic Distributions

Cross-sectional one-way ECM

a) T is fixed. $\overline{y}_{i.}$ are i.i.d. with $E\left(\overline{y}_{i.}\right)=\theta$ and $Var\left(\overline{y}_{i.}\right)=\sigma_{\mu}^{2}+\frac{\sigma_{\varepsilon}^{2}}{T}$. *Applying a standard CLT, the result follows.*

b)

$$\sqrt{N}\left(\overline{\overline{y}}-\theta\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mu_{i}+\sqrt{N}\overline{\overline{\varepsilon}}$$

$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mu_{i}\underset{N\rightarrow\infty}{\Longrightarrow}N\left(0,\sigma_{\mu}^{2}\right)$ by CLT and $\sqrt{N}\overline{\overline{\varepsilon}}\underset{N,T\rightarrow\infty}{\overset{P}{\rightarrow}}0$ (Assumption C')

The result follows

Temporal one-way ECM

a)$T \to \infty$

$$\sqrt{T}\left(\overline{\overline{y}}-\theta\right)=\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\right)+\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{\varepsilon}_{.t}\right)$$

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\right)\underset{T\to\infty}{\Longrightarrow}N\left(0,V_f^{\infty}\right)\ \ (\textit{Proposition 6})\quad\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{\varepsilon}_{.t}\right)\underset{T\to\infty}{\Longrightarrow}N\left(0,\frac{\sigma_\varepsilon^2}{N}\right)$$

thus

$$\sqrt{T}\left(\overline{\overline{y}}-\theta\right)\underset{T\to\infty}{\Longrightarrow}N\left(0,V_f^{\infty}+\frac{\sigma_\varepsilon^2}{N}\right)$$

b)$N,T\to\infty$

$$\sqrt{T}\left(\overline{\overline{y}}-\theta\right)=\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\right)+\sqrt{T}\overline{\overline{\varepsilon}}$$

$$\sqrt{N}\overline{\overline{\varepsilon}}\underset{N,T\to\infty}{\overset{P}{\to}}0\ \text{(C')}$$

thus

$$\sqrt{T}\left(\overline{\overline{y}}-\theta\right)\underset{N,T\to\infty}{\Longrightarrow}N\left(0,V_f^{\infty}\right)$$

Two-way ECM

a)$\frac{N}{T}\to\delta\in[0,\infty)$

$$\sqrt{N}\left(\overline{\overline{y}}-\theta\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mu_i+\frac{\sqrt{N}}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\right)+\sqrt{N}\overline{\overline{\varepsilon}}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mu_i\underset{N\to\infty}{\Longrightarrow}N\left(0,\sigma_\mu^2\right)\ \ \text{by CLT;}\quad\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\right)\underset{N,T\to\infty}{\Longrightarrow}N\left(0,V_f^{\infty}\right)$$

.

$$\sqrt{N}\overline{\overline{\varepsilon}}\underset{N,T\to\infty}{\overset{P}{\to}}0\ \text{(C')}$$

The result follows[4].

[4]When the vector$(X_n,Y_n)^{'}$ converges to a normal distribution, the asymptotic distribu-

b) $\frac{N}{T} \to \infty$

$$\sqrt{T}\left(\overline{\overline{y}}-\theta\right)=\frac{\sqrt{T}}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mu_i\right)+\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\right)+\sqrt{T}\overline{\overline{\varepsilon}}$$

$$\sqrt{N}\overline{\overline{\varepsilon}}\underset{N,T\to\infty}{\overset{P}{\to}}0;\quad \frac{\sqrt{T}}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mu_i\right)\underset{N,T\to\infty}{\overset{m.s.}{\to}}0$$

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\right)\underset{T\to\infty}{\Longrightarrow}N\left(0,V_f^{\infty}\right)$$

The result follows.

Factor Models

$$\sqrt{N}\left(\overline{\overline{y}}-\theta\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mu_i+\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\right)\left(\frac{1}{T}\sum_{t=1}^{T}F_t\right)+\sqrt{N}\overline{\overline{\varepsilon}}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mu_i\underset{N\to\infty}{\Longrightarrow}N\left(0,\sigma_\mu^2\right)$$

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t\right)\underset{N,T\to\infty}{\Longrightarrow}\left[N\left(0,\sigma_\lambda^2\right)\right]*\left[N\left(0,V_f^{\infty}\right)\right]$$

thus

$$\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\right)\left(\frac{1}{T}\sum_{t=1}^{T}F_t\right)\underset{N,T\to\infty}{\overset{m.s.}{\to}}0$$

$$\sqrt{N}\overline{\overline{\varepsilon}}\underset{N,T\to\infty}{\overset{P}{\to}}0\ (\text{ C'})$$

and we have

$$\sqrt{N}\left(\overline{\overline{y}}-\theta\right)\underset{N,T\to\infty}{\Longrightarrow}N\left(0,\sigma_\mu^2\right)$$

---

tion of any linear combination of the elements of the vector (in particular the sum) can be deduced. The fact that X_n and Y_n are independent and converge to a normal distribution, implies that their sum converge to the sum of their asymptotic normal distributions.

Proof. [Proof of Proposition 1] *$\overline{\overline{y}}$ and $\overline{\overline{y}}^{*}$ having the same asymptotic distribution, implies that $|P^{*}(..)-P(..)|$ converges to zero. Under the continuity assumption, uniform convergence is given by the Pólya theorem (Pólya (1920) or Serfling (1980), p. 18)* ■

Proof. [Proof of Proposition 2] *An analysis of variance gives :*

Using CBB, there is the time dimension, K blocks are chosen from T possible blocks. The bootstrap-mean $\overline{\overline{y}}^{**}$ rewritten as

$$\overline{\overline{y}}^{**}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it}^{**}=\frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K}z_{ik}^{**}$$

where

$$z_{ik}^{**}=\frac{1}{l}\sum_{t\,\in\,block\;k}y_{it}^{**}$$

$$Var^{**}\left(\overline{\overline{y}}^{**}\right)=Var^{**}\left(\overline{\overline{z}}^{**}\right)=\frac{1}{NK}Var^{**}\left(z_{it}^{**}\right)+\frac{1}{(NK)^{2}}\sum_{(i,k)\neq}\sum_{(j,s)}Cov^{**}\left(z_{ik}^{**},z_{js}^{**}\right)$$

z_{ik}^{**} can take any of the $N*T$ values of elements of $[Z]$ with probability $1/NT$ then the expectation and the variance are identical to those obtained with i.i.d. bootstrap accommodated to panel data $[Z]$: $E^{**}\left(z_{it}^{**}\right)=E^{*}\left(z_{it}^{*}\right)$; $Var^{**}\left(z_{it}^{**}\right)=Var^{*}\left(z_{it}^{*}\right)$.

$$\text{For } i\neq j \text{ and } k\neq s,\ Cov^{**}\left(z_{ik}^{**},z_{js}^{**}\right)=0$$

$$\begin{aligned}\frac{1}{(NK)^{2}}\sum_{(i,k)\neq}\sum_{(j,s)}Cov^{**}\left(z_{ik}^{**},z_{js}^{**}\right) &= \frac{1}{(NK)^{2}}\sum_{k=1}^{K}\sum_{i\neq}\sum_{j}Cov^{**}\left(z_{ik}^{**},z_{jk}^{**}\right)\\ &\quad+\frac{1}{(NK)^{2}}\sum_{i=1}^{N}\sum_{t\neq}\sum_{s}Cov^{**}\left(z_{ik}^{**},z_{is}^{**}\right)\end{aligned}$$

$$
\begin{aligned}
Cov^{**}\left(z_{ik}^{**}, z_{jk}^{**}\right) &= \frac{1}{N^2T}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N} z_{ik}z_{jk} - \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{k=1}^{T} z_{ik}\right)^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} z_{ik}\right)^2 - \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\right)^2 \\
&= \frac{1}{T}\sum_{k=1}^{T}\left(\overline{z}_{.k}\right)^2 - \left(\frac{1}{T}\sum_{k=1}^{T}\overline{z}_{.k}\right)^2 \\
&= Var^*\left(\overline{z}_{.k}^*\right)
\end{aligned}
$$

Similary

$$
Cov^{**}\left(z_{ik}^{**}, z_{is}^{**}\right) = Var^*\left(\overline{z}_{i.}^*\right)
$$

$$
\begin{aligned}
\overline{z}_{i.}^* &= \frac{1}{K}\sum_{k=1}^{K} z_{ik} = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{1}{l}\sum_{t\,\in\, block\ k} y_{it}^{**}\right] \\
\overline{z}_{i.}^* &= \frac{1}{Kl}\sum_{t=1}^{T} y_{it}^{**} = \frac{1}{T}\sum_{t=1}^{T} y_{it}^{**} = \overline{y}_{i.}^*
\end{aligned}
$$

There are $T\left(N^2 - N\right)$ possibilities of $Cov^{**}\left(z_{ik}^{**}, z_{jk}^{**}\right)$. There are $N\left(T^2 - T\right)$ possibilities of $Cov^{**}\left(z_{ik}^{**}, z_{is}^{**}\right)$ then :

$$
\begin{aligned}
Var^{**}\left(\overline{\overline{y}}^{**}\right) &= \frac{Var^*\left(z_{ik}^*\right)}{NT} + \left(1 - \frac{1}{T}\right)\frac{Var^*\left(\overline{z}_{i.}^*\right)}{N} + \left(1 - \frac{1}{N}\right)\frac{Var^*\left(\overline{z}_{.k}^*\right)}{T} \\
Var^{**}\left(\overline{\overline{y}}^{**}\right) &= Var^*\left(\overline{\overline{z}}^*\right) + \left(1 - \frac{1}{T}\right)\frac{Var^*\left(\overline{y}_{i.}^*\right)}{N} + \left(1 - \frac{1}{N}\right)\frac{Var^*\left(\overline{z}_{.k}^*\right)}{T} \\
Var^{**}\left(\overline{\overline{y}}^{**}\right) &= Var^*\left(\overline{\overline{z}}^*\right) + \left(1 - \frac{1}{T}\right)Var^*\left(\overline{\overline{y}}_{cros}^*\right) + \left(1 - \frac{1}{N}\right)Var^*\left(\overline{\overline{y}}_{bl}^*\right) \\
Var^{**}\left(\overline{\overline{y}}^{**}\right) &> \left(1 - \frac{1}{T}\right)Var^*\left(\overline{\overline{y}}_{cros}^*\right) \\
Var^{**}\left(\overline{\overline{y}}^{**}\right) &> \left(1 - \frac{1}{N}\right)Var^*\left(\overline{\overline{y}}_{bl}^*\right)
\end{aligned}
$$

■

Proof. [Proof of Proposition 3] *Variance decomposition in the proof of Proposition 2 gives :*

$$Var^{**}\left(\sqrt{NT}\;\overline{\overline{\varepsilon}}^{**}\right) = Var^{*}\left(\varepsilon_{it}^{*}\right)+\left(1-\frac{1}{T}\right)\left[T.Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right)\right]+\left(1-\frac{1}{N}\right)\left[N.Var^{*}\left(\overline{\varepsilon}_{.t}^{*}\right)\right]$$

$$Var^{*}\left(\varepsilon_{it}^{*}\right)\underset{NT\to\infty}{\overset{P}{\to}}\sigma_{\varepsilon}^{2}\;,\;\;\left[T.Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right)\right]\underset{N\to\infty}{\overset{P}{\to}}\sigma_{\varepsilon}^{2}\;,\;\;\left[N.Var^{*}\left(\overline{\varepsilon}_{.t}^{*}\right)\right]\underset{T\to\infty}{\overset{P}{\to}}\sigma_{\varepsilon}^{2}$$

therefore

$$Var^{**}\left(\sqrt{NT}\;\overline{\overline{\varepsilon}}^{**}\right)\underset{N,T\to\infty}{\overset{P}{\to}}3.\sigma_{\varepsilon}^{2}$$

■

Proof. [Proof of Proposition 4]

$$\begin{aligned} Var^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right) &= E^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right)^{2}-\left[E^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right)\right]^{2} \\ &= E^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right)^{2}-\left[\sqrt{N}\overline{\overline{\varepsilon}}\right]^{2} \end{aligned}$$

By assumption $\sqrt{N}\overline{\overline{\varepsilon}}\overset{P}{\to}0$ thus $\left[\sqrt{N}\overline{\overline{\varepsilon}}\right]^{2}\overset{P}{\to}0$. Let's study now the behavior of $E^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right)^{2}$.

The block size is l and we have K blocks.

$$\overline{\overline{\varepsilon}}^{**}=\frac{1}{NT}\sum_{i}\sum_{t}\varepsilon_{it}^{**}=\frac{1}{NK}\sum_{i}\sum_{k}e_{ik}^{**}=\frac{1}{NK}\sum_{i}\sum_{k}e_{it}\alpha_{i}\beta_{k}$$

where e_{ik}^{**} denote for given i the average of observations in the block k.

$\alpha_i\sim Multinomial\left(N,p_1=p_2=......=p_N=\frac{1}{N}\right)$ and

$\beta_k\sim Multinomial\left(K,p_1=p_2=......=p_K=\frac{1}{\tau}\right)$ τ denotes the number of potential block to resample from Y. For the case of CBB $\tau=T$ α_i (resp. β_k)denotes how much time individual i (respectively block k)appears in the pseudo-sample Y^{*}. α_i is independent of β_k and both are indepndent of the observations :

$$\begin{aligned}
E^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right)^2 &= N * E^{**}\left(\frac{1}{NK}\sum_i\sum_k e_{ik}\alpha_i\beta_k\right)^2 \\
&= N * E^{**}\left(\left(\frac{1}{NK}\right)^2 \sum_i\sum_j\sum_k\sum_s e_{ik}e_{js}\alpha_i\alpha_j\beta_k\beta_s\right) \\
&= N * \left(\frac{1}{NK}\right)^2 \sum_i\sum_j\sum_k\sum_s e_{ik}e_{js}E^{**}\left(\alpha_i\alpha_j\beta_k\beta_s\right) \\
&= N * \left(\frac{1}{NK}\right)^2 \sum_i\sum_j\sum_k\sum_s e_{ik}e_{js}E^{*}\left(\alpha_i\alpha_j\right)E^{*}\left(\beta_k\beta_s\right)
\end{aligned}$$

if $i = j,\ E^*(\alpha_i\alpha_j) = E^*(\alpha_i^2) = Var^*(\alpha_i) + [E^*(\alpha_i)]^2$

$= N * p_i(1 - p_i) + N * p_i$

$= N * \frac{1}{N}\left(1 - \frac{1}{N}\right) + 1 = \left(1 - \frac{1}{N}\right) + 1$

if $i \neq j\ E^*(\alpha_i\alpha_j) = Cov^*(\alpha_i\alpha_j) + E^*(\alpha_i)E^*(\alpha_j)$

$= -N * p_i * p_j + N * p_i * N * p_j$

$= -\frac{1}{N} + 1 = \left(1 - \frac{1}{N}\right)$

if $k = s,\ E^*(\beta_k\beta_s) = E^*\left(\beta_k^2\right) = Var^*(\beta_k) + [E^*(\beta_k)]^2$

$= K * p_k(1 - p_k) + [K * p_k]^2$

$= K * \frac{1}{T}\left(1 - \frac{1}{T}\right) + \left(K * \frac{1}{T}\right)^2 = \frac{1}{l}\left(1 - \frac{1}{T}\right) + \left(\frac{1}{l}\right)^2$

$= \frac{1}{l}\left[\left(\frac{1}{l} - \frac{1}{T}\right) + 1\right]$

if $k \neq s\ E^*(\beta_k\beta_s) = -K * p_k * p_s + K * p_k * K * p_s$

$= -K * \frac{1}{T} * \frac{1}{T} + K * \frac{1}{T} * K * \frac{1}{T} = -\frac{1}{l}\frac{1}{T} + \left(\frac{1}{l}\right)^2$

$= \frac{1}{l}\left[\frac{1}{l} - \frac{1}{T}\right]$

$$\begin{aligned}
E^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right)^2 &= N\left(\frac{1}{NK}\right)^2\sum_i\sum_j\sum_k\sum_s e_{it}e_{js}E^*\left(\alpha_i\alpha_j\right)E^*\left(\beta_k\beta_s\right)\\
&= N*\left(\frac{1}{NK}\right)^2\sum_i\sum_j\sum_k\sum_s e_{ik}e_{js}\left(1-\frac{1}{N}\right)\frac{1}{l}\left[\frac{1}{l}-\frac{1}{T}\right]\\
&\quad +N*\left(\frac{1}{NK}\right)^2\sum_i\sum_k\sum_s e_{ik}e_{is}+N*\left(\frac{1}{NK}\right)^2\sum_i\sum_j\sum_k e_{ik}e_{jk}\left(\frac{1}{l}\right)\\
&= N*\left(1-\frac{1}{N}\right)\frac{1}{l}\left[\frac{1}{l}-\frac{1}{T}\right]\left[\frac{1}{NK}\sum_i\sum_k e_{ik}\right]^2\\
&\quad +N*\left(\frac{1}{K}\right)^2\sum_k\left[\frac{1}{N}\sum_i e_{ik}\right]^2\left(\frac{1}{l}\right)\\
&\quad +N*\left(\frac{1}{N}\right)^2\sum_i\left[\frac{1}{K}\sum_k e_{ik}\right]^2
\end{aligned}$$

$$\begin{aligned}
\frac{1}{NK}\sum_i\sum_t e_{it} &= \frac{1}{NT}\sum_i\sum_t\varepsilon_{it}\equiv\overline{\overline{\varepsilon}}\\
\frac{1}{K}\sum_k e_{ik} &= \frac{1}{T}\sum_t\varepsilon_{it}\equiv\overline{\varepsilon}_{i.},
\end{aligned}$$

$thus$

$$\begin{aligned}
E^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right)^2 &= N*\left(1-\frac{1}{N}\right)\frac{1}{l}\left[\frac{1}{l}-\frac{1}{T}\right]\left[\overline{\overline{\varepsilon}}\right]^2+\frac{1}{N}\sum_i\left[\overline{\varepsilon}_{i.}\right]^2+\frac{N}{K}\left(\frac{1}{l}\right)\frac{1}{K}\sum_t\left[\overline{\varepsilon}_{.t}\right]^2\\
&= \left(1-\frac{1}{N}\right)\frac{1}{l}\left[\frac{1}{l}-\frac{1}{T}\right]\underbrace{\left[\sqrt{N}\overline{\overline{\varepsilon}}\right]^2}_{\xrightarrow{P}0}+\underbrace{\frac{1}{N}\sum_i\left[\overline{\varepsilon}_{i.}\right]^2}_{\xrightarrow{P}0}+\frac{N}{T}\underbrace{\frac{1}{K}\sum_k\left[\overline{\varepsilon}_{.k}\right]^2}_{\xrightarrow{P}0}\\
&\xrightarrow{P}0
\end{aligned}$$

thus $Var^{**}\left(\sqrt{N}\overline{\overline{\varepsilon}}^{**}\right)\xrightarrow{P}0$

Similary,

$$
\begin{aligned}
E^{**}\left(\sqrt{T}\overline{\overline{\varepsilon}}^{**}\right)^2 &= T*\left(1-\frac{1}{N}\right)\frac{1}{l}\left[\frac{1}{l}-\frac{1}{T}\right]\left[\overline{\overline{\varepsilon}}\right]^2+\frac{1}{N}\sum_i\left[\overline{\varepsilon}_{i.}\right]^2+\frac{N}{K}\left(\frac{1}{l}\right)\frac{1}{K}\sum_t\left[\overline{\varepsilon}_{.t}\right]^2 \\
&= \left(1-\frac{1}{N}\right)\frac{1}{l}\left[\frac{1}{l}-\frac{1}{T}\right]\underbrace{\left[\sqrt{T}\overline{\overline{\varepsilon}}\right]^2}_{\xrightarrow{P}0}+\frac{T}{N}\underbrace{\frac{1}{N}\sum_i\left[\overline{\varepsilon}_{i.}\right]^2}_{\xrightarrow{P}0}+\underbrace{\frac{1}{K}\sum_k\left[\overline{\varepsilon}_{.k}\right]^2}_{\xrightarrow{P}0} \\
&\xrightarrow{P} 0
\end{aligned}
$$

thus $Var^{**}\left(\sqrt{T}\overline{\overline{\varepsilon}}^{**}\right)\xrightarrow{P}0$. ■

Proof. [Proof of Proposition 5] *With the cross-sectional resampling, the centering eliminates the temporal heterogeneity* $\{f_t\}$. *Its behavior does not appear neither in finite sample properties nor in the asymptotic distribution and therefore causes inconsistency. With block resampling, the centering eliminates the cross-sectional heterogeneity* $\{\mu_i\}$, *therefore insconsistency. With the double resampling, all the properties demonstrated for the i.i.d. bootstrap or for the various block bootstrap methods are transferred to the appropriate errors terms without restriction. With the different specifications, the consistency or the inconsistency holds comparing the bootstrap asymptotic distribution and the classic asymptotic distribution, according Proposition 1.* ■

Proof. [Proof of Proposition 6] *We have already the consistency of Proposition 5 in hand.*

with $\widehat{V}ar\left(\overline{\overline{y}}\right)=Var^{**}\left(\overline{\overline{y}}^{**}\right)$ *and* $\widehat{V}ar^{**}\left(\overline{\overline{y}}^{**}\right)$ *is the analog of* $Var^{**}\left(\overline{\overline{y}}^{**}\right)$ *on the pseudo-sample* Y^{**}, *the proposition 4.1 of Shao and Tu (1995) ensures that the consistency of Proposition 5 implies the result in Proposition 6.* ■

Proof. [Proof of Proposition 7] For the three bootstrap methods we have :

$$\sup_{x\in\mathbb{R}} |P^* (t_b^* \leq x) - \Phi(x)| \underset{NT\to\infty}{\overset{P}{\to}} 0$$

For N and$T$ large enough, $t_{1-\frac{\alpha}{2}}^{*cros} \sim t_{1-\frac{\alpha}{2}}^{*bl} \sim t_{1-\frac{\alpha}{2}}^{**} > 0$ and $t_{\frac{\alpha}{2}}^{*cros} \sim t_{\frac{\alpha}{2}}^{*bl} \sim t_{\frac{\alpha}{2}}^{**} <$

0

$Var^{**}\left(\overline{\overline{y}}^{**}\right) \geq Var^{*}\left(\overline{\overline{y}}^{*}_{cross}\right)$
$Var^{**}\left(\overline{\overline{y}}^{**}\right) \geq Var^{*}\left(\overline{\overline{y}}^{*}_{bl}\right)$

$$\begin{aligned}
CI_{1-\alpha}^{*cros} &= \left[\overline{\overline{y}} - \sqrt{Var^{*}\left(\overline{\overline{y}}^{*}_{cros}\right)}.t_{1-\frac{\alpha}{2}}^{*cros}; \overline{\overline{y}} - \sqrt{Var^{*}\left(\overline{\overline{y}}^{*}_{cros}\right)}.t_{\frac{\alpha}{2}}^{*cros}\right] \\
&\in \left[\overline{\overline{y}} - \sqrt{Var^{**}\left(\overline{\overline{y}}^{**}\right)}.t_{1-\frac{\alpha}{2}}^{*cros}; \overline{\overline{y}} - \sqrt{Var^{**}\left(\overline{\overline{y}}^{**}_{cros}\right)}.t_{\frac{\alpha}{2}}^{*cros}\right] \\
&\sim \left[\overline{\overline{y}} - \sqrt{Var^{**}\left(\overline{\overline{y}}^{**}\right)}.t_{1-\frac{\alpha}{2}}^{**}; \overline{\overline{y}} - \sqrt{Var^{**}\left(\overline{\overline{y}}^{**}_{cros}\right)}.t_{\frac{\alpha}{2}}^{**}\right] = CI_{1-\alpha}^{**}
\end{aligned}$$

Similary

$$CI_{1-\alpha}^{*bl} \in CI_{1-\alpha}^{**}$$

■

Cross-sectional Resampling Bootstrap

$$\left(\overline{\overline{y}}_{cros}^{*}-\overline{\overline{y}}\right)=\left(\overline{\mu}^{*}-\overline{\mu}\right)+\left(\overline{\lambda}^{*}\overline{F}-\overline{\lambda F}\right)+\left(\left[\overline{\varepsilon}_{inter}\right]^{*}-\overline{\overline{\varepsilon}}\right)$$

$$\begin{aligned} Var^{*}\left(\sqrt{N}\left[\overline{\varepsilon}_{inter}\right]^{*}\right) &= N*Var^{*}\left[\frac{1}{N}\sum_{i}\overline{\varepsilon}_{i.}^{*}\right]=N\frac{1}{N^{2}}\sum_{i}Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right) \\ &= N\frac{1}{N^{2}}N*Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right)=Var^{*}\left(\overline{\varepsilon}_{i.}^{*}\right) \\ &= \underbrace{\frac{1}{N}\sum_{i}\overline{\varepsilon}_{i.}^{2}}_{\xrightarrow{P}0(C2'')}-\underbrace{\left(\overline{\overline{\varepsilon}}\right)^{2}}_{\xrightarrow{P}0(C1'')}\quad\xrightarrow{P}0 \end{aligned}$$

Thus under assumption C', the behavior $\left(\left[\overline{\varepsilon}_{inter}\right]^{*}-\overline{\overline{\varepsilon}}\right)$ doesn't appear in the asymptotic distribution of $\sqrt{N}\left(\overline{\overline{y}}_{cros}^{*}-\overline{\overline{y}}\right)$.

$$\sqrt{N}\left(\overline{\lambda}^{*}\overline{F}-\overline{\lambda F}\right)=\sqrt{N}\left(\overline{\lambda}^{*}-\overline{\lambda}\right)\overline{F}$$

$\overline{F}\xrightarrow{P}0.$ and $\sqrt{N}\left(\overline{\lambda}^{*}-\overline{\lambda}\right)\Rightarrow N\left(0,\sigma_{\lambda}^{2}\right)$ thus $\sqrt{N}\left(\overline{\lambda}^{*}\overline{F}-\overline{\lambda F}\right)\xrightarrow{P}0.$

Remarks

A - For the negligibility of ε_{it} using the cross-sectional resampling bootstrap, we need

1 - $C''1$: $\sqrt{N}\overline{\overline{\varepsilon}}\xrightarrow{P}0$ for the classical asymptotic distribution of $\overline{\overline{y}}$.

2 - $C''2$: $\frac{1}{N}\sum_{i}\overline{\varepsilon}_{i.}^{2}\xrightarrow{P}0$ for the classical asymptotic distribution of $\overline{\overline{y}}_{cros}^{*}$.

3 - N to go do infinity.

B - Validity of the cross-sectional resampling bootstrap :

For the sample mean, the cross-sectional resampling bootstrap is equivalent to i.i.d. bootstrap of averages $[\overline{y}_{1.}, \overline{y}_{2.}, ..., \overline{y}_{N.}]$. The minimal assumptions of validity is that $\overline{y}_{i.}$ are i.i.d. of decomposable on a i.i.d. term μ_i plus a asymptotically negligible term ε_{it}.

Table 4 summarizes the asymptotic distributions of $\left(\overline{\overline{y}}^*_{cros} - \overline{\overline{y}}\right)$ for the different panel models.

Table 4 : Asymptotic distributions with cross sectional resampling bootstrap

Model	*Assumptions*	*Asymptotic distribution*	ω
Cross − sect.	$A,\ C$	$\sqrt{N}\left(\overline{\overline{y}}^{*}-\overline{\overline{y}}\right)\underset{N\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\omega\right)$	$\sigma_{\mu}^{2}+\frac{\sigma_{\varepsilon}^{2}}{T}$
One − way *ECM*	$A,\ C''$	$\sqrt{N}\left(\overline{\overline{y}}^{*}-\overline{\overline{y}}\right)\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\omega\right)$	σ_{μ}^{2}
Temporal	$B,\ C$	$\sqrt{NT}\left(\overline{\overline{y}}^{*}-\overline{\overline{y}}\right)\underset{N\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\omega\right)$	σ_{ε}^{2}
One − way *ECM*	$B,\ C''$	$\sqrt{N}\left(\overline{\overline{y}}^{*}-\overline{\overline{y}}\right)\underset{N,T\to\infty}{\overset{m.s.*}{\longrightarrow}}0$	
Two − way	$A,\ B,\ C,D$	$\sqrt{N}\left(\overline{\overline{y}}^{*}-\overline{\overline{y}}\right)\underset{N\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\omega\right)$	$\sigma_{\mu}^{2}+\frac{\sigma_{\varepsilon}^{2}}{T}$
ECM	$A,\ B,\ C'',D$	$\sqrt{N}\left(\overline{\overline{y}}^{*}-\overline{\overline{y}}\right)\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\omega\right)$	σ_{μ}^{2}
Factor model	A,C'',E	$\sqrt{N}\left(\overline{\overline{y}}^{*}-\overline{\overline{y}}\right)\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\omega\right)$	σ_{μ}^{2}

Block Resampling Bootstrap

$$\left(\overline{\overline{y}}_{bl}^{*}-\overline{\overline{y}}\right)=\left(\overline{f}_{bl}^{*}-\overline{f}\right)+\left(\overline{\lambda F}_{bl}^{*}-\overline{\lambda F}\right)+\left(\left[\overline{\varepsilon}_{inter}\right]_{bl}^{*}-\overline{\overline{\varepsilon}}\right)$$

Assume for simplicity that T/l is an integer. We draw with replacement $K=T/l$ blocks

$$Var^{*}\left(\sqrt{T}\left[\overline{\varepsilon}_{inter}\right]_{bl}^{*}\right)=T*Var^{*}\left[\frac{1}{K}\sum_{k}\overline{e}_{k.}^{*}\right]$$

where $\overline{e}_{k.}^{*}=\frac{1}{l}\sum_{\in\ block\ k}\overline{\varepsilon}_{.t}^{*}$ is the average of the elements of the block k. Conditionally on the sample, $\overline{e}_{k.}^{*}$ are i.i.d. thus :

$$Var^{*}\left(\sqrt{T}\left[\overline{\varepsilon}_{inter}\right]_{bl}^{*}\right)=T*Var^{*}\left[\frac{1}{K}\sum_{k}\overline{e}_{k.}^{*}\right]=T*\frac{Var^{*}\left(\overline{e}_{k.}^{*}\right)}{K}$$

$$Var^{*}\left(\overline{e}_{k.}^{*}\right)=\left[\frac{1}{\tau}\sum_{k}\left(\overline{e}_{k.}\right)^{2}-\left(\overline{\overline{\varepsilon}}\right)^{2}\right]$$

τ depends on the block bootstrap method. For CBB there are $\tau=T$ possible values.

$$Var^{*}\left(\sqrt{T}\left[\overline{\varepsilon}_{inter}\right]_{bl}^{*}\right)=\frac{T}{K}\left[\frac{1}{T}\sum_{k}\left(\overline{e}_{k.}\right)^{2}-\left(\overline{\overline{\varepsilon}}\right)^{2}\right]=\underbrace{\frac{1}{K}\sum_{k}\left(\overline{e}_{k.}\right)^{2}}_{\xrightarrow{P}0(C')}-\frac{1}{K}\underbrace{\left(\sqrt{T}\overline{\overline{\varepsilon}}\right)^{2}}_{\xrightarrow{P}0(C')}$$

where

$$\frac{1}{K}\sum_{k}\left(\overline{e}_{k.}\right)^{2}=\frac{1}{K}\sum_{k}\left(\frac{1}{l}\sum_{\in\ block\ k}\overline{\varepsilon}_{.t}\right)^{2}$$
$$=\underbrace{\frac{1}{K}\sum_{k}\left(\overline{e}_{k.}\right)^{2}}_{\xrightarrow{P}0(C''3)}-\underbrace{\left(\overline{\overline{\varepsilon}}\right)^{2}}_{\xrightarrow{P}0(C''1)}\quad\xrightarrow{P}0$$

Thus under assumption C', the behavior $\left([\overline{\varepsilon}_{inter}]^{*} - \overline{\overline{\varepsilon}}\right)$ doesn't appear in the asymptotic distribution of $\sqrt{T}\left(\overline{\overline{y}}_{bl}^{*} - \overline{\overline{y}}\right)$.

Remarks

A - For the negligibility of ε_{it} using the block resampling bootstrap, we need :

1 - $C''1$: $\sqrt{T}\overline{\overline{\varepsilon}} \xrightarrow{P} 0$ for the classical asymptotic distribution of $\overline{\overline{y}}$

2 - $C''3$: $\frac{1}{K}\sum_{k}\left(\overline{e}_{k.}\right)^{2} \xrightarrow{P} 0$ for the classical asymptotic distribution of $\overline{\overline{y}}_{bl}^{*}$

3 - T to go do infinity.

B - Validity of the block resampling bootstrap, we need :

For the sample mean, the block resampling bootstrap is equivalent to block bootstrap on averages by time periods $[\overline{y}_{.1}, \overline{y}_{.2}, ..., \overline{y}_{.T}]$, the minimal assumption of validity is that $\overline{y}_{.t}$ is decomposable in a $\alpha-$mixing process f_{t} (verifying Ibragimov's assumptions) plus a asymptotically negligible error term ε_{it}.

Table 5 summarizes the asymptotic distributions of $\left(\overline{\overline{y}}_{bl}^{*} - \overline{\overline{y}}\right)$ for the different panel model specifications.

Table 5 : Asymptotic distributions with block resampling bootstrap

Model	*Assump.*	*Asymptotic distribution*	ω
Cross − sect. One − way ECM	A, B	$\sqrt{NT}\left(\overline{\overline{y}}^{*}_{bl} - \overline{\overline{y}}\right) \underset{T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \omega\right)$	σ^2_{ε}
	A, C''	$\sqrt{T}\left(\overline{\overline{y}}^{*}_{bl} - \overline{\overline{y}}\right) \underset{N,T\to\infty}{\overset{m.s.*}{\to}} 0$	
Temporal One − way ECM	B, C	$\sqrt{T}\left(\overline{\overline{y}}^{*}_{bl} - \overline{\overline{y}}\right) \underset{T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \omega\right)$	$V^{\infty}_{f} + \frac{\sigma^2_{\varepsilon}}{N}$
	B, C''	$\sqrt{T}\left(\overline{\overline{y}}^{*}_{bl} - \overline{\overline{y}}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \omega\right)$	V^{∞}_{f}
Two − way	A, B, C, D	$\sqrt{T}\left(\overline{\overline{y}}^{*}_{bl} - \overline{\overline{y}}\right) \underset{T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \omega\right)$	$V^{\infty}_{f} + \frac{\sigma^2_{\varepsilon}}{N}$
ECM	A, B, C'', D	$\sqrt{T}\left(\overline{\overline{y}}^{*}_{bl} - \overline{\overline{y}}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \omega\right)$	V^{∞}_{f}
Factor model	A, C'', E	$\sqrt{T}\left(\overline{\overline{y}}^{*}_{bl} - \overline{\overline{y}}\right) \underset{N,T\to\infty}{\overset{m.s.*}{\to}} 0$	

Double Resampling Bootstrap

$$\left(\overline{\overline{y}}^{**} - \overline{\overline{y}}\right) = \left(\overline{\mu}^{*} - \overline{\mu}\right) + \left(\overline{f}^{*}_{bl} - \overline{f}\right) + \left(\overline{\lambda}^{*}\overline{F}^{*}_{bl} - \overline{\lambda F}\right) + \left(\overline{\overline{\varepsilon}}^{**} - \overline{\overline{\varepsilon}}\right)$$

Cross-sectional One-way ECM

$$\sqrt{N}\left(\overline{\overline{y}}^{**} - \overline{\overline{y}}\right) = \sqrt{N}\left(\overline{\mu}^{*} - \overline{\mu}\right) + \sqrt{N}\left(\overline{\overline{\varepsilon}}^{**} - \overline{\overline{\varepsilon}}\right)$$

$$\sqrt{N}\left(\overline{\mu}^{*} - \overline{\mu}\right) \underset{N\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \sigma^2_{\mu}\right)$$

$$\left[\sqrt{N}\left(\overline{\overline{\varepsilon}}^{**} - \overline{\overline{\varepsilon}}\right)\right] \underset{N,T\to\infty}{\overset{m.s.*}{\longrightarrow}} 0 \quad \text{(Proposition 4)}$$

The result follows.

Two-way ECM

a) $\frac{N}{T} \to \delta \in [0, \infty)$:

$$\sqrt{N}\left(\overline{\overline{y}}^{**} - \overline{\overline{y}}\right) = \sqrt{N}\left(\overline{\mu}^{*} - \overline{\mu}\right) + \frac{\sqrt{N}}{\sqrt{T}}\sqrt{T}\left(\overline{f}^{*}_{bl} - \overline{f}\right) + \left[\sqrt{N}\left(\overline{\overline{\varepsilon}}^{**} - \overline{\overline{\varepsilon}}\right)\right]$$

$$\sqrt{N}\left(\overline{\mu}^{*} - \overline{\mu}\right) \underset{N\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \sigma^2_{\mu}\right)$$

$$\frac{\sqrt{N}}{\sqrt{T}}\sqrt{T}\left(\overline{f}^{*}_{bl} - \overline{f}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \delta.V^{\infty}_{f}\right)$$

$$\left[\sqrt{N}\left(\overline{\overline{\varepsilon}}^{**} - \overline{\overline{\varepsilon}}\right)\right] \underset{N,T\to\infty}{\overset{m.s.*}{\longrightarrow}} 0 \text{ (Proposition 4)}$$

thus

$$\sqrt{N}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right)\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\sigma_{\mu}^{2}+\delta.V_{f}^{\infty}\right)$$

b) $\frac{N}{T}\to\infty$:

$$\sqrt{T}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right)=\frac{\sqrt{T}}{\sqrt{N}}\left[\sqrt{N}\left(\overline{\mu}^{*}-\overline{\mu}\right)\right]+\sqrt{T}\left(\overline{f}_{bl}^{*}-\overline{f}\right)+\frac{\sqrt{T}}{\sqrt{N}}\left(\frac{1}{\sqrt{T}}\left[\sqrt{NT}\left(\overline{\overline{\varepsilon}}^{**}-\overline{\overline{\varepsilon}}\right)\right]\right)$$

$$\sqrt{T}\left(\overline{f}^{*}-\overline{f}\right)\underset{T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,V_{f}^{\infty}\right)$$

$$\frac{1}{\sqrt{N}}\left[\sqrt{NT}\left(\overline{\overline{\varepsilon}}^{**}-\overline{\overline{\varepsilon}}\right)\right]\underset{N,T\to\infty}{\overset{m.s.*}{\longrightarrow}}0$$

$$;\quad\frac{\sqrt{T}}{\sqrt{N}}\left[\sqrt{N}\left(\overline{\mu}^{*}-\overline{\mu}\right)\right]\underset{N,T\to\infty}{\overset{m.s.*}{\longrightarrow}}0$$

The result follows.

Factor Model

$$\sqrt{N}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right)=\sqrt{N}\left(\overline{\mu}^{*}-\overline{\mu}\right)+\sqrt{N}\left(\overline{\lambda}^{*}\overline{F}_{bl}^{*}-\overline{\lambda F}\right)+\left[\sqrt{N}\left(\overline{\overline{\varepsilon}}^{**}-\overline{\overline{\varepsilon}}\right)\right]$$

$$\sqrt{N}\left(\overline{\mu}^{*}-\overline{\mu}\right)\underset{N\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\sigma_{\mu}^{2}\right)$$

$$\left[\sqrt{N}\left(\overline{\overline{\varepsilon}}^{**}-\overline{\overline{\varepsilon}}\right)\right]\underset{N,T\to\infty}{\overset{m.s.*}{\longrightarrow}}0;\qquad\sqrt{N}\left(\overline{\lambda}^{*}\overline{F}_{bl}^{*}-\overline{\lambda F}\right)\underset{N,T\to\infty}{\overset{m.s.*}{\longrightarrow}}0$$

then

$$\sqrt{N}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right)\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\sigma_{\mu}^{2}\right)$$

Table 6 summarizes the asymptotic distributions of $\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right)$ for the different panel model specifications.

Table 6 : Asymptotic distributions with double resampling bootstrap

$Model$	$Assumptions$	$Asymptotic\ distribution$	ω
$Cross-sect.$ $One-way$ ECM	A, C''	$\sqrt{N}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N(0,\omega)$	σ_μ^2
$Temporal$ $One-way$ ECM	B, C''	$\sqrt{T}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N(0,\omega)$	V_f^∞
$Two-way$ ECM	A, B, C'', D	$\sqrt{N}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right) \underset{\substack{N,T\to\infty \\ \frac{N}{T}\to\delta\in[0,\infty)}}{\overset{*}{\Longrightarrow}} N(0,\omega)$	$\sigma_\mu^2 + \delta.V_f^\infty$
		$\sqrt{T}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right) \underset{\substack{N,T\to\infty \\ \frac{N}{T}\to\infty}}{\overset{*}{\Longrightarrow}} N(0,\omega)$	V_f^∞
$Factor$ $model$	A, C'', E	$\sqrt{N}\left(\overline{\overline{y}}^{**}-\overline{\overline{y}}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N(0,\omega)$	σ_μ^2

Chapitre 2

Bootstrap for panel regression models with random effects

Abstract

This paper considers bootstrap methods for panel linear regression models with random effects. Three kinds of regressors are considered : individual characteristics, temporal characteristics and regressors varying among periods and cross-section units. Using a two-way error component model, ordinary least squares estimator and the residual based bootstrap, it is shown that the double resampling that combines cross-sectional and temporal resampling is valid for the whole vector of parameters, under general conditions on cross-sectional and temporal heterogeneity as well as cross-sectional dependence. On the other hand, resampling only in the cross section dimension is only valid for the coefficients associated with individual characteristics, while block resampling only in the time series dimension is only valid for the coefficients associated with temporal characteristics. The bootstrap does not require the researcher to choose one of several asymptotic approximations available for panel models. Simulations confirm these theoretical results.

JEL Classification : C15, C23.

Keywords : Bootstrap, Panel Data Models.

2.1 Introduction

The purpose of this paper is to develop bootstrap methods for panel linear regression models and to prove their validity. The parameter of interest is the vector of the coefficients in a linear regression model β. Three different methods are considered : the first takes into account the cross-section dimension, the second use the time dimension and the third combines the two previous.

Because of their two dimensions, 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. These data have traditionally been used in many applied microeconomic fields such as labour economics and public finance, but more recently, the analysis of macro-level panel data sets has become common. For an overview of panel data models, see for example Baltagi (2008) or Hsiao (2003).

The exact 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. For an overview of bootstrap methods for dependent data, see Lahiri (2003). Application of bootstrap methods to panel data is an embryonic research field.

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 (2010) 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 (2011) explores the validity of several resampling methods for panel data. The main result of that 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. Using a two-way error component model, ordinary least squares estimator and the residual based bootstrap, it is shown that the double resampling that combines cross-sectional and temporal resampling is valid for the whole vector of parameters, under general conditions on cross-sectional and temporal heterogeneity as well as cross-sectional dependence. On the other hand, resampling only in the cross section dimension is only valid for the coefficients associated with individual characteristics, while block resampling only in the time series dimension is only valid for the coefficients associated with temporal characteristics.

The paper is organized as follows. In the second section, the panel data model studied is presented : a two-way error component model. Section 3 presents three bootstrap methods for panel models : the first takes into account the cross-section dimension, the second use the time dimension and the third combines the two previous. The fourth section presents theoretical results and analyzes the validity of each resampling methods using a heuristic mimic analysis and the asymptotic consistency. The sixth section concludes. Proofs of propositions are given in the appendix.

2.2 Panel Data Models

Consider a panel linear model

$$y_{it} = Z_{it}\beta + \nu_{it}, i = 1, 2, ...N, t = 1, 2, ...T. \tag{2.1}$$

y_{it} is the cross-sectional $i's$ observation at period t. Three kinds of regressors are considered : cross-section varying variables V_i, time varying variables W_t and double dimension varying variables X_{it}. β is an unknown vector of parameters.

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

where we collect all the parameters in a vector

$$\underset{(K,1)}{\beta} = \begin{pmatrix} \underset{(1,1)}{\theta} \\ \underset{(K_1,1)}{\tau} \\ \underset{(K_2,1)}{\gamma} \\ \underset{(K_3,1)}{\zeta} \end{pmatrix} \tag{2.3}$$

It is convenient to represent panel data as a matrix. By convention, in this paper, 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.

$$\underset{(1,K)}{Z_{(it)}} = \left[\underset{(1,1)}{1} \quad \underset{(1,K_1)}{V_i} \quad \underset{(1,K_2)}{W_t} \quad \underset{(1,K_3)}{X_{it}} \right] \tag{2.4}$$

The$^{\sim}$ is used to denote vectors obtained pooling the elements of matrices.

$$\widetilde{Z} = \left[\underset{(NT,1)}{1} \quad \underset{(NT,K_1)}{\widetilde{V}} \quad \underset{(NT,K_2)}{\widetilde{W}} \quad \underset{(NT,K_3)}{\widetilde{X}} \right] \tag{2.5}$$

Subbar and upbar refer respectively to the average in the cross-section dimension and the temporal dimension.

$$\underset{(1,K)}{\overline{Z}_{(i)}} = \frac{1}{T}\sum_{t=1}^{T} Z_{(it)} \quad , \quad \underset{(1,K)}{\underline{Z}_{(t)}} = \frac{1}{N}\sum_{i=1}^{N} Z_{(it)} \tag{2.6}$$

Assumptions about ν_{it} define different panel data models. Assume the following decomposition

$$\nu_{it} = \mu_i + f_t + \lambda_i F_t + \varepsilon_{it} \tag{2.7}$$

(2.10) 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.

Assumption A (individual effects)

The individual effects μ_i are drawn independently across i from some distribution with mean 0 and variance σ_μ^2 where $0 < \sigma_\mu^2 < \infty$;

Assumption A requires the individual heterogeneities to be independent and identically distributed with finite variance. The assumption of a zero mean is an identification assumption as any non-zero mean could be subsumed into the overall mean θ. The *i.i.d.* assumption, strong for classical asymptotic distribution is however important for bootstrap validity because i.i.d. bootstrap will be used in the cross-sectional dimension.

Assumption B (time effects)

$\{f_t\}$ is a stationary and α-mixing process with mixing coefficients $\alpha(j)$; $E(f_t) = 0$ and $\{f_t\}$ verifies Ibragimov's assumptions, that is $\exists\ \delta \in (0, \infty)$ such that $E|f_t|^{2+\delta} < \infty$ and $\sum_{j=1}^{\infty} \alpha(j)^{\delta/(2+\delta)} < \infty$ with finite long-run variance $V_f^\infty = \sum_{h=-\infty}^{\infty} Cov(f_t, f_{t+h}) \in (0, \infty)$;

Assumption B imposes some conditions on the time-series heterogeneity of our panel data. In particular, it requires it to be generated from a stationary process and that the dependence between f_t and f_k vanishes sufficiently fast as the distance between them increases.

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

Assumptions C (idiosyncratic error)

C : The idiosyncratic error ε_{it} is drawn independently across i and over t from some distribution with mean 0 and variance σ_ε^2 where $0 < \sigma_\varepsilon^2 < \infty$; [2]

Assumption D (independence)

The two processes $(\mu_1, .., \mu_N)$ and $(f_1, ..., f_T)$ are independent.

Assumption D imposes independence between the vector of individual heterogeneities and the vector of temporal heterogeneities. It is essential that there is no dependence between the two types of heterogeneity because the double resampling bootstrap method we will present later would destroy any dependence between the two dimensions..

Assumptions E (factor)

E1 : The factor loadings λ_i are drawn independently across i from some distribution with mean 0 and variance σ_λ^2 where $0 < \sigma_\lambda^2 < \infty$;

E2 : The factors (F_t) are a stationary and α-mixing process with mean 0 verifying Ibragimov's assumptions.

E3 : The two processes $(\lambda_1, .., \lambda_N)$ and $(F_1, ..., F_T)$ are independent.

Assumptions E are about factor model. Assumption E1 requires the loadings in a factor model be independent and identically distributed with finite variance. Assumption E2 is similar to assumption B, but applied to the factors in a factor model. Assumption E3 imposes independence between the vector of loadings and the vector of factors in an factor model. The reason is similar to B.

[2] It must be possible to assume a weaker version of assumption C as in chapter one. It will be a part of our next research.

Assumptions F (regressors)

F1 : The regressors are strictly exogenous

F2 : $\frac{\widetilde{Z}'\widetilde{Z}}{NT} \underset{NT\to\infty}{\overset{P}{\longrightarrow}} \underset{(K,K)}{Q} > 0$

F3 : $\frac{\overline{Z}'\overline{Z}}{N} \underset{N\to\infty}{\overset{P}{\longrightarrow}} \overline{Q}$

F4 : $\frac{\underline{Z}'\underline{Z}}{T} \underset{T\to\infty}{\overset{P}{\longrightarrow}} \underline{Q}$

Assumption F1 : The choice of strictly exogenous regressors is motivated by the fact that this chapter address residuals based bootstrap methods. The resampling uses first step estimations residuals thus any correlation between the regressors and the error terms would be destroyed in case of non exogeneity. Assumptions F3-F4 ensure that the product of regressors converge to non stochastic matrices. These matrices will be useful to derive asymptotic distributions.

One difficulty with asymptotic theory for panel data is the assumption made on the size of N and T. Traditionally, because panel data was mostly used in microeconometrics with large cross-sectional dimension but short time dimension, the assumption was made that N was large (approaching infinity) but that T remained finite. Conversely, in multiple time series models, the asymptotic analysis typically assumes that the number of series N is small while the number of time series observations T is large. Of course, these two asymptotic scenarios lead to different approximations and one is left to wonder which one is most appropriate for a given application at hand.

Recently. the analysis of large macro-type panels where both dimensions are reasonably large has allowed both dimensions to diverge. Phillips and Moon (1999) have provided underpinnings for these asymptotic analyses and have defined different frameworks. A sequential limit is obtained when an index is fixed at first, and the other passes to infinity, to have intermediate result. Next, the final result is obtained by passing the fixed index to infinity. On the other hand, in a diagonal path limit*,* N and T pass to infinity along a specific path, for example $T = T(N)$ and $N \to \infty$. Finally, in a joint limit*,* N and T pass to infinity simultaneously. Sometimes, it is necessary to control the relative expansion rate of N and T. For equivalence conditions between sequential and joint limits, see Phillips and Moon (1999).

Again, in practice, when faced with a particular application, it is not always obvious how to choose among these multiple asymptotic distributions, which may very be different. One of the advantages of the bootstrap approach we are analyzing is that it avoids having to choose between these competing approximations.

Proposition 9 *: Asymptotic distribution*

1 - Assume that A - F hold. When $N, T \to \infty$*, with* $\frac{N}{T} \to \delta \in [0, \infty)$

$$\sqrt{N}\left(\widehat{\beta} - \beta\right) \Longrightarrow N\left(0, \sigma_\mu^2\left(Q^{-1}\overline{Q}Q^{-1}\right) + \delta.Q^{-1}\Omega_{f\underline{Z}}^{\infty}Q^{-1}\right) \tag{2.11}$$

where $\Omega_{f\underline{Z}}^{\infty} = \lim_{T\to\infty}\left[Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_t\right)\right]$

2 - Assume that A-F hold. When $N, T \to \infty$*, with* $\frac{N}{T} \to \infty$

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) \Longrightarrow N\left(0, Q^{-1}\Omega_{f\underline{Z}}^{\infty}Q^{-1}\right) \tag{2.12}$$

The asymptotic distribution depends on δ, the relative convergence rate between the two indexes N and T With this asymptotic distributions in hand, the next step is to present the bootstrap methods, the bootstrap estimators in order to analyze their validity.

2.3 Bootstrap Methods

Methodology

In this section, we present the bootstrap methods. From initial data (Y, Z), we 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 sub-vector of β. Then inference is possible only with parameters that are not cancelled 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 that it reduces drastically the number of observations. Inference becomes 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 their framework does not cover the specifications with 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. Thus our choice is to use an unbiased and consistent estimator the pooled OLS estimator. 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 $\widehat{u}_{it}$

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

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

The first step residuals are arranged in matrix U :

$$\underset{(N,T)}{U} = \begin{pmatrix} \widehat{u}_{11} & \widehat{u}_{12} & .... & \widehat{u}_{1T} \\ \widehat{u}_{21} & \widehat{u}_{22} & .... & \widehat{u}_{2T} \\ .... & .... & ...... & .... \\ \widehat{u}_{N1} & \widehat{u}_{N2} & .... & \widehat{u}_{NT} \end{pmatrix}$$

Step 2 : From U, use a resampling method to create pseudo-sample of residuals U^* and pseudo-values of the dependent variable y^*_{it} .

$$y^*_{it} = Z_{it}\widehat{\beta} + u^*_{it} \qquad or \qquad \widetilde{Y}^* = \widetilde{Z}\widehat{\beta} + \widetilde{U}^* \tag{3.3}$$

Run pooling regression with (Y^*, Z) to obtain the bootstrap estimate :

$$\widehat{\beta}^* = \left(\widetilde{Z}^{/}\widetilde{Z}\right)^{-1}\widetilde{Z}^{/}\widetilde{Y}^* \tag{3.4}$$

Step 3 : Repeat *step 2* 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 with P^*. The resampling methods used to compute pseudo-panel-data are exposed below.

Cross-sectional Resampling Bootstrap

For a $N \times T$ matrix Y, *cross-sectional resampling* constructs a new $N \times T$ matrix Y^* with rows obtained by resampling with replacement the rows of U. In other words, we resample the vectors of T observations for each individual. U^* takes the form :

$$\underset{(N,T)}{U^*} = \begin{pmatrix} u_{11}^* = \widehat{u}_{i_1 1} & u_{12}^* = \widehat{u}_{i_1 2} & \dots & u_{1T}^* = \widehat{u}_{i_1 T} \\ u_{21}^* = \widehat{u}_{i_2 1} & u_{22}^* = \widehat{u}_{i_2 2} & \dots & u_{2T}^* = \widehat{u}_{i_2 T} \\ \dots & \dots & .. & \dots \\ u_{N1}^* = \widehat{u}_{i_N 1} & u_{N2}^* = \widehat{u}_{i_N 2} & \dots & u_{NT}^* = \widehat{u}_{i_N T} \end{pmatrix}$$

where each of the indices $(i_1, i_2, ..., i_N)$ is obtained by *i.i.d.* drawing with replacement from $(1, 2, ..., N)$. As a consequence, conditionally on U, the rows of U^* are independent and identically distributed.

Block Resampling Bootstrap

It is an accommodation of block bootstrap methods designed for time series. The idea is to resample in the time dimension blocks of consecutive periods in order to capture temporal dependence. All the observations at each time period are kept together in the hope of preserving their dependence. In this chapter, the original data are not directly resampled. The resampling is about the residuals from a first step OLS regression. Paparoditis and Politis (2003) exposed theoretical results using residual-based block bootstrap (RBB) for unit root testing with time series. Paparoditis and Politis (2003) resamples the residuals using the moving block bootstrap (MBB) (Kunsch (1989), Liu and Singh (1992)). [3]

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. U^* takes the following form :

$$\underset{(N,T)}{U^*} = \begin{pmatrix} u^*_{11} = \widehat{u}_{1t_1} & u^*_{12} = \widehat{u}_{1t_2} & ... & u^*_{1T} = \widehat{u}_{1t_T} \\ u^*_{21} = \widehat{u}_{2t_1} & u^*_{22} = \widehat{u}_{2t_2} & ... & u^*_{2T} = \widehat{u}_{2t_T} \\ ... & ... & .. & ... \\ u^*_{N1} = \widehat{u}_{Nt_1} & u^*_{N2} = \widehat{u}_{Nt_2} & ... & u^*_{NT} = \widehat{u}_{Nt_T} \end{pmatrix}$$

where in the setup of CBB resampling, $(t_1, t_2, ., t_T)$ takes the form

$$\underbrace{\tau_1, \tau_1 + 1, .., \tau_1 + l - 1}_{block1} \ \underbrace{\tau_2, \tau_2 + 1, .., \tau_2 + l - 1}_{block2}, ..., \underbrace{\tau_{[T/l]}, \tau_{[T/l]} + 1, .., \tau_{[T/l]} + l - 1}_{block[T/l]}$$

where the vector of indices $\left(\tau_1, \tau_2, ..., \tau_{[T/l]}\right)$ is obtained by $i.i.d.$ drawing with replacement from $(1, 2, ..., T)$, l denoting the block length. The name Circular come from the fact that when $\tau_t > T - l,$ the index of some observations exceed T and are replace using the rule : $T + t \leftrightarrow t$, as if the original data are around a circle and after T we continue with the first observation $t = 1$.

Double Resampling Bootstrap

This method is a combination of the two previous resampling methods. The term *double* comes from the fact that the resampling can be made in

[3] Non-overlapping block bootstrap (NMB) (Carlstein (1986)), circular block bootstrap (CBB) (Politis and Romano (1992)) and stationary block bootstrap (SB) (Politis and Romano (1994)) can also be adapted to panel data.

two steps. In a first step, one dimension is taken into account : from U, an intermediate matrix U^* is obtained either by cross-sectional resampling or block resampling. It turns out that the resampling is symmetric so it does not matter which dimension is resampled first. Then, another resampling is made in the second dimension : from U^* the final matrix U^{**}[4] is obtained. If we resampled in the cross-sectional dimension in the first step, then we resampled columns of the intermediate matrix in order to get our resampled matrix U^{**}.

Carvajal (2000) and Kapetanios (2008) have both suggested this double resampling in the special case where the block length is 1. They also analyze this resampling method by Monte Carlo simulations but give no theoretical support. The idea is that by drawing in one dimension, we preserve the dependence in that dimension in the first step. In the second step, we reproduce the properties in the other dimension by preserving the vectors drawn in the first step. U^{**} takes the following form :

$$\underset{(N,T)}{U^{**}} = \begin{pmatrix} u_{11}^{**} = \widehat{u}_{i_1 t_1} & u_{12}^{**} = \widehat{u}_{i_1 t_2} & \dots & u_{1T}^{**} = \widehat{u}_{i_1 t_T} \\ u_{21}^{**} = \widehat{u}_{i_2 t_1} & u_{22}^{**} = \widehat{u}_{i_2 t_2} & \dots & u_{2T}^{**} = \widehat{u}_{i_2 t_T} \\ \dots & \dots & .. & \dots \\ u_{N1}^{**} = \widehat{u}_{i_N t_1} & u_{N2}^{**} = \widehat{u}_{i_N t_2} & \dots & u_{NT}^{**} = \widehat{u}_{i_N t_T} \end{pmatrix}$$

where the indices $(i_1, i_2, ..., i_N)$ and $(t_1, t_2, ., t_T)$ are chosen as described in the the two previous sub-sections.

[4]We will use double asterisks** denote the quantities induced by double resampling.

Bootstrap Confidence Interval

In the literature, there are several bootstrap confidence intervals. In this chapter we study the percentile confidence interval [5]., constructed as following :

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 :

$$R_k^*(x) = \frac{1}{B+1} \sum_{b=1}^{B+1} I\left(r_k^{b*} \leq x\right) \tag{3.5}$$

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

$$CI_{1-\alpha,k}^* = \left[\widehat{\beta}_k - r_{k,1-\alpha/2}^*; \widehat{\beta}_k - r_{k,\alpha/2}^*\right] \tag{3.6}$$

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 $\left\{\widehat{\beta}_k^{*b}\right\}_{b=1..B+1}$.

The next section analyzes the validity of the bootstrap methods exposed above.

---

[5] In this chapter we do not address the percentile-t interval issue. The complexity to found a valid bootstrap variance to studentize the test statictic comes from the fact that the regressors are not resampled, avoiding to generalize easely some theoretical results founded in Chapter 1, specificallly with the block bootstrap.

2.4 Theoretical Results

This section presents theoretical results about resampling methods exposed in section 3.

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 much 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.

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

$$U^*_{cros} = [\mu]^*_{cros} + [f] + [\lambda]^*_{cros}[F] + [\varepsilon]^*_{cros} \tag{4.2}$$

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

$$U^{**} = [\mu]^*_{cros} + [f]^*_{bl} + [\lambda]^*_{cros}[F]^*_{bl} + [\varepsilon]^{**} \tag{4.4}$$

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 strength of the double resampling is thus to replicate the behavior of the temporal and cross-sectional components of the error terms without having to separate them and then induces a good inference. This analysis is heuristic and is about the validity of the resampling methods for the whole vector of the parameters. The next sub-section presents the analysis the asymptotic validity of the resampling methods for the vector whole vector of the parameters and some specific sub-vectors of parameters.

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. Because of multiple asymptotic distributions, there are several consistency definitions. A bootstrap method is said to be *consistent* for β if :

$$\sup_{x\in\mathbb{R}^K}\left|P^*\left(\sqrt{M}\left(\widehat{\beta}^* - \widehat{\beta}\right)\leq x\right) - P\left(\sqrt{M}\left(\widehat{\beta}-\beta\right)\leq x\right)\right| \underset{NT\to\infty}{\overset{P}{\to}} 0 \quad (4.5)$$

where $M \in \{N, T\}$

Definition *4.5* is 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.*

Using the asymptotic distributions exposed in section 2, the consistency of the resampling methods presented above is demonstrated when the asymptotic distributions of the bootstrap estimators are identical to their classic counterfactual. The following proposition analyzes the double resampling bootstrap.

Proposition 10 *: Consistency of the double resampling bootstrap*

Assume that A -F hold and $l \to \infty$ *such that* $\frac{l}{\sqrt{T}} \to 0$ *as* $T \to \infty$*, then :*

$$\sup_{x\in\mathbb{R}^K}\left|P^{**}\left(\sqrt{N}\left(\widehat{\beta}^{**} - \widehat{\beta}\right)\leq x\right) - P\left(\sqrt{N}\left(\widehat{\beta}-\beta\right)\leq x\right)\right| \underset{\substack{N,T\to\infty \\ \frac{N}{T}\to\delta\in[0,\infty)}}{\overset{P}{\to}} 0 \quad (4.6)$$

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

Proposition 10 is a extension of the result of the previous chapter to linear regression model with random effect in the two dimensions. The double resampling takes into account the two dimensions of the panels and thus induce a correct inference for all the vector of the parameters β without having to impose a restriction or the convergence of N and T, except what is necessary to have classical asymptotic distribution.

As a consequence of Chapter 1, resampling in one dimension would lead to invalid inference about the vectors of parameters β. However the next two propositions analyze respectively the validity of the cross-sectional resampling bootstrap and block resampling bootstrap for components of β, even if the inference is not valid for the whole vector.

To do this, let's introduce new assumptions about the regressors.

Assumptions F' (regressors)

F1' : The regressors are strictly exogenous and the means of stochastic regressors converge to zero.

F2' : $\frac{\tilde{Z}'\tilde{Z}}{NT} \underset{NT\to\infty}{\overset{P}{\to}} \underset{(K,K)}{Q} > 0$

F3' : $\frac{\overline{Z}'\overline{Z}}{N} \underset{N\to\infty}{\overset{P}{\to}} \overline{Q}$

F4 ' : $\frac{\underline{Z}'\underline{Z}}{T} \underset{T\to\infty}{\overset{P}{\to}} \underline{Q}$

F1' imposes the means of stochastic regressors to converge to zero inducing a particular variance covariance in that the asymptotic distributions of temporal (resp. individual) characteristic coefficients are independent of individuals effects (time effects). The zero mean assumption hold in finite sample when fixed effects are remove by centering, or using Frisch-Waugh-Lovell theorem to center the regressors and the dependent variable, or center the stochastic regressors only and maintain the constant. Assumptions F2', F3' and F4' are identical to F2, F3 and F4.

Proposition 11 *: Consistency of the cross-sectional bootstrap*

Assume that A - F' hold. When $N, T \to \infty$, the cross-sectional bootstrap is consistent for the sub-vector of the parameters associated with regressors varying only by individuals

$$\sup_{x \in \mathbb{R}^{K1}} \left| P^* \left(\sqrt{N} \left(\widehat{\tau}^*_{cros} - \widehat{\tau} \right) \leq x \right) - P \left(\sqrt{N} \left(\widehat{\tau} - \tau \right) \leq x \right) \right| \underset{N,T \to \infty}{\overset{P}{\to}} 0 \qquad (4.8)$$

The presence of the temporal heterogeneity f_t induces the inconsistency of the cross-sectional resampling bootstrap for β when N and T goes to infinity. However, a sub-vector analysis shows that the asymptotic behavior of the parameter τ depends on the process μ_i, whose behavior is correctly replicated by the cross-sectional resampling, inducing a correct inference for the individual characteristics even if it is not correct for all the coefficients. In practice, Proposition 11 allows the use of the cross-sectional bootstrap when the inference is only about the parameters associated with individual characteristics. The next proposition analyzes the consistency of the temporal block resampling.

Proposition 12 *: Consistency of the block resampling bootstrap*

Assume that A - F' hold and $l \to \infty$ such that $\frac{l}{\sqrt{T}} \to 0$ as $T \to \infty$, then :

$$\sup_{x \in \mathbb{R}^{K2}} \left| P^* \left(\sqrt{T} \left(\widehat{\gamma}^*_{bl} - \widehat{\gamma} \right) \leq x \right) - P \left(\sqrt{T} \left(\widehat{\gamma} - \gamma \right) \leq x \right) \right| \underset{N,T \to \infty}{\overset{P}{\to}} 0 \qquad (4.8)$$

This proposition means that the block resampling bootstrap is consistent for sub-vector of the parameters associated with time varying regressors. As shown by the mimic analysis, block bootstrap resampling bootstrap does not replicate the behavior of the cross-sectional heterogeneity μ_i, inducing its inconsistency for β when N and T goes to infinity. However, a sub-vector analysis shows that the asymptotic behavior of the parameter τ depends on the process f_t , whose behavior is correctly replicated by the temporal block resampling, inducing a correct inference for the temporal characteristics even if it is not correct for all the coefficients. In practice, the the previous proposition allows the use of the block bootstrap when the inference is only about the parameters associated with temporal characteristics.

The consistency of bootstrap methods as defined in *(4.5)* implies the validity of percentile confidence interval to make inference. With all the theoretical results in hand, in the next section we will see the behavior of the bootstrap methods in finite sample, using simulations.

2.5 Simulations

This section presents results from a small simulation experiment to illustrate our theoretical results. The data generating process is *(2.2)* and *(2.10)*. The individual effects are standard normal and independent across units :

$$\mu_i \sim i.i.d.N\left(0,1\right),$$

while both the time effect and common factor are $AR\left(1\right)$ process with parameter 0.25 :

$$\begin{aligned} f_t &= \rho f_{t-1} + \varpi_t \\ F_t &= \rho F_{t-1} + \eta_t \\ \eta_t, \varpi_t &\sim i.i.d.N\left(0, \left(1-\rho^2\right)\right) \\ \rho &= 0.25 \end{aligned}$$

The factor loadings are standard normal :

$$\lambda_i \sim i.i.d.N\left(0,1\right)$$

and the idiosyncratic errors are also standard normal :

$$\varepsilon_{it} \sim i.i.d.N\left(0,1\right).$$

Each component of β=$(\theta, \tau, \gamma, \zeta)'$ is equal 1. The regressors are generated standard normal :

$$\begin{aligned} V_i &\sim i.i.d.N\left(0,1\right) \\ W_t &\sim i.i.d.N\left(0,1\right) \\ X_{it} &\sim i.i.d.N\left(0,1\right) \end{aligned}$$

Four panel dimensions are considered : $(N,T) = (10,10)\,,(20,20)\,,(30,30)\,,$ and $(50,50)$. Temporal resampling is carried out with the Circular Block Bootstrap (CBB) with block length l = 2, 2, 3, 5 for $T = 10, 20, 30$ and 50. For each bootstrap resampling scheme, B is equal to 999 and the number of simulations is 1000.

Tables 1 and 2 gives rejection rates for a two-tailed test for the null hypothesis that $\beta_k = 0$ at nominal level 5%.

Several specifications are considered combining different processes in *(2.10)* : 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 two-way ECM with spatial dependence $(\mu_i + f_t + \lambda_i F_t + \varepsilon_{it})$.

The analysis of the simulations results can be made at two levels. The first one is about the sub-model specifications when the two dimensions are similar. The fact that the failure on the resampling methods in only one dimension is due to the non-replication of the heterogeneity on the other dimension. Thus when there is one kind of heterogeneity, the resampling in only this specific dimension produces good inferences. 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 second level is about sub-vector analysis. With the two-way ECM, the resampling in one dimension is not valid for all the component of β when N and T have similar sizes. However, the first parts of Proposition 2 and 3 affirms that the cross-sectional resampling and the temporal block resampling allows to have a good inference with respectively τ and γ. The results *(in italics)* are near the theoretical five percents for the coefficient associated with individual characteristic when the cross-sectional resampling is used. The same thing is observed with the coefficient associated with the temporal characteristic using the block resampling.

Table 1 : Simulations results with percentile interval

			$(N;T)$	$=$	$(10;10)$	$(N;T)$	$=$	$(20;20)$
			Cros.	Bloc.	D-Res	Cros.	Bloc.	D-Res
Cross	1	θ	11.6	63.1	**8.5**	6.0	69.7	**5.5**
1-way	V_i	τ	12.2	63.8	**9.0**	*8.1*	69.1	**7.5**
ECM	W_t	γ	8.6	13.9	**1.8**	6.7	*6.6*	**0.4**
$\mu_i + \varepsilon_{it}$	X_{it}	ζ	5.8	34.8	**4.2**	6.3	37.5	**5.8**
Temp.	1	θ	67.9	23.0	**9.0**	75.8	16.1	**8.7**
1-way	V_i	τ	12.4	13.1	**2.0**	*7.7*	8.8	**1.0**
ECM	W_t	γ	57.6	12.7	**8.6**	68.1	*8.6*	**7.5**
$f_t + \varepsilon_{it}$	X_{it}	ζ	32.7	8.0	**6.5**	35.3	6.0	**5.6**
	1	θ	31.4	34.4	**9.4**	25.2	24.4	**8.9**
2-way	V_i	τ	12.6	59.4	**6.0**	*8.1*	67.5	**6.9**
ECM	W_t	γ	58.5	12.2	**9.9**	67.3	*7.8*	**7.2**
$\mu_i + f_t + \varepsilon_{it}$	X_{it}	ζ	26.3	28.7	**7.0**	26.4	27.1	**5.5**
2-way ECM	1	θ	27.0	35.8	**9.9**	23.8	25.4	**8.5**
with spatial	V_i	τ	12.7	53.8	**9.5**	*7.8*	59.8	**6.4**
dependence	W_t	γ	45.9	11.0	**6.8**	60.8	*8.6*	**6.7**
$\mu_i + f_t + \lambda_i F_t + \varepsilon_{it}$	X_{it}	ζ	18.4	24.1	**5.5**	21.4	19.8	**5.7**

In bold the rejection rates with the double resampling bootstrap.

In italics the rejection rates with valid subvector inference

Table 2 : Simulations results with percentile interval

			$(N;T)$	$=$	$(30;30)$	$(N;T)$	$=$	$(50;50)$
			Cros.	Bloc.	D-Res	Cros.	Bloc.	D-Res
Cross	1	θ	7.4	75.2	**6.1**	5.8	79.4	**5.6**
1-way	V_i	τ	*6.1*	75.9	**6.0**	*5.5*	80.4	**5.5**
ECM	W_t	γ	7.1	*7.8*	**0.7**	6.6	*6.1*	**0.4**
$\mu_i + \varepsilon_{it}$	X_{it}	ζ	5.5	37.4	**5.7**	4.9	36.0	**4.4**
Temp.	1	θ	77.9	9.1	**8.1**	83.5	7.0	**6.5**
1-way	V_i	τ	*7.5*	9.4	**1.1**	*5.8*	8.7	**0.8**
ECM	W_t	γ	73.3	*7.0*	**6.5**	78.7	*6.2*	**5.9**
$f_t + \varepsilon_{it}$	X_{it}	ζ	36.7	5.7	**5.4**	36.9	5.2	**4.9**
	1	θ	26.0	23.8	**6.5**	24.3	20.5	**6.0**
2-way	V_i	τ	*8.7*	73.8	**5.2**	*5.5*	81.6	**5.5**
ECM	W_t	γ	73.8	*7.3*	**4.7**	78.2	*5.2*	**5.6**
$\mu_i + f_t + \varepsilon_{it}$	X_{it}	ζ	24.4	28.2	**5.8**	24.8	25.3	**5.4**
2-way ECM	1	θ	24.2	22.5	**6.7**	22.6	20.9	**6.0**
with spatial	V_i	τ	*6.8*	65.2	**6.0**	*6.0*	73.2	**5.8**
dependence	W_t	γ	68.6	*7.4*	**5.9**	77.3	*5.2*	**4.7**
$\mu_i + f_t + \lambda_i F_t + \varepsilon_{it}$	X_{it}	ζ	20.5	21.2	**5.5**	19.5	20.4	**4.9**

In bold the rejection rates with the double resampling bootstrap.

In italics the rejection rates with valid subvector inference.

2.6 Conclusion

This paper considers the issue of bootstrap methods for panel data models. Three bootstrap methods are explored : cross-sectional bootstrap, temporal block bootstrap and double resampling bootstrap. The cross-sectional bootstrap resamples the cross-section units. The temporal block bootstrap resamples blocks of consecutive time periods for all the cross-section units. The double resampling bootstrap combines the two previous methods. It is shown that the presence of temporal heterogeneity invalids the use of the cross-sectional bootstrap for the make inference about the whole coefficient of the model. The reason of this failure is that the resampling in only the cross-sectional dimension, treat the processes in the time dimension as constants and does not replicate their behavior. However, an appropriate analysis of the covariance matrix shows that the asymptotic distribution of the coefficients associated with individual characteristics, does not depends to the temporal heterogeneity. The cross-sectional bootstrap is then valid for the sub-vector of the parameters associated with temporal regressors. The temporal block resampling only in the time dimension fails because of the presence of the cross-sectional heterogeneity. However, the inference about the coefficients associated with temporal characteristics is valid using the temporal block bootstrap. The double resampling bootstrap replicates simultaneously the behavior of the cross-sectional and temporal processes without having to separate them. This property induces the validity of the double resampling bootstrap for all the coefficients of the model. The implementation of bootstrap methods for panel models does not explicitly take into account how N and T goes to infinity, avoiding the multiple asymptotics problem that sometimes arises with large panel models.

APPENDIX

Proof of Proposition 9

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

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

$$\begin{aligned}Var\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\varepsilon_{it}\mid Z\right) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Var\left(Z_{(it)}^{/}\varepsilon_{it}\right)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}Z_{(it)}Var\left(\varepsilon_{it}\right) \\ &= \frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\sigma_\varepsilon^2 \underset{N,T\to\infty}{\to} Q\sigma_\varepsilon^2 \\ Var\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\varepsilon_{it}\right) &: \underset{N,T\to\infty}{\to} Q\sigma_\varepsilon^2\end{aligned}$$

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

$$Var\left(\frac{1}{\sqrt{T}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\lambda_i F_t \mid Z\right) = Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\lambda_i\overline{F} \mid Z\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=i}^{N}Cov\left(\overline{Z}_{(i)}^{/}\lambda_i\overline{F},\overline{Z}_{(j)}^{/}\lambda_j\overline{F} \mid Z\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=i}^{N}\overline{Z}_{(i)}^{/}Cov\left(\lambda_i\overline{F},\lambda_j\overline{F}\right)\overline{Z}_{(j)}^{/}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=i}^{N}\overline{Z}_{(i)}^{/}E\left(\lambda_i\lambda_j\overline{F}^2\right)\overline{Z}_{(j)}^{/}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}Var\left(\lambda_i\right)Var\left(\overline{F}\right)\overline{Z}_{(j)}^{/}$$

$$= \left(\frac{\overline{Z}^{'}\overline{Z}}{N}\sigma_\lambda^2\right)\left(Var\left(\overline{F}\right)\right)$$

$\frac{\overline{Z}^{'}\overline{Z}}{N}\sigma_\lambda^2 \underset{N\to\infty}{\overset{P}{\to}} \overline{Q}\sigma_\lambda^2$.

$Var\left(\overline{F}\right) \underset{T\to\infty}{\to} 0$ because $V_F^\infty \sum_{h=-\infty}^{\infty} Cov\left(F_t,F_{t+h}\right) \in (0,\infty)$

thus $\left(\frac{1}{\sqrt{T}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\lambda_i F_t\right) \underset{N,T\to\infty}{\overset{m.s.}{\to}} 0$

Applying Lindeberg-Feller CLT to $\left(\overline{Z}_{(1)}^{/}\mu_1,\overline{Z}_{(2)}^{/}\mu_2,...,\overline{Z}_{(N)}^{/}\mu_N\right)$ gives

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\mu_i \underset{N\to\infty}{\Longrightarrow} N\left(0,\sigma_\mu^2\overline{Q}\right).$$

A multivariate case of Ibragimov(1962) applied to $\left(\underline{Z}_{(1)}^{/}f_1,\underline{Z}_{(2)}^{/}f_2,...,\underline{Z}_{(T)}^{/}f_T,\right)$ gives

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_t \underset{T\to\infty}{\Longrightarrow} N\left(0,\Omega_{f\underline{Z}}^\infty\right)$$

The previous results allow to write :

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

$2-$ When $\frac{N}{T} \to \infty$ the scaling factor of $\left(\widehat{\beta}-\beta\right)$ is $\sqrt{T}$.

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

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\mu_i \underset{N\to\infty}{\Longrightarrow} N\left(0,\sigma_\mu^2\overline{Q}\right) \text{ and } \frac{\sqrt{T}}{\sqrt{N}} \to 0 \text{ then } \frac{\sqrt{T}}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\mu_i\right) \underset{N,T\to\infty}{\overset{m.s.}{\to}} 0$$

$$\frac{1}{\sqrt{N}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\lambda_i F_t = \frac{\sqrt{T}}{\sqrt{N}}\left(\frac{1}{\sqrt{T}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\lambda_i F_t\right)$$

$$\left(\frac{1}{\sqrt{T}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\lambda_i F_t\right) \underset{N,T\to\infty}{\overset{m.s.}{\to}} 0 \text{ and } \frac{\sqrt{T}}{\sqrt{N}} \to 0 \text{ thus } \frac{1}{\sqrt{N}}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}Z_{(it)}^{/}\lambda_i F_t \underset{N,T\to\infty}{\overset{m.s.}{\to}} 0$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_t \underset{T\to\infty}{\Longrightarrow} N\left(0,\Omega_{f\underline{Z}}^{\infty}\right)$$

The previous results allow to write :

$$\sqrt{T}\left(\widehat{\beta}-\beta\right) \Longrightarrow N\left(Q^{-1}\Omega_{f\underline{Z}}^{\infty}Q^{-1}\right)$$

Proof Proposition 10

$$\sqrt{N}\left(\widehat{\beta}^{**}-\widehat{\beta}\right) = \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\begin{array}{c}\underbrace{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}\left(\mu_i^{*}\right)}_{\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\sigma_\mu^2\overline{Q}\right)}+\frac{\sqrt{N}}{\sqrt{T}}\underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}\left(f_{bl,t}^{*}\right)}_{\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\Omega^{\infty}_{f\underline{Z}}\right)} \\ +\underbrace{\sqrt{N}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z^{/}_{(it)}e_{it}^{**}}_{\underset{N,T\to 0}{\overset{P*}{\to}}0}\end{array}\right]$$

$$\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0,\sigma_\mu^2Q^{-1}\overline{Q}Q^{-1}+\delta Q^{-1}\Omega^{\infty}_{\underline{Z}^{/}f}Q^{-1}\right)$$

By Lemma 1,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}\left(\mu_i^{*}\right)\underset{N\to\infty}{\overset{*}{\Longrightarrow}} N\left(0,\sigma_\mu^2\overline{Q}\right)$$

By Lemma 2,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}\left(f_{bl,t}^{*}\right)\underset{T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0,\Omega^{\infty}_{f\underline{Z}}\right)$$

First case : When $N,T\to\infty$ with $\frac{N}{T}\to\delta\in[0,\infty)$, $\sqrt{N}\left(\widehat{\beta}^{**}-\widehat{\beta}\right)\overset{*}{\Longrightarrow} N\left(0,\sigma_\mu^2Q^{-1}\overline{Q}Q^{-1}+\delta Q^{-1}\Omega^{\infty}_{\underline{Z}^{/}f}Q^{-1}\right)$.

Second case : $N,T\to\infty$ with $\frac{N}{T}\to\infty$, $\sqrt{T}\left(\widehat{\beta}^{**}-\widehat{\beta}\right)\overset{*}{\Longrightarrow} N\left(0,Q^{-1}\Omega^{\infty}_{\underline{Z}^{/}f}Q^{-1}\right)$

The consistency follows.

Matrix Notations

Under Assumption F1'

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

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

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

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

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

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

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

Lemma 1

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\left(\mu_i^*\right)\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\sigma_\mu^2\overline{Q}\right)$$

Proof of Lemma 1

Conditionally on Z and Y, $\left(\overline{Z}_{(1)}^{/}\left(\mu_1^*\right),\overline{Z}_{(2)}^{/}\left(\mu_2^*\right),.....\overline{Z}_{(N)}^{/}\left(\mu_N^*\right)\right)$ are independent with respect of the cross-section resampling. $Var^*\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\left(\mu_i^*\right)\right]=\frac{1}{N}Var^*\left[\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\left(\mu_i^*\right)\right]=\frac{1}{N}\left[\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}Var^*\left(\mu_i^*\right)\overline{Z}_{(i)}\right]=\frac{1}{N}\left[\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\overline{Z}_{(i)}\right]Var^*\left(\mu_i^*\right)=\frac{\overline{Z}^{/}\overline{Z}}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\mu}_i^2\right)\underset{N\to\infty}{\overset{P}{\to}}\sigma_\mu^2\overline{Q}$ because $\frac{\overline{Z}^{/}\overline{Z}}{N}\underset{N\to\infty}{\to}\overline{Q}$ and $\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{\mu}_i^2\right)\underset{N\to\infty}{\overset{P}{\to}}\sigma_\mu^2$.

The application of Lindeberg-Feller CLT gives : $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}_{(i)}^{/}\left(\mu_i^*\right)\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}}N\left(0,\sigma_\mu^2\overline{Q}\right)$.

Lemma 2

Under assumption A2, if $l\to\infty$ such that $\frac{l}{\sqrt{T}}\to 0$ as $T\to\infty$, then

$$\sup_{x\in\mathbb{R}^{K2}}\left|P^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_{bl,t}^*\leq x\right)-P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_t\leq x\right)\right|\underset{T\to\infty}{\overset{P}{\to}}0$$

Proof of Lemma 2

The residual based block bootstrap of Paparoditis and Politis (2003) gives

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_{bl,t}^*\leq x\right)-P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t\leq x\right)\right|\underset{T\to\infty}{\overset{P}{\to}}0$$

Under assumption strict exogeneity assumption, and conditionally to the regressors, the previous result implies :

$$\sup_{x\in\mathbb{R}^{K2}}\left|P^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_{bl,t}^*\leq x\right)-P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}_{(t)}^{/}f_t\leq x\right)\right|\underset{T\to\infty}{\overset{P}{\to}}0$$

Proof of Proposition 11

Let's define $e_{it} = \lambda_i F_t + \varepsilon_{it}$ and $\widehat{e}_{it} = \widehat{\lambda}_i \widehat{F}_t + \widehat{\varepsilon}_{it}$

$$\sqrt{N}\left(\widehat{\beta}^{*}_{cros} - \widehat{\beta}\right) = \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\underbrace{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}\left(\mu^{*}_{i}\right)}_{\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0,\sigma^2_\mu \overline{Q}\right)} + \underbrace{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}\left(\overline{e}^{*}_{i.}\right)}_{\underset{N,T\to\infty}{\overset{m.s.^*}{\to}} 0} + \underbrace{\sqrt{N}\frac{1}{T}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}\widehat{f}_t}_{=0}\right]$$

$$\begin{aligned} Var^{*}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}\left(\overline{e}^{*}_{i.}\right)\right) &= \frac{1}{N}\left[\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}Var^{*}\left(\overline{e}^{*}_{i.}\right)\overline{Z}_{(i)}\right] = \frac{1}{N}\left[\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}\overline{Z}_{(i)}\right]Var^{*}\left(\overline{e}^{*}_{i.}\right) \\ &= \frac{\overline{Z}^{\prime}\overline{Z}}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\overline{\widehat{e}}_{i.}\right)^{2}\right) \underset{N,T\to\infty}{\overset{P}{\to}} 0 \end{aligned}$$

Then $\frac{1}{\sqrt{N}}\sum\limits_{i=1}^{N}\overline{Z}^{/}_{(i)}\left(\overline{e}^{*}_{i.}\right) \underset{N,T\to\infty}{\overset{m.s.^*}{\to}} 0$.

$\frac{1}{\sqrt{N}}\sum\limits_{i=1}^{N}\overline{Z}^{/}_{(i)}\left(\mu^{*}_{i}\right) \underset{N\to\infty}{\overset{*}{\Longrightarrow}} N\left(0,\sigma^2_\mu\overline{Q}\right)$ by Lemma 1. $\frac{1}{T}\sum\limits_{t=1}^{T}\underline{Z}^{/}_{(t)}\widehat{f}_t = 0$ by the properties of the decomposition of the residuals. Then

$$\sqrt{N}\left(\widehat{\beta}^{*}_{cros} - \widehat{\beta}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0,\sigma^2_\mu Q^{-1}\overline{Q}Q^{-1}\right)$$

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

$$\underset{(K,1)}{\beta} = \begin{pmatrix} \underset{(1,1)}{\theta} \\ \underset{(K1,1)}{\tau} \\ \underset{(K2,1)}{\gamma} \\ \underset{(K3,1)}{\zeta} \end{pmatrix}$$

The sub-matrix of $\underline{Q}$ associated with τ is a block of zeros then the asymptotic distribution of τ depends only on $\sigma_\mu^2 Q^{-1}\overline{Q}Q^{-1}$ and we have.

$$\sqrt{N}\left(\widehat{\tau}-\tau\right) \underset{N,T\to\infty}{\Longrightarrow} N\left(0, \sigma_\mu^2 Q_V^{-1}\overline{Q}_V Q_V^{-1}\right)$$

A sub-vector convergence for $\left(\widehat{\beta}^*_{cros}-\widehat{\beta}\right)$ gives

$$\sqrt{N}\left(\widehat{\tau}^*_{cros}-\widehat{\tau}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0, \sigma_\mu^2 Q_V^{-1}\overline{Q}_V Q_V^{-1}\right)$$

The result of consistency follows.

Proof of Proposition 12

$$\sqrt{T}\left(\widehat{\beta}^*_{bl}-\widehat{\beta}\right) = \left(\frac{\widetilde{Z}^{/}\widetilde{Z}}{NT}\right)^{-1}\left[\underbrace{\sqrt{T}\frac{1}{N}\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}\widehat{\mu}_i}_{=0} + \underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}\left(f^*_{bl,t}\right)}_{\underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0,\Omega^{\infty}_{f\underline{Z}}\right)} + \underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}\left(e^*_{bl,.t}\right)}_{\underset{N,T\to 0}{\overset{ms^*}{\to}} 0}\right]$$

$\sqrt{T}\frac{1}{N}\sum_{i=1}^{N}\overline{Z}^{/}_{(i)}\widehat{\mu}_i$ by the properties of the decomposition of the residuals.
$Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}\left(e^*_{bl,.t}\right)\right) \underset{N,T\to\infty}{\overset{P}{\to}} 0$ then $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}\left(e^*_{bl,.t}\right) \underset{N,T\to 0}{\overset{ms^*}{\to}} 0$.
By Lemma 2 $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}\left(f^*_{bl,t}\right) \underset{T\to\infty}{\overset{*}{\Longrightarrow}} N\left(0,\Omega^{\infty}_{f\underline{Z}}\right)$, thus

$$\sqrt{T}\left(\widehat{\beta}^{*}_{bl}-\widehat{\beta}\right)\underset{N,T\rightarrow\infty}{\overset{*}{\Longrightarrow}}N\left(Q^{-1}\Omega^{\infty}_{f\underline{Z}}Q^{-1}\right)$$

(3.1).

$$\underline{Z}^{/}_{(t)}=\left[\begin{array}{cccc}\underset{(1,1)}{1} & \underset{(0,K_1)}{0} & \underset{(1,K_2)}{W_t} & \underset{(1,K_3)}{\underline{X}_t}\end{array}\right]$$

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}f_t &= \left[\begin{array}{cccc}\underset{(1,1)}{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_t} & \underset{(1,K_1)}{\overline{V}} & \underset{(1,K_2)}{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}W_tf_t} & \underset{(1,K_3)}{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{X}_tf_t}\end{array}\right]\\ &= \left[\begin{array}{cccc}A & \underset{(1,K_1)}{D} & \underset{(1,K_2)}{B} & \underset{(1,K_3)}{C}\end{array}\right]\end{aligned}$$

$$\begin{aligned}Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underline{Z}^{/}_{(t)}f_t\right) &= \left[\begin{array}{cccc}\underset{(1,1)}{Var\left(A\right)} & \underset{(K1,K_1)}{Cov(A,D)} & Cov(A,B) & Cov(A,C)\\ \underset{(K1,K_1)}{Cov(D,A)} & \underset{(K1,K_1)}{Var(D)} & \underset{(K1,K_1)}{Cov(D,B)} & \underset{(K1,K_1)}{Cov(D,C)}\\ Cov\left(B,A\right) & \underset{(K1,K_1)}{Cov(B,D)} & \underset{(K_2,K_2)}{Var\left(B\right)} & Cov\left(B,C\right)\\ Cov\left(C,A\right) & \underset{(K1,K_1)}{Cov(C,D)} & Cov\left(C,B\right) & \underset{(K_3,K_3)}{Var\left(C\right)}\end{array}\right]\\ &\rightarrow\ \Omega^{\infty}_{f\underline{Z}}=\left[\begin{array}{cccc}1 & 0 & \underline{Q}_{XW} & \underline{Q}_{X1}\\ 0 & 0 & 0 & 0\\ \underline{Q}_{1W} & 0 & \underline{Q}_{W} & \underline{Q}_{WX}\\ \underline{Q}_{1X} & 0 & \underline{Q}_{W1} & \underline{Q}_{X}\end{array}\right]\end{aligned}$$

The sub-matrix of $\overline{Q}$ associated with γ is a block of zeros then the asymptotic distribution of γ depends only on $\Omega^{\infty}_{f\underline{Z}}$ and we have.

$$\sqrt{T}\left(\widehat{\gamma}-\gamma\right)\underset{N,T\rightarrow\infty}{\Longrightarrow}N\left(Q^{-1}_{W}\underline{Q}_{W1}Q^{-1}_{W}\right)$$

A subvector convergence for $\left(\widehat{\beta}^*_{bl} - \widehat{\beta}\right)$ gives

$$\sqrt{T}\left(\widehat{\gamma}^*_{bl} - \widehat{\gamma}\right) \underset{N,T\to\infty}{\overset{*}{\Longrightarrow}} N\left(Q_W^{-1}\underline{Q}_{W1}Q_W^{-1}\right)$$

The result of consistency follows.

Chapitre 3

Bootstrapping differences-in-differences estimator

Abstract

This paper re-examines the analysis of differences-in-differences estimators by Bertrand, Duflo and Mullainathan (2004). Their empirical application uses panel data from the Current Population Survey on wages of women in the 50 states. Placebo laws are generated at the state level, and the authors measure their impact on wages. By construction, no impact should be found. Bertrand, Duflo and Mullainathan (2004) show that neglected heterogeneity and temporal correlation lead to spurious findings of an effect of the Placebo laws. The double resampling bootstrap method taking into account the temporal and the cross-section dimension of the panel dataset, corrects these size distortions very well and gives more reliable evaluation of public policies.

JEL Classification : C15, C23, C21

Keywords : Bootstrap, Panel data models, Differences-in-differences estimates, Evaluation of Public Policies

3.1 Introduction

Since the work of Ashenfelter and Card (1985), differences-in-differences (DD) estimation methods are commonly used to evaluate the effect of an treatment or intervention such as a change in policy [1]. The basic set up is the case with two groups and two periods. One group (the treated group) is exposed to a treatment in the second period but not the first one. The second group (the control group) is not exposed to the treatment. The impact of this treatment is evaluated by comparing changes in the response of the treated with the changes among the control group. This basic framework can be easily extended to several time periods and heterogeneity through the introduction of covariates.

This paper explores the application of bootstrap methods to make more accurate inference in DD estimation. It is motivated by results in Bertrand, Duflo and Mullainathan (2004), henceforth BDM, who document inference problems with the use of standard OLS inference. Their evidence comes from panel data from the Current Population Survey on wages of women in the 50 states. Placebo laws are generated at the state level, and the authors measure their impact on wages. By construction, no impact should be found. They show that neglected heterogeneity and temporal correlation lead to spurious findings of an effect of the Placebo laws. Typically, instead of the theoretical 5% rejection rate, simulations with the OLS fixed effect estimator lead to 45% rejection rate. This size distortion means that many evaluations will incorrectly conclude that the analyzed public policy has an effect when it has no impact in reality.

[1] For an overview of impact evaluation methods, see for example Shahidur, Koolwal, and Hussain (2009).

To correct these problems, we suggest using bootstrap methods. Application of the bootstrap in this context is complicated by the presence of possible cross-sectional dependence, serial correlation and neglected heterogeneity. We propose using the double resampling method developed in the previous chapter that is robust to some forms of these difficulties to generate bootstrap samples. Our simulation results suggest that the size distortions reported in BDM are corrected to a large degree by our methods.

The paper is organized as follows. The next section presents differences-in-differences estimation. Section 3 presents bootstrap methods for panel data and presents conditions for their validity. Section 4 revisits the empirical exercise of BDM, while the section 5 concludes.

3.2 Differences-in-differences Estimation

The simplest setup of DD estimation is the case with two groups and two periods. One group (treatment group) is exposed to a treatment (or public policy) in a second period not the first one. The second group (control) is not exposed to the treatment. Using the second group as a control group, the basic idea is to evaluate the impact of this treatment. The model is :

$$y = \alpha_0 + \alpha_1 I_2 + \delta I + u$$

where y is the outcome of interest, I_2 is a dummy variable for the second time period which captures changes in the outcome even in the absence of the treatment. I is a binary program indicator, which is unity if unit i is affected by the public policy (treatment). The parameter of interest is δ. The impact of the treatment is defined as :

$$DD = E(y_{T,2} - y_{T,1}) - E(y_{U,2} - y_{U,1})$$

where $y_{T,2}$ (resp. $y_{T,1}$) is the value of y for the treatment group in the second period (resp. first period). Similar definitions are associated to the control group when T is replaced by U. The DD estimator is :

$$\widehat{DD} = \hat{\delta} = (\overline{y}_{T,2} - \overline{y}_{T,1}) - (\overline{y}_{U,2} - \overline{y}_{U,1})$$

where $\overline{y}_{T,2}$ (resp. $\overline{y}_{T,1}$) is the mean of y for the treatment group at the second period (resp. first period). The impact of the treatment is then evaluated by the change in the treated group $(\overline{y}_{T,2} - \overline{y}_{T,1})$ minus the change in the untreated group $(\overline{y}_{U,2} - \overline{y}_{U,1})$ which is considered as the evolution of y that is not induced by the treatment[2]. The name differences-in-differences is due to this double difference. The DD estimator is exactly the OLS estimator δ in the model (2.1) : this analogy is useful in the general case.

In a more general setup, there are several periods for the two groups and covariates. Then model (2.1) becomes :

$$y_{it} = \alpha X_{it} + \delta I_{it} + u_{it} = Z_{it}\beta + u_{it} \tag{2.4}$$

where y_{it} is the outcome of interest, X_{it} a range of covariates, I_{it} a binary program indicator, which is unity if unit i is affected by the public policy (treatment) at time t. The parameter of interest is δ which is assumed constant among units and periods. Using the analogy between the DD estimator and the OLS estimator, the impact of the treatment effect is evaluated by the OLS estimator of δ. The problem is not about pointwise estimation of δ because under general assumptions about the error term, $\hat{\delta}$ is unbiased and consistent. The main problem is about statistical inference, specially in samples of small or moderate sample as would be the case if the policy is

[2] This paper will not address the issue of possible selection bias. As in BMD, the paper will treat the possible bias in the estimation the confidence interval of δ

subject to a pilot-project. The next section presents bootstrap methods in order to have a correct inference.

3.3 Bootstrap Method

The basic idea of bootstrap methods consists in drawing many random samples that resemble the observed sample as much as possible and estimating the distribution of the object of interest over these random samples. The original bootstrap procedure has been proposed by Efron (1979) for statistical analysis of independent and identically distributed (i.i.d.) observations. 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 and more recently, the application of bootstrap methods to panel data models. In the setup of linear regression, there are two main approaches of bootstrap methods : the residual based bootstrap and the pair bootstrap.

3.3.1 Residual-based Bootstrap

The residual-based method uses a first step regression in order to obtain residuals which are used to create the pseudo data and the pseudo values of the parameter of interest. In details, this method can be summarized in the followings steps.

Step 1 : Run the pooled regression to obtain the OLS estimator $\hat{\beta}$ and the residuals $\hat{u}_{it}$

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

$$\hat{u}_{it} = y_{it} - Z_{it}\hat{\beta} \qquad or \qquad U = Y - Z\hat{\beta} \tag{3.2}$$

Step 2 : From the matrix of residuals U, use a resampling method to create pseudo-sample of residuals U^* and pseudo-values of the dependent variable :

$$y_{it}^* = Z_{it}\hat{\beta} + u_{it}^* \qquad or \qquad Y^* = Z\hat{\beta} + U^* \tag{3.3}$$

Run pooling regression with (Y^*, Z)

$$\hat{\beta}^* = \left(Z^{/}Z\right)^{-1} Z^{/}Y^* \tag{3.4}$$

Step 3 : Repeat *step 2* B times in order to have $B+1$ realizations of the vector of parameters $\hat{\beta}^*$ and thus the coefficient of interest $\hat{\delta}^*$: $\left\{\hat{\delta}_b^*\right\}_{b=1,2,\ldots,B+1}$. The empirical distribution of these $(B+1)$ realizations is used to make inference.

3.3.2 Pair bootstrap

The pair bootstrap method does not require a preliminary regression. The pseudo data are created directly from the real data. The term pair comes from the fact that to create the pseudo data, the methods resample the dependent variable and covariates jointly. In details, this method can be summarized in the following steps.

Step 1 : From the original data matrix Y, Z, use a resampling method to create pseudo-sample of regressors Z^* and pseudo-values of the dependent variable Y^* and run pooling regression with (Y^*, Z^*)

$$\hat{\beta}^* = \left(Z^{*/}Z^*\right)^{-1} Z^{*/}Y^* \tag{3.5}$$

Step 2 : Repeat *step 2* B times in order to have $B+1$ realizations of the vector of parameters $\hat{\beta}^*$ and thus the coefficient of interest $\hat{\delta}^*$: $\left\{\hat{\delta}_b^*\right\}_{b=1,2,\ldots,B+1}$.

3.3.3 Bootstrap Confidence Intervals

Two bootstrap confidence intervals are commonly used : the percentile confidence interval and the percentile-t confidence interval. ***A equal-tailed percentile confidence interval*** of level $(1-\alpha)$ for the parameter $\hat{\delta}$ is :

$$CI^*_{1-\alpha,\delta} = \left[\hat{\delta} - r^*_{1-\alpha/2} ; \hat{\delta} - r^*_{\alpha/2}\right] \tag{3.6}$$

where $r^*_{\alpha/2}$ and $r^*_{1-\alpha/2}$ are respectively the $\alpha/2$-percentile and $(1-\alpha/2)$-percentile of the empirical distribution of $\left\{r^{b*}_{\delta} = \hat{\delta}^{b*} - \hat{\delta}\right\}_{b=1,2,\ldots,B+1}$.

The construction of a percentile-t interval is based on a studentized statistic. In practice, with each pseudo-sample, compute the statistic :

$$t^*_b = \frac{\hat{\delta}^* - \hat{\delta}}{\sqrt{\widehat{V}ar^*\left(\hat{\delta}^*\right)}} \tag{3.7}$$

A equal-tailed percentile-t confidence interval of level $(1-\alpha)$ is

$$CI^*_{1-\alpha,\delta} = \left[\hat{\delta} - \sqrt{\widehat{V}ar\left(\hat{\delta}\right)}.t^*_{1-\frac{\alpha}{2}} ; \hat{\delta} - \sqrt{\widehat{V}ar\left(\hat{\delta}\right)}.t^*_{\frac{\alpha}{2}}\right] \tag{3.8}$$

where $t^*_{\alpha/2}$ and $t^*_{1-\alpha/2}$ are respectively the $\alpha/2$-percentile and $(1-\alpha/2)$-percentile of the empirical distribution of $\{t^*_b\}_{b=1,2,\ldots,B+1}$. The percentile-t confidence interval allow theoretical results for asymptotic refinements in some situations.

3.3.4 Panel Resampling Methods

The first bootstrap resampling methods have been developed for one dimension data : cross-section units, time series. The two dimension of panel data need appropriate resampling methods. It is practical to represent panel data as a matrix. By convention, 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 X of N rows and T columns, x_{it} *is the cross-sectional* $i's$ *observation at period* t.

$$\underset{(N,T)}{X} = \begin{pmatrix} x_{11} & x_{12} & ... & ... & x_{1T} \\ x_{21} & x_{22} & ... & ... & x_{2T} \\ ... & ... & ... & ... & ... \\ x_{N1} & x_{N2} & .. & ... & x_{NT} \end{pmatrix} \tag{3.9}$$

We will present three resampling methods designed for panel data using a generic matrix X ; these resampling methods exposed can be adapted to Y, Z or U depending of the bootstrap method (residual-based or pair).

Cross-sectional Resampling Bootstrap

It is an accommodation of the original i.i.d. bootstrap method to the rows of X. For a $N \times T$ matrix X, *cross-sectional resampling* is the operation of constructing a $N \times T$ matrix X^* with rows obtained by resampling with replacement rows of X. Conditionally on X, the rows of X^* are independent and identically distributed.

For a $N \times T$ matrix Y, *cross-sectional resampling* constructs a new $N \times T$ matrix Y^* with rows obtained by resampling with replacement the rows of Y. In other words, we resample the vectors of T observations for each individual. X^* takes the form :

$$\underset{(N,T)}{X^*} = \begin{pmatrix} x^*_{11} = x_{i_1 1} & y^*_{12} = y_{i_1 2} & ... & x^*_{1T} = x_{i_1 T} \\ x^*_{21} = x_{i_2 1} & x^*_{22} = x_{i_2 2} & ... & x^*_{2T} = x_{i_2 T} \\ ... & ... & .. & ... \\ x^*_{N1} = x_{i_N 1} & x^*_{N2} = x_{i_N 2} & ... & x^*_{NT} = x_{i_N T} \end{pmatrix} \tag{3.10}$$

where each of the indices $(i_1, i_2, ..., i_N)$ is obtained by *i.i.d.* drawing with replacement from $(1, 2, ..., N)$. As a consequence, conditionally on X, the rows of X^* are independent and identically distributed.

Block Resampling Bootstrap

This methods is a direct generalization 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 the time dimension blocks of consecutive periods in order to capture temporal dependence. All the observations at each time period are kept together in the hope of preserving their dependence.

The block bootstrap resampling constructs a new $N \times T$ matrix X^* with columns obtained by resampling with replacement blocks of columns of X. X^* takes the following form :

$$\underset{(N,T)}{X^*} = \begin{pmatrix} x^*_{11} = x_{1t_1} & x^*_{12} = x_{1t_2} & ... & x^*_{1T} = x_{1t_T} \\ x^*_{21} = x_{2t_1} & x^*_{22} = x_{2t_2} & ... & x^*_{2T} = x_{2t_T} \\ ... & ... & .. & ... \\ x^*_{N1} = x_{Nt_1} & x^*_{N2} = x_{Nt_2} & ... & x^*_{NT} = x_{Nt_T} \end{pmatrix} \tag{3.11}$$

The choice of $(t_1, t_2, .., t_T)$ depends on the which block bootstrap method is used in the time dimension. For example with the CBB bootstrap resampling, we have $(t_1, t_2, ., t_T)$ taking the form

$$\underbrace{\tau_1, \tau_1+1, .., \tau_1+l-1}_{block1} \; \underbrace{\tau_2, \tau_2+1, .., \tau_2+l-1}_{block2}, ..., \underbrace{\tau_{[T/l]}, \tau_{[T/l]}+1, .., \tau_{[T/l]}+l-1}_{block[T/l]}$$

where the vector of indices $\left(\tau_1, \tau_2, ..., \tau_{[T/l]}\right)$ is obtained by *i.i.d.* drawing with replacement from $(1, 2, ..., T-l)$, where the vector of indices $\left(\tau_1, \tau_2, ..., \tau_{[T/l]}\right)$ is obtained by *i.i.d.* drawing with replacement from $(1, 2, ..., T)$, l denoting the block length. The name Circular come from the fact that when $\tau_t > T-l$, the index of some observations exceed T and are replace using the rule : $T+t \longleftrightarrow t$, as if the original data are around a circle and after T we continue with the first observation $t=1$. Others block bootstrap methods can also be accommodated in the time dimension to panel data.

Double Resampling Bootstrap

This method is a combination of the two previous resampling methods. 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 X, an intermediate matrix X^* is obtained either by cross-sectional resampling or block resampling. It turns out that the resampling is symmetric so it does not matter which dimension is resampled first. Then, another resampling is made in the second dimension : from X^* the final matrix X^{**} is obtained. If we resampled in the cross-sectional dimension in the first step, then we resampled columns of the intermediate matrix in order to get our resampled matrix X^{**}[3].

Carvajal (2000) and Kapetanios (2008) have both suggested this double resampling in the special case where the block length is 1. They also analyze this resampling method by Monte Carlo simulations but give no theoretical support. The idea is that by drawing in one dimension, we preserve the dependence in that dimension in the first step. In the second step, we reproduce

[3] We will use double asterisks** denote the quantities induced by double resampling.

the properties in the other dimension by preserving the vectors drawn in the first step. X^{**} takes the following form :

$$\underset{(N,T)}{X^{**}} = \begin{pmatrix} x_{11}^{**} = x_{i_1 t_1} & x_{12}^{**} = x_{i_1 t_2} & ... & x_{1T}^{**} = x_{i_1 t_T} \\ x_{21}^{**} = x_{i_2 t_1} & x_{22}^{**} = x_{i_2 t_2} & ... & x_{2T}^{**} = x_{i_2 t_T} \\ ... & ... & .. & ... \\ x_{N1}^{**} = x_{i_N t_1} & x_{N2}^{**} = x_{i_N t_2} & ... & x_{NT}^{**} = x_{i_N t_T} \end{pmatrix} \tag{3.12}$$

where the indices $(i_1, i_2, ..., i_N)$ and $(t_1, t_2, ., t_T)$ are chosen as described in *(3.10)* and *(3.11)*.

Bootstrap Methods Validity

Let's consider the validity of bootstrap methods exposed above for good inference about the vector of parameters β in general and coefficient of the impact of the treatment effect δ in particular. To justify to validity of a bootstrap percentile confidence interval, it must be shown that the asymptotic distribution of $(\hat{\delta} - \delta)$ coincides with the asymptotic distribution of its bootstrap counterpart $(\hat{\delta}^* - \hat{\delta})$. The validity of the percentile-t interval is proved in a similar way. It must be shown that the behavior of $(\hat{\delta} - \delta)/\sqrt{\widehat{V}ar\left(\hat{\delta}\right)}$ asymptotically equivalent to the behavior of its bootstrap analog $(\hat{\delta}^* - \hat{\delta})/\sqrt{\widehat{V}ar^*\left(\hat{\delta}^*\right)}$.

In the literature, there are many applications of the bootstrap with panel data, but several are carried out without rigorous theoretical justification. Recently some theoretical papers appeared in the literature. Kapetanios (2008) presents theoretical results about panel regression models N goes to infinity, under the assumption that cross-sectional vectors of regressors and errors terms are i.i.d. Hounkannounon (2011) shows that the double resampling bootstrap is valid in the presence of some forms of temporal and or cross-sectional random heterogeneity and cross-sectional dependence. It also shows that the presence of temporal random heterogeneity leads to invalid inference using cross-sectional resampling bootstrap and the presence of cross-sectional random heterogeneity leads to incorrect inference using the temporal resampling bootstrap. The double resampling bootstrap is valid in presence of both temporal and or cross-sectional random heterogeneity. Gonçalves (2010) explores the accommodation of the moving blocks bootstrap to panel linear model with individual fixed effects.

3.4 Empirical Application

3.4.1 Specification

This section re-examines the differences-in-differences estimates exercise of BDM (2004). Their empirical application uses data from the Current Population Survey (CPS) on wages of women between 25 and 50 in the fourth month of the Merged Outgoing Rotation Group for years 1979 to 1999, in the 50 American States. Placebo laws are generated at the state level, and the authors measure their impact on wages. Formally, consider the model

$$Y_{ist} = A_s + B_t + cX_{ist} + \delta I_{st} + \varepsilon_{ist} \qquad (4.1)$$

where A_s and B_t are effects for states and years respectively, I_{st} a dummy for whether the intervention has affected group s (state) at period t (year) and Y_{ist} the outcome (wage) for individual i in group s by time t. X_{ist} are individual controls including four education dummies (less than high school, high school, some college and college or more) and a quartic in age. Model (4.1) is called a multilevel (or hierarchical) linear model because of the presence of three indexes. In order to have a panel dataset, log weekly earnings are first regressed on the individual controls X_{ist}. A panel is constructed with mean of these residuals by state and year.

$$\overline{Y}_{st} = \alpha_s + \gamma_t + \delta I_{st} + \varepsilon_{st} \tag{4.2}$$

By construction, $\overline{Y}_{st}$ presents the evolution of wages that do not depend on the education level and age. Figure 1 presents the evolution of $\overline{Y}_{st}$ and is presented for a subsample of states. The relative position of the aggregate wage among states seems stable (state heterogeneity) and the temporal evolution of the series are similar (temporal heterogeneity). To take into account these heterogeneities, two main approaches are used : fixed-effect vs random effects.

In the setup of model (4.2), fixed-effect models assumed that $\{\alpha_s\}_{s=1,2,..,N}$ and $\{\gamma_t\}_{t=1,2,..,T}$ belong to the space of interest parameters. The random effect models treat $\{\alpha_s\}_{s=1,2,..,N}$ and $\{\gamma_t\}_{t=1,2,..,T}$ as random variables like ε_{st} with exogeneity conditions. It is also possible to assume heterogeneities fixed in one dimension and random in the other. In the model (4.2), taking for each state the time average, we have $\overline{Y}_{s.} = \alpha_s + \overline{\gamma} + \delta \overline{I}_{s.} + \overline{\varepsilon}_{s.}$, taking for each time period the average across states, we have $\overline{Y}_{.t} = \overline{\alpha} + \gamma_t + \delta \overline{I}_{.t} + \overline{\varepsilon}_{.t}$. Finally the overall mean is defined as $\overline{\overline{Y}} = \overline{\alpha} + \overline{\gamma} + \delta \overline{\overline{I}} + \overline{\overline{\varepsilon}}$. By Frisch-Waugh-Lovell theorem, the OLS estimation of δ in the model (4.2) assuming fixed effects in time and cross-section dimensions is exactly the OLS estimation of δ with the transformed model :

$$(\overline{Y}_{st} - \overline{Y}_{s.} - \overline{Y}_{.t} + \overline{\overline{Y}}) = \delta(I_{st} - \overline{I}_{s.} - \overline{I}_{.t} + \overline{\overline{I}}) + (\varepsilon_{st} - \overline{\varepsilon}_{s.} - \overline{\varepsilon}_{.t} + \overline{\overline{\varepsilon}}) \tag{4.3}$$

Assuming only fixed effect in the time dimension, defining $u_{st} = \alpha_s + \varepsilon_{st}$ as the new error term, the appropriate transformed model is :

$$(\overline{Y}_{st} - \overline{Y}_{.t}) = \delta(I_{st} - \overline{I}_{.t}) + (u_{st} - \overline{u}_{.t}) \tag{4.4}$$

The specifications *(4.3)* and *(4.4)* will be useful to remove fixed effects and time effects from the original data.

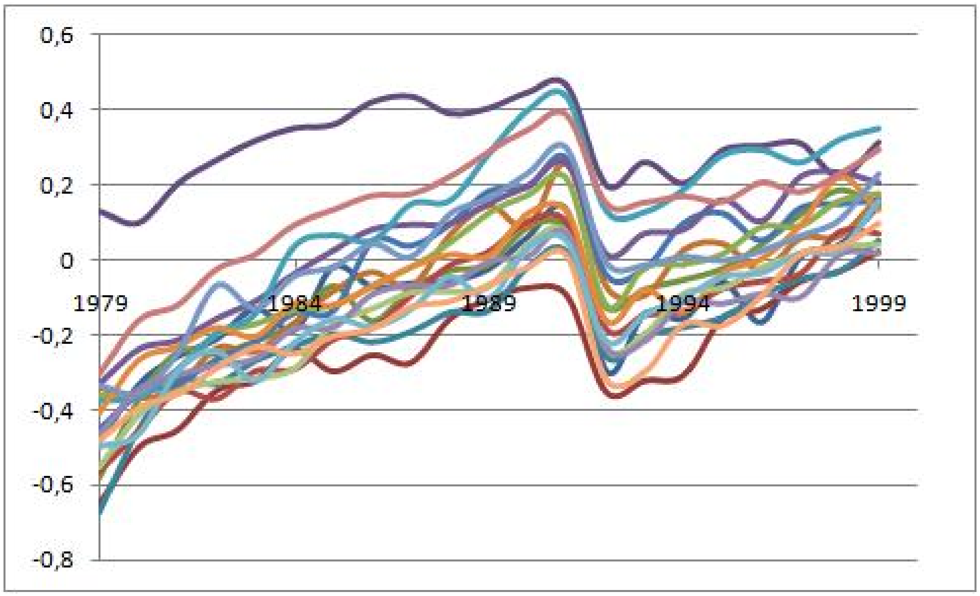

FIG. 3.1 – Time Evolution of Wage by State

3.4.2 Placebo Laws

At each simulation, the selection of the treatment group states is similar : half of the states are randomly chosen to form the treatment group. A passage date is randomly chosen (uniformly drawn between 1985 and 1995) identically for each state in the treatment group : from the passage date all states in the treatment group are assumed affected by the public intervention for all the remaining periods.

3.4.3 Simulation Results

By construction of the placebo laws, no impact is expected, i.e. the true value of δ is zero. Simulations with a correct inference method to construct confidence interval of level 95 % will reject the null hypothesis ($H0 : \delta = 0$) approximately 5 % times. $H0$ is rejected if the true value 0 doesn't belong to the confidence interval. Five inference methods are considered. The first is the BDM fixed effect OLS with usual estimation of the standard error (BDM-OLS). The second is BDM *block bootstrap* method[4] (BDM-BSP). The third is Pair Bootstrap with resampling in the cross-section dimension and percentile-t confidence interval applied to transformed model (4.3) (extraction of fixed effects) [5] (Pair-BSP). This bootstrap method is modified version of the bootstrap method called *block bootstrap* in BDM. [6] The fourth is the Residual based bootstrap with double resampling (block size=3, C.B.B. in time dimension) and percentile confidence interval, applied to transformed model (4.4) (extraction time effect) (D-Res-BSP-R). The fifth method is the Pair bootstrap with double resampling (block size=3, C.B.B. in time dimension) and percentile-t confidence interval, applied to transformed model (4.4) (extraction time effect) (D-Res-BSP-P).

Table 1 presents rejection rates based on 2000 simulations, and 999 bootstrap replications. The time periods are maintained constant (1979-1999)

[4]BDM *block bootstrap* method resamples States that is in our notation cross-sectional resampling bootstrap.

[5]The choice to extract the time fixed effect instead of random is due to a positive trend visible in Figure 1, implying a non-stationary process.

[6]BDM uses the usual OLS estimate standard error in the bootstrap world. The problem with this methodology is the contradiction between the resampling method and the estimation of $Var^*(\hat{\delta}^*)$. The resampling method assumes that the states are i.i.d. and keeps the time dependence but the estimation of the variance ignores the temporal correlation.

while the number of States takes the values 6,10, 20 and 50 and the states are selected as presented in the subsection Placebo laws.

Table 1 : Simulations Results

States	BDM-OLS	BDM-BSP	Pair-BSP	D-Res-BSP-R	D-Res-BSP-P
6	48.0	43.5	17.1	15.0	4.9
10	38.5	22.5	13.3	9.6	5.3
20	38.5	13.5	8.1	6.3	5.1
50	43.0	6.5	6.5	5.1	5.1

Standard OLS fixed effect estimator clearly over-rejects the null hypothesis. In practice, that means that researchers conclude that a public policy has an impact, when in fact, there is no impact. The bad performance of the commonly used OLS fixed effect estimator is due to serial correlation remaining in the specification (4.3). The correlation in the original data is not completely eliminated by the extraction the time effect. The BDM bootstrap method performs better than fixed effect OLS, but the results are only acceptable when N=50. The disappointing performance of the BDM bootstrap method comes from a bad specification of the variance in the bootstrap world. BDM uses the usual OLS estimate standard error in the bootstrap world. The problem with this methodology is the contradiction between the resampling method and the estimation of $Var^*(\hat{\delta}^*)$. The resampling method assumes that the states are i.i.d. and keeps the time dependence but the estimation of the variance ignores the temporal correlation. A modification of BDM bootstrap method, using a correct variance estimator, gives better results, what is visible comparing columns BDM-BSP and Pair-BSP. The two last columns present results with the double resampling bootstrap : residual-based and pair bootstrap. In all the cases, the double resampling bootstrap performances are the best with a advantage the pair version. This advantage is probably due to the use of a percentile-t confidence interval while a percentile confidence interval is used in the residual based method of the double resampling bootstrap.

3.5 Conclusion

The evaluation of public policies when data are available for the outcome during several periods, leads to panel models. The necessity to have very quickly some information about the effect of the treatment implies restrictions in the time dimension. Sometimes, the public program can be implemented with a test sample before a large application, thus a moderate number of individuals is also involved. This double restriction in the cross-section and time dimension gives to the researcher in charge of this evaluation very moderate sample size dataset. Despite theses restrictions, the researcher has the obligation to do his best to evaluate properly the potential impact of the treatment. For this purpose, bootstrap methods for linear panel models can be useful. In this paper we give a justification of the disappointing performance of BDM bootstrap method carried-out without theoretical justification. We compare a modified version BDM bootstrap and the double resampling bootstrap method based on resampling in time and cross-section dimensions. Simulation results with time dependent placebo laws, these bootstrap methods correct very well size distortions in moderate size samples. The double resampling bootstrap outperforms the other methods and corrects size distortions even in small samples. In practice, DD method can use when data are available before and after the public intervention. Other methodologies are available for impact evaluation when the information is available only after the intervention (matching estimator, propensity matching estimator,...). The theoretical justification of resampling methods for inference with these impact evaluation methods is a research field to explore.

Conclusion générale

Dans cette thèse nous fournissons les fondements théoriques des méthodes de bootstrap appliquées aux données de panel. Comme il est courant dans la littérature, nous avons commencé par un estimateur de la moyenne en postulant dans le premier chapitre, un modèle avec un seul paramètre. Nous montrons les méthodes de rééchantillonnage qui tiennent compte seulement dans une dimension échoue à répliquer le comportement du processus générateur de données dans la seconde dimension. Ainsi, lorsque qu'il y a de l'hétérogénéité aléatoire dans la dimension individuelle(resp. temporelle), la méthode réchantillonnage dans la dimension temporelle (resp. individuelle) seulement, échoue à produire une inférence valide. Nous proposons la méthode du bootstrap de double rééchantillonnage qui tient compte des deux dimensions du panel et avec laquelle on obtient des résultats valides là où les autres échouent.

Le second chapitre étend le premier à un modèle de regression linéaire de panel. En utilisant un modèle à erreurs composées doubles, l'estimateur des moindres carrés ordinaires et la méthode de bootstrap des résidus, on montre que le rééchantillonnage dans la seule dimension individuelle est valide pour l'inférence sur les coefficients associés aux régresseurs qui changent uniquement par individu. Le rééchantillonnage dans la dimension temporelle est valide seulement pour le sous vecteur des paramètres associé aux régresseurs qui évoluent uniquement dans le temps. Le double rééchantillonnage est quand à lui est valide pour faire de l'inférence pour tout le vecteur des paramètres.

Le troisième chapitre re-examine l'exercice de l'estimateur des doubles différences de Bertrand, Duflo et Mullainathan (2004). L'exercice empirique utilise des données de panel provenant du Current Population Survey sur le salaire des femmes dans les 50 états des Etats-Unis d'Amérique de 1979 à 1999. Des variables de pseudo-interventions publiques au niveau des états sont générées et on s'attend à ce que les tests arrivent à la conclusion qu'il n'y a pas d'effet de ces politiques placebos sur le salaire des femmes. Bertrand, Duflo et Mullainathan (2004) montre que la non-prise en compte de la dépendance temporelle entraîne d'importantes distorsions de niveau de test lorsqu'on évalue l'impact de politiques publiques en utilisant des données de panel. La méthode de double rééchantillonnage développée dans cette thèse permet de corriger le problème de niveau de test et donc d'évaluer correctement l'impact des politiques publiques.

Les perspectives de recherche pour l'avenir peuvent être empiriques ou théoriques. Dans les simulations et les applications empiriques, le choix de la longueur du bloc du double resampling bootstrap est arbitraire. Une méthodologie de choix optimal de choix reste à développer. Une piste à explorer serait d'adapter les méthodes de choix de bloc optimal, développés pour les séries temporelles, au pseudo échantillon intermédiaire obtenue après application du bootstrap i.i.d. dans la dimension individuelle. D'un autre côté, les méthodologies bootstrap ont été validées au premier degré. Le recours à des expansions aux ordres supérieurs permettra d'apporter des preuves de raffinements asymptotiques. Sur le plan pratique, la méthodologie bootstrap pourrait être utilisée pour évaluer de réelles politiques publiques lorsqu'on a des observations plusieurs périodes, avant et après la mise en place de la politique.

Bibliographie

[1] Andersson, M. K. and Karlsson, S. (2001). Bootstrapping Error Component Models. *Computational Statistics 16*, 2, 221-231.

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

[3] Ashenfelter O. and Card D. (1985) Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs, *The Review of Economics and Statistics*, MIT Press, vol. 67(4), 648-60

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

[5] 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.

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

[7] Bellman, L., Breitung, J. and Wagner, J. (1989). Bias Correction and Bootstrapping of Error Component Models for Panel Data : Theory and Applications. *Empirical Economics 14*, 329-342.

[8] Bickel, P. J. and Freedman, D. A. (1981). Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics 9*, 6, 1196-1217.

[9] Bertrand M., Duflo E. and Mullainathan S. (2004) How much should we trust differences-in-differences estimates? *The Quartely Journal of Economics 119*, 1, 249-275.

[10] 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.

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

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

[13] 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.

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

[15] Everaert, G. and Pozzi L. (2007). Bootstrap-Based Bias Correction for Dynamic Panels. *Journal of Economic Dynamics & Control 31*, 1160-1184.

[16] Focarelli, D. (2005). Bootstrap Bias-Correction Procedure in Estimating Long-run Relationships from Dynamic Panels, with an Application to Money Demand in the Euro Area. *Economic Modelling 22*, 305–325.

[17] Freedman, D.A. (1981). Bootstrapping Regression Models. *The Annals of Statistics 9*, 6, 1218-1228.

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

[19] Gonçalves S. (2010) The Moving Blocks Bootstrap for Panel Regression Models with Individual Fixed Effects, *Forthcoming in Econometric Theory.*

[20] Herwartz, H. (2007). Testing for Random Effects in Panel Models with Spatially Correlated Disturbances. *Statistica Neerlandica 61*, 4, 466–487.

[21] Herwartz, H. (2006). Testing for Random Effects in Panel Data under Cross-sectional Error Correlation – A Bootstrap Approach to the Breusch Pagan Test. *Computational Statistics and Data Analysis 50*, 3567–3591.

[22] Hounkannounon B. (2011). Double Resampling Bootstrap for the Mean of a Panel, Mimeo, Department of Economics, Université de Montréal.

[23] Hounkannounon B. (2011) Bootstrap for Panel Regression Models with Random Effects, Mimeo, Department of Economics, Université de Montréal.

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

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

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

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

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

[29] 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.

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

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

[32] Paparoditis and Politis (2003) Residual-based Block Bootstrap fou Unit Root Testing. *Econometrica 71*, 3, 813-855.

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

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

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

[36] 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.

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

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

[39] Shahidur R., Koolwal G. and Hussain S. (2009) Handbook on Impact Evaluation : Quantitatives Methods and Pratices, World Bank Publications.

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

[41] Singh, K. (1981). On the Asymptotic Accuracy of Efron's Bootstrap. *The Annals of Statistics 9*, 6, 1187-1195.