Department of Economics

A Bootstrap Procedure for Panel Datasets with Many Cross-Sectional Units

George Kapetanios

Working Paper No. 523 October 2004 ISSN 1473-0278

A Bootstrap Procedure for Panel Datasets with Many Cross-Sectional Units

G. Kapetanios*
Queen Mary, University of London

September 28, 2004

Abstract

This paper considers the issue of bootstrap resampling in panel datasets. The availability of datasets with large temporal and cross sectional dimensions suggests the possibility of new resampling schemes. We suggest one possibility which has not been widely explored in the literature. It amounts to constructing bootstrap samples by resampling whole cross sectional units with replacement. In cases where the data do not exhibit cross sectional dependence but exhibit temporal dependence, such a resampling scheme is of great interest as it allows the application of i.i.d. bootstrap resampling rather than block bootstrap resampling. It is well known that the former enables superior approximation to distributions of statistics compared to the latter. We prove that the bootstrap based on cross sectional resampling provides asymptotic refinements. A Monte Carlo study illustrates the superior properties of the new resampling scheme compared to the block bootstrap.

Keywords: Bootstrap, Panel Data

JEL Codes: C32, C33

1 Introduction

Panel datasets have been increasingly used in economics to analyse complex economic phenomena. One of the attractions of panel datasets is the ability to use an extended dataset to obtain information about parameters of interest which are assumed to have common values across panel units. The existing literature on panel data is huge and rapidly expanding. Good but inevitably somewhat partial reviews may be found, among others, in Baltagi (2001) and Hsiao (2003). Traditionally, panel analysis focussed on datasets with large cross sectional dimension (N) and smaller time series dimension (T). But more recently, and with the emergence of rich datasets both in N and T, focus rests on the theoretical analysis of

*Department of Economics, Queen Mary, University of London, Mile End Rd., London E1 4NS. Email: G.Kapetanios@qmul.ac.uk

large N - T datasets.

Inference in panel datasets has mainly used asymptotic approximations for the construction of test statistics and estimation of variances of estimators. The use of the bootstrap as an alternative to such asymptotic approximations has been considered but its properties have not received the same amount of attention as in the time series literature. Here, we need to note the well known fact that bootstrap methods can provide better approximations to the exact distributions of various statistics compared to asymptotic approximations leading to the conclusion that the analysis of the bootstrap for panel data merits further attention. This property of the bootstrap is well documented in the literature (see, e.g., Hall (1992), for independent data or Lahiri (2003), for weakly dependent data) The consideration of the bootstrap for panel data has mainly focussed on resampling in the time dimension extending the work on the bootstrap in time series. Resampling in the cross sectional dimension has received less attention and, in particular no rigorous treatment of such resampling for large N - T panel datasets seems to be currently available in the econometric literature.

This paper aims to provide an initial treatment of the bootstrap when resampling occurs either in the cross sectional dimension or more generally in both cross sectional and time series dimensions. In a nutshell, cross sectional resampling consists of resampling cross sectional units as wholes rather than resampling within the units across the time dimension. The motivation for such resampling is clear when N is large compared to T. In particular, it is the only kind of resampling that will provide asymptotically valid bootstrap procedures when N increases but T remains fixed. Nevertheless, this is not very interesting as treatment of this case bears analogies to the treatment of the bootstrap for multivariate time series with N and T transposed.

The analysis becomes more interesting when both N and T are large. There, cross sectional resampling is an alternative to time series resampling. Both are asymptotically valid. The paper will discuss the asymptotic validity of cross sectional resampling in this context. The question of what sort of resampling to use becomes more interesting when dependence is considered. Allowing for temporal dependence in panel data is of course essential in the large N-T context. On the other hand, the analysis of cross sectional dependence is much less developed and assuming no such dependence is quite common in the literature. This is crucial for the bootstrap. Dependent data cannot be resampled in the same way as independent data and methods such as the block bootstrap need to be employed. Furthermore, the use of

the block bootstrap or its variants has been shown to provide less accurate approximations that the bootstrap in i.i.d. context as discussed in, e.g., Lahiri (2003) or Andrews (2002). We show that if there exists temporal dependence but no cross sectional dependence then cross sectional resampling can be more accurate than temporal resampling.

The structure of the paper is as follows: Section 2 provides a discussion of cross sectional resampling. Section 3 provides theoretical results for the bootstrap based on cross sectional resampling for a particular estimator. Section 4 presents a Monte Carlo analysis of the new bootstrap procedure. Finally, Section 5 concludes.

2 The Bootstrap for Panel Datasets

In this section we discuss various possibilities for bootstrap resampling schemes that can be applied in large N - T panel datasets. In order to do this we introduce a general panel model. This is given by

$$y_{i,t} = \boldsymbol{z}_t'\boldsymbol{a}_i + \boldsymbol{x}_{i,t}'\boldsymbol{\beta} + \epsilon_{i,t}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{1}$$

The focus of attention is inference on the vector $\boldsymbol{\beta}$. $\boldsymbol{z}_t$ is a vector of variables that enter all cross sectional units. In many applications it will contain deterministic terms such as a constant or a trend. $\boldsymbol{x}_{i,t} = (x_{1,i,t}, \ldots, x_{k,i,t})'$ contains explanatory variables that are particular to a given cross-sectional unit. We will regulate the behaviour of the explanatory variables, the coefficients and the error term, $\epsilon_{i,t}$ via appropriate assumptions in the next section, but keep the discussion heuristic at this stage to concentrate on the intuition.

We assume the existence of an estimator $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ which is consistent and, suitably normalised, asymptotically normal. The exact nature of the estimator will depend on the assumptions made about (1). Define $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i, \ldots, \boldsymbol{y}_N) = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t \ldots, \boldsymbol{y}_T)'$, $\boldsymbol{X}_j = (\boldsymbol{x}_{j,1}, \ldots, \boldsymbol{x}_{j,i}, \ldots, \boldsymbol{x}_{j,N}) = (\boldsymbol{x}_{j,1}, \ldots, \boldsymbol{x}_{j,t} \ldots, \boldsymbol{x}_{j,T})'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_i \ldots, \boldsymbol{\epsilon}_N) = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t \ldots, \boldsymbol{\epsilon}_T)'$ $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T)'$, $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,T})'$, $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{N,t})'$, $\boldsymbol{x}_{j,i} = (x_{j,i,1}, \ldots, x_{j,i,T})'$, $\boldsymbol{x}_{j,t} = (x_{j,1,t}, \ldots, x_{j,N,t})'$, $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,T})'$ and $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \ldots, \epsilon_{N,t})'$. We first consider a 'fixed effects' interpretation of the model and assume the existence of consistent estimates for each $\boldsymbol{a}_i$, denoted by $\hat{\boldsymbol{a}}_i$. Define $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N)$ and $\hat{\boldsymbol{A}} = (\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_N)$

We now consider the definition of a bootstrap sample. We distinguish between the parametric and the nonparametric bootstrap. There is the obvious tradeoff between the two

depending on how realistic one considers the assumed model to be. We define the nonparametric bootstrap sample to be given by the following set: $\{\boldsymbol{Y}^*, \boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_k^*\}$. Likewise, the parametric bootstrap sample is given by $\left\{ \boldsymbol{\epsilon}^*, \boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_k^*, \hat{\boldsymbol{A}}^* \middle| \hat{\boldsymbol{\beta}} \right\}$ where starred entries have been obtained by some sort of resampling from their non-starred counterparts. We now focus on possible resampling schemes.

Dealing first with the nonparametric bootstrap, the most common scheme for resampling both $y_{i,t}$ and $\boldsymbol{x}_{i,t}$ operates in the time dimension and consists of drawing with replacement either individual rows, or, in the case where the data are assumed to be dependent, blocks of contiguous rows from $\boldsymbol{Y}$ and $\boldsymbol{X}_j$, where the block size is assumed to depend solely on and grow with T. So, for example in the case of independent data, $\boldsymbol{Y}^* = (\boldsymbol{y}_{t_1}, \ldots, \boldsymbol{y}_{t_t} \ldots, \boldsymbol{y}_{t_T})'$ where each element of the vector of indices $(t_1, \ldots, t_T)'$ is obtained by drawing with replacement from $(1, \ldots, T)'$. The same vector of indices is used to obtain $\boldsymbol{X}_j^*$, $j = 1, \ldots, k$.

Cross-sectional resampling on the other hand resamples columns of $\boldsymbol{Y}$ with replacement. Thus, in this case, $\boldsymbol{Y}^* = (\boldsymbol{y}_{i_1}, \ldots, \boldsymbol{y}_{i_i} \ldots, \boldsymbol{y}_{i_N})$ where each element of the vector of indices $(i_1, \ldots, i_N)'$ is obtained by drawing with replacement from $(1, \ldots, N)'$. The same vector of indices is used to obtain $\boldsymbol{X}_j^*$, $j = 1, \ldots, k$. In the case of cross sectional dependence blocks of columns of $\boldsymbol{Y}$ can be randomly resampled with replacement. In this case, $\boldsymbol{Y}^* = (\boldsymbol{y}_{i_1}, \boldsymbol{y}_{i_1+1}, ..., \boldsymbol{y}_{i_1+b}, \ldots, \boldsymbol{y}_{i_i}, \boldsymbol{y}_{i_i+1}, ..., \boldsymbol{y}_{i_i+b}, \ldots, \boldsymbol{y}_{i_{[N/b]}}, \boldsymbol{y}_{i_{[N/b]}+b})$ where the vector of indices $(i_1, \ldots, i_{[N/b]})'$ is obtained by drawing with replacement from $(1, \ldots, N-b)'$ and b denotes the block size.

Of course, a combination of the two resampling schemes is also possible. The combination is obtained as follows: Let the temporally resampled bootstrap sample be denoted by $\tilde{\boldsymbol{Y}}^* = (\boldsymbol{y}_{t_1}, \ldots, \boldsymbol{y}_{t_t} \ldots, \boldsymbol{y}_{t_T})' \equiv (\tilde{\boldsymbol{y}}_1^*, \ldots, \tilde{\boldsymbol{y}}_i^* \ldots, \tilde{\boldsymbol{y}}_N^*)$. Then, the bootstrap sample from the combination of temporal and cross-sectional resampling is given by $\boldsymbol{Y}^* = (\tilde{\boldsymbol{y}}_{i_1}^*, \ldots, \tilde{\boldsymbol{y}}_{i_i}^* \ldots, \tilde{\boldsymbol{y}}_{i_N}^*)$ In this case, temporal resampling occurs first followed by cross-sectional resampling. It is equivalent to transpose the order of the two resampling operations. Block resampling is straightforwardly defined.

In summary, formal definitions for the various resampling schemes suggested above, are provided below.

Definition 1 (cross-sectional resampling) *For a $T \times N$ matrix of random variables $\boldsymbol{Y}$, cross sectional resampling is defined as the operation of constructing a $T \times N^*$ matrix $\boldsymbol{Y}^*$*

where the columns of $\boldsymbol{Y}^*$ are a random resample with replacement of blocks of the columns of $\boldsymbol{Y}$ and N^* is not necessarily equal to N.

Definition 2 (temporal resampling) *For a $T \times N$ matrix of random variables $\boldsymbol{Y}$, temporal resampling is defined as the operation of constructing a $T^* \times N$ matrix $\boldsymbol{Y}^*$ where the rows of $\boldsymbol{Y}^*$ are a random resample with replacement of blocks of the rows of $\boldsymbol{Y}$ and T^* is not necessarily equal to T.*

Definition 3 (cross-sectional/temporal resampling) *For a $T \times N$ matrix of random variables $\boldsymbol{Y}$, cross sectional/temporal resampling is defined as the operation of constructing a $T^* \times N^*$ matrix $\boldsymbol{Y}^*$ where the columns and rows of $\boldsymbol{Y}^*$ are a random resample with replacement of blocks of the columns and rows of $\boldsymbol{Y}$ and N^*, T^* are not necessarily equal to N, T.*

The parametric bootstrap can be implemented similarly with the residual matrix $\boldsymbol{\epsilon}$, rather than $\boldsymbol{Y}$ being resampled together with the $\boldsymbol{X}_j$ in the manner discussed above. Then, the estimates of the model parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{A}}$ are used to construct $\boldsymbol{Y}^*$.

Moving on to a 'random effects' interpretation of the panel model, we abstract from the issue of estimating $\boldsymbol{\beta}$ but simply assume that some appropriate estimator has been used. Of course, the dichotomy between the 'random effects' and 'fixed effects' interpretation is not relevant for the nonparametric bootstrap. For the parametric bootstrap, we note that by assumption $\boldsymbol{a}_i$ is independent of $\boldsymbol{a}_j$, $\forall j \neq i$. Then, we define the residual term $\hat{\varepsilon}_{i,t} = y_{i,t} - \boldsymbol{x}'_{i,t}\hat{\boldsymbol{\beta}}$. Note that, conditional on $\boldsymbol{z}_t$, $\hat{\varepsilon}_{i,t}$ is independent of $\hat{\varepsilon}_{j,t}$ $\forall j \neq i$. Similarly to $\boldsymbol{\epsilon}$, we construct the matrix $\boldsymbol{\varepsilon}$ and simply resample from it either cross-sectionally, temporally or both.

3 Theoretical Results

In this section we provide some theoretical results for the bootstrap based on cross-sectional resampling. We will deal with the nonparametric bootstrap. Similar treatments for the parametric bootstrap can also be considered. The following assumptions are made:

Assumption 1 *For each i the regressors, $\boldsymbol{x}_{i,t}$, are covariance stationary with absolutely summable autocovariances, zero means and finite fourth-order moments and are distributed independently of the individual-specific errors, $\epsilon_{i,t'}$, for all t and t'. The regressors are independent across i.*

Assumption 2 *The observed common effects, $\boldsymbol{z}_t$, are covariance stationary with absolute summable autocovariances, distributed independently of the individual-specific errors, $\epsilon_{i,t'}$, for all t and t'.*

Assumption 3 *The slope coefficients of the individual-specific effects, $\boldsymbol{\beta}_i$ are restricted to be equal to a common value, $\boldsymbol{\beta}$ The coefficients of the observed common effects, $\boldsymbol{\alpha}_i$, are bounded (lie on a compact set).*

Assumption 4 *The individual specific error, $\epsilon_{i,t}$, is distributed independently across i and t with mean zero, variance, σ_i^2, and a finite fourth-order moment, $E(\epsilon_{i,t}^4) \leq K$.*

The discussion needs for concreteness some given estimators for $\boldsymbol{\beta}$. Of course, alternative estimators could be analysed. We define the following two estimators for $\boldsymbol{\beta}$ which are common in the literature (see, e.g., Pesaran (2002) and the references cited therein). First, we define the pooled estimator given by

$$\hat{\boldsymbol{\beta}}_P = \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i'\boldsymbol{M}\boldsymbol{x}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i'\boldsymbol{M}\boldsymbol{y}_i\right) \tag{2}$$

where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. Secondly, we define the mean group estimator

$$\hat{\boldsymbol{\beta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\beta}_i \tag{3}$$

Define

$$T_j = N^{1/2}\hat{\boldsymbol{\Sigma}}_{jx}^{-1/2}(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}) \tag{4}$$

where

$$\hat{\boldsymbol{\Sigma}}_{Px} = \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i'\boldsymbol{M}\boldsymbol{x}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i\boldsymbol{x}_i'\boldsymbol{M}\boldsymbol{x}_i\left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i'\boldsymbol{M}\boldsymbol{x}_i\right)^{-1} \tag{5}$$

$$\hat{\boldsymbol{\Sigma}}_{MGx} = \frac{1}{N^2}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta})' \tag{6}$$

and

$$T_j^* = N^{1/2}\hat{\boldsymbol{\Sigma}}_{jx}^{*-1/2}(\hat{\boldsymbol{\beta}}_j^* - \hat{\boldsymbol{\beta}}) \tag{7}$$

where

$$\hat{\boldsymbol{\Sigma}}_x^* = \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i^{*\prime}\boldsymbol{M}^*\boldsymbol{x}_i^*\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^*\boldsymbol{x}_i'\boldsymbol{M}\boldsymbol{x}_i\left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i^{*\prime}\boldsymbol{M}^*\boldsymbol{x}_i^*\right)^{-1} \tag{8}$$

$$\hat{\boldsymbol{\Sigma}}_{MGx}^* = \frac{1}{N^2}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_{MG}^* - \hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_{MG}^* - \hat{\boldsymbol{\beta}})' \tag{9}$$

and $j = P, MG$.

Theorem 1 *Let assumptions 1-4 hold. Then, for fixed T, $T_j^* \xrightarrow{d} N(0, \boldsymbol{I})$. Further, as $T \to \infty$ and $N \to \infty$ sequentially, $T^{1/2} T_j^* \xrightarrow{d} N(0, \boldsymbol{I})$.*

Proof:

We deal with the pooled estimator first. Substituting the true model for $y_{i,t}^*$ in the estimator gives

$$N^{1/2}(\hat{\boldsymbol{\beta}}_P^* - \hat{\boldsymbol{\beta}}) = \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{x}_i^* \right)^{-1} \left(\frac{1}{N^{1/2}} \sum_{i=1}^{N} \boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{\epsilon}_i^* \right) \tag{10}$$

For fixed T, $\{\boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{\epsilon}_i^*\}_i$ is a sequence of i.i.d. random variables with expectation 0 and variance $\sigma_i E(\boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{x}_i^*)$ by assumptions 1, 2 and 4. Hence, by a standard central limit theorem for i.d. random variables (see, e.g., theorem 25.2 of Davidson (1994)),

$$\left(\frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_i^* \boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{x}_i^* \right) \frac{1}{N^{1/2}} \sum_{i=1}^{N} \boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{\epsilon}_i^* \xrightarrow{d} N(0, I) \tag{11}$$

Hence, the result follows. If $T \to \infty$, then

$$(NT)^{1/2}(\hat{\boldsymbol{\beta}}_P^* - \hat{\boldsymbol{\beta}}) = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{x}_i^*}{T} \right)^{-1} \left(\frac{1}{N^{1/2}} \sum_{i=1}^{N} \frac{\boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{\epsilon}_i^*}{T^{1/2}} \right) \tag{12}$$

Then, by assumption 1

$$\frac{\boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{x}_i^*}{T} \to Q_i \tag{13}$$

where Q_i is a positive definite matrix and

$$\frac{\boldsymbol{x}_i^{*\prime} \boldsymbol{M}^* \boldsymbol{\epsilon}_i^*}{T^{1/2}} \xrightarrow{d} N(0, \sigma_i^* Q_i) \tag{14}$$

again leading to the required result.

Moving on to the mean group estimator, we see that the result for fixed T is obvious. To see this note that

$$\hat{\boldsymbol{\beta}}_{MG}^* = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\beta}_i^* \tag{15}$$

But $\boldsymbol{\beta}_i^*$ is simply an i.i.d. resample from $\{\hat{\boldsymbol{\beta}}_i\}_{i=1}^N$. Since $\hat{\boldsymbol{\beta}}_i$ are i.d. random variables with finite second moments and equal mean (by assumption 3) the result follows, noting that the variance estimator given in (9) converges to the true variance following a standard law of large numbers for i.d. random variables. For $T \to \infty$, $\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})$ converges to a normal distribution and hence the above argument for fixed T can be applied to a resample from $\{\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})\}_{i=1}^N$ when $T \to \infty$ followed by $N \to \infty$. Q.E.D.

Assumption 5 *The joint distribution of* $\{\boldsymbol{x}_{i,t}\}_{t=1}^{T}$ *and the distribution of* $\epsilon_{i,t}$ *is the same for all* i*.*

Assumption 6 *Let* $\boldsymbol{u}_{it} = (\boldsymbol{x}'_{i,t}, \epsilon_{i,t})'$*.* $E(||\boldsymbol{u}_{it}||)^l < \infty$*,* $l < \infty$

Assumption 7 *Let* $\chi(\boldsymbol{u})$ *denote the characteristic function of* $\boldsymbol{u}_{it}$*.* $\chi(\boldsymbol{u})$ *satisfies Cramer's condition*

$$\lim \sup_{||\boldsymbol{u}|| \to \infty} |\chi(\boldsymbol{u})| < 1 \tag{16}$$

Theorem 2 *Let the assumptions underlying theorem 1 hold. Assume further, assumptions 5-7. Then, the bootstrap estimate of the distribution of* T_j *is* $O_p(N^{-1})$ *consistent.*

Proof:

We will provide a proof for the simple case of only the constant belonging to $\boldsymbol{z}_t$ and a single x regressor. The general case of multiple z and x follows with appropriate modifications. We first consider the fixed T case. The estimator $\hat{\beta}_P$ is given by

$$\hat{\beta}_P = \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)^2 \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)(y_{i,t} - \bar{y}_i) \right) \tag{17}$$

where $\bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{i,t}$. Substituting in the true model for $y_{i,t}$ gives

$$\hat{\beta}_P - \beta = \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)^2 \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)\epsilon_{i,t} \right) \tag{18}$$

Denote

$$\mathcal{X}_i = \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)^2 \tag{19}$$

and

$$\mathcal{Y}_i = \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)\epsilon_{i,t} \tag{20}$$

and remember the definition of

$$\hat{\Sigma}_x = (\frac{1}{N} \sum_{i=1}^{N} \mathcal{X}_i)^{-2} (\frac{1}{N} \sum_{i=1}^{N} \sigma_{x,i}^2) \tag{21}$$

where $\sigma_{x,i}^2 = \sum_{t=1}^{T} \hat{\sigma}_i^2 (x_{i,t} - \bar{x}_i)^2$. Then, the quantity whose distribution we are estimating is easily seen to be a function of means of i.i.d. random variables, by assumption. These random variable sequences are $\mathcal{X}_i$, $\mathcal{Y}_i$ and $\sigma_{x,i}^2$. Their means over i are denoted by $\bar{\mathcal{X}}_i$ $\bar{\mathcal{Y}}_i$ and $\bar{\sigma}_{x,i}^2$. They are i.i.d. by assumption 4. Denote the function of the means by $N^{1/2} A(\bar{\mathcal{Z}})$, where

$\bar{\mathcal{Z}} = (\bar{\mathcal{X}}_i, \bar{\mathcal{Y}}_i, \bar{\sigma}^2_{x,i})$. Further, denote the bootstrap equivalent of $N^{1/2}A(\bar{\mathcal{Z}})$ by $N^{1/2}A^*(\bar{\mathcal{Z}})$. We then consider Edgeworth expansions for $N^{1/2}A(\bar{\mathcal{Z}})$ and $N^{1/2}A^*(\bar{\mathcal{Z}})$. By assumption 6, we have that x_i, t and $\epsilon_{i,t}$ possess moments of sufficiently high order, denoted l. It then follows that, for fixed T, $\bar{\mathcal{X}}_i$ $\bar{\mathcal{Y}}_i$ and $\bar{\sigma}^2_{x,i}$ possess moments of the same order. Then, under assumptions 5-7, it follows from theorem 5.1 of Hall (1992) that

$$sup_{-\infty<w<\infty}\left|P(N^{1/2}A(\bar{\mathcal{Z}}) \leq w) - \Phi(w) - \sum_{j=1}^{\nu} N^{-j/2}q_j(w)\phi(w)\right| = O(N^{-\nu/2}) \tag{22}$$

and

$$sup_{-\infty<w<\infty}\left|P^*(N^{1/2}A^*(\bar{\mathcal{Z}}) \leq w) - \Phi(w) - \sum_{j=1}^{\nu} N^{-j/2}\hat{q}_j(w)\phi(w)\right| = O(N^{-(\nu+1)/2}) \tag{23}$$

where $\Phi(.)$, and $\phi(.)$ denote the standard normal distribution and density functions respectively, $q_j(w)$ are polynomials of population cumulants of $\mathcal{X}_i$, $\mathcal{Y}_i$ and $\sigma^2_{x,i}$ and $\hat{q}_j(w)$ are as $q_j(w)$ but where the population quantities are replaced by sample ones. These are the Edgeworth expansions corresponding to $N^{1/2}A(\bar{\mathcal{Z}})$ and $N^{1/2}A^*(\bar{\mathcal{Z}})$. Inverting these expansions gives Cornish-Fisher expansions of the distribution quantiles given by

$$v_\alpha = z_\alpha + \sum_{j=1}^{\nu} N^{-j/2}q_{j1}(z_\alpha) \tag{24}$$

$$\hat{v}_\alpha = z_\alpha + \sum_{j=1}^{\nu} N^{-j/2}\hat{q}_{j1}(z_\alpha) \tag{25}$$

where v_α and z_α are the solutions of $P(N^{1/2}A(\bar{\mathcal{Z}}) \leq v_\alpha) = \alpha$ and $\Phi(z_\alpha) = \alpha$ and $q_{j1}(.)$ and $\hat{q}_{j1}(.)$ are polynomials defined in terms of q_j and $\hat{q}_j(.)$. Since, sample moments and cumulants are $O_p(N^{-1/2})$ consistent estimators of population moments it follows that $\hat{q}_j(w) = q_j(w) + O_p(N^{-1/2})$ and so $\hat{v}_\alpha - v_\alpha = O_p(N^{-1})$ completing the proof for fixed T.

For sequential (N, T) asymptotics we consider a sequence where $T \to \infty$ followed by $N \to \infty$. This case is much simplified since as $T \to \infty$, $\frac{1}{T}\mathcal{X}_i$ and $\frac{1}{T}\sigma^2_{x,i}$ tend in probability to constants. Further, $\frac{1}{\sqrt{T}}\mathcal{Y}_i$ tends to a normal distribution which automatically satisfies assumptions 4-6. Hence, the result follows via a similar treatment to the fixed T case. Q.E.D.

4 Monte Carlo Study

In this section we carry out a Monte Carlo analysis of the various resampling schemes described in Section 2. Two different models are considered. The first is given by (1). The

Monte Carlo design for this model (Case I) is designed to satisfy the assumptions of Section 3. The second model (Case II) extends (1) by allowing $\epsilon_{i,t}$ to contain a common factor effect along the lines of Pesaran (2002). In other words

$$\epsilon_{i,t} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{i,t}$$

where $\boldsymbol{f}_t$ is an m-dimensional weakly dependent process which satisfies the same assumption as z_t (i.e. assumption 2), $\varepsilon_{i,t}$ has the same properties as $\epsilon_{i,t}$ did previously and γ_i can be interpreted as either fixed bounded constants or i.i.d. random variables across i. Note that although the factor introduces cross sectional dependence in the panel, this is not a problem for i.i.d. cross sectional resampling. The reason for this is that the factor is basically another z_t variable which does not introduce any spatial structure to the panel. Hence, a random i.i.d. cross sectional resample of the original sample will replicate its properties as long as $N \to \infty$. In order to estimate $\boldsymbol{\beta}$ for the model with the factors we follow Pesaran (2002) and use the following estimators. Define

$$\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{i,t}, \;\; \bar{\boldsymbol{x}}_t = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_{i,t}. \tag{26}$$

Let $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{y}}$ be $T \times k$ and $T \times 1$ observation matrices on the aggregates $\bar{\boldsymbol{x}}_t$ and $\bar{y}_t$, respectively. Then , $\hat{\boldsymbol{\beta}}_i$ for the mean group estimator is defined as

$$\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{X}_i)^{-1} \boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{y}_i, \tag{27}$$

where

$$\bar{\boldsymbol{M}} = \boldsymbol{I}_T - \bar{\boldsymbol{H}}(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}})^{-1}\bar{\boldsymbol{H}}', \tag{28}$$

and $\bar{\boldsymbol{H}} = (\boldsymbol{Z}, \bar{\boldsymbol{X}}, \bar{\boldsymbol{y}})$.The pooled estimator is defined as

$$\hat{\boldsymbol{\beta}}_{PC} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \bar{\boldsymbol{M}} \boldsymbol{y}_i. \tag{29}$$

These estimators are proven to be consistent and asymptotically normal by Pesaran (2002).

We first provide specifications for the Monte Carlo experiments for Case I. We allow $m = 1, 3$. z_t and k are set to 1. We set

$$f_{s,t} = \rho_{s,f} f_{s,t-1} + \varepsilon_{f,s,t}, \;\; s = 1, ..., m$$

and

$$x_{i,t} = \rho_{i,x} x_{i,t-1} + \sum_{s=1}^{m} \phi_{i,s} f_{s,t} + + v_{i,t} \; i = 1, ..., N; \;\; s = 1, ..., m$$

Let

$$\boldsymbol{\xi}_{it} = \begin{pmatrix} \varepsilon_{i,t} \\ v_{i,t} \\ \boldsymbol{\varepsilon}_{f,i,t} \end{pmatrix},$$

$\boldsymbol{\xi}_{it}$ is generated as iid $N(\mathbf{0}, \boldsymbol{\Sigma}_{\xi i})$, where

$$\boldsymbol{\Sigma}_{\xi i} = diag\left(\sigma^2_{i\varepsilon}, \sigma^2_{iv}, \sigma^2_{i\varepsilon_f}\right).$$

where $\sigma^2_{i\varepsilon_f} = 1 - \rho^2_{if}$. We let $\rho_{sf} = 0.5$, $\rho_{ix} \sim U[0.2, 0.9]$, $\phi_{i,s} \sim U[0.5, 1.5]$, $\sigma^2_{i\varepsilon} \sim U[0.5, 1.5]$, $\sigma^2_{iv} \sim U[1.5, 2.5]$ and $\gamma_{i,s} \sim N(1, 0.04)$. The final set of parameters to be fixed is $\boldsymbol{\gamma}_{is}$. $\boldsymbol{\gamma}_{is}[k] \sim N(1, 0.04)$. $N, T = 50, 100, 150$. 1000 replications are carried out. For Case II we simply set $\phi_{i,s} = \gamma_{i,s} = 0$. We carry out cross sectional resampling, temporal resampling with a block structure where the block size is set to $[T^{1/4}]$ as suggested by, e.g., Lahiri (2003) or Andrews (2002), and combined cross sectional and temporal resampling. We focus on the properties of the bootstrap variance estimator and report root mean squared errors as performance measures of the estimators. The estimators are denoted by V^N, V^T and $V^{N,T}$ for the cross sectional, temporal and combined cross sectional and temporal resampling, respectively. Results are reported in Tables 1-4.

Results make interesting reading. Clearly, the cross sectional resampling does much better than the temporal resampling for all cases considered apart from the case $N = 50, T = 150$. This is expected as cross sectional resampling improves with N. But, the relative performance of cross sectional resampling improves as both N and T increase together as we can see from the diagonal elements of the panels of Tables 1 and 3. This implies that the i.i.d. resampling nature of cross sectional resampling is superior to the block temporal resampling scheme. This provides some evidence supporting the theoretical result in Theorem 2.

Results are similar for both Cases I and II and both pooled and mean groups estimators, leading us to suggest that the performance of cross sectional resampling is good in a variety of panel data models. In particular, the result obatined in Case II is of considerable interest as it implies that strong forms of cross-sectional dependence that do not have a local cross-sectional structure such as factor structures can still be dealt with i.i.d. resampling making the applicabilty of the new procedures much wider. This is indeed of practical significance given the recent work of, among others, Stock and Watson (1998), Bai and Ng (2002), Bai (2003), Bai and Ng (2004) and Pesaran (2002).

Tables 2 and 4 report results on the combined cross sectional and temporal resampling. We report absolute RMSE results there because the combined resampling scheme performs

much worse that either of the other two resampling schemes. In particular the variance estimator is considerably upwards biased. Neverthless, its performance improves when either N and T increase.

5 Conclusions

This paper has considered the issue of bootstrap resampling in panel datasets. The availability of datasets with large temporal and cross sectional dimensions suggests the possibility of new resampling schemes. We suggest one possibility which has not been widely explored in the literature. It amounts to constructing bootstrap samples by resampling whole cross sectional units with replacement.

In cases where the data do not exhibit cross sectional dependence but exhibit temporal dependence, such a resampling scheme is of interest as it allows the application of i.i.d. bootstrap resampling rather than block bootstrap resampling. It is well known that the former enables superior approximation to distributions of statistics compared to the latter. We prove that the bootstrap based on cross sectional resampling provides asymptotic refinements. A Monte Carlo study illustrates the superior properties of the new resampling scheme compared to the block bootstrap.

References

Andrews, D. W. K. (2002): "Higher Order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators," *Econometrica*.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dinensions," *Econometrica*, 71, 135–173.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

——— (2004): "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, Forthcoming.

Baltagi, B. H. (2001): *Econometric Analysis of Panel Data*. Wiley.

Davidson, J. (1994): *Stochastic Limit Theory*, Advanced Tests in Econometrics. Oxford University Press.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics. Springer-Verlag.

Hsiao, C. (2003): *Analysis of Panel Data*. Cambridge University Press.

Lahiri, S. N. (2003): *Resampling Methods for Dependent Data*. Springer.

Pesaran, M. H. (2002): "Estimation and Inference in Large Heterogenous Panels with Cross-Section Dependence," *University of Cambridge DAE Working Paper No.0305 and CESifo Working Paper Series No 869 (Revised Version)*.

Stock, J., and M. W. Watson (1998): "Diffusion Indexes," *NBER Working Paper 6702*.

Table 1: Results for Model with Factors						
Pooled estimator $\frac{RMSE(V^N)}{RMSE(V^T)}$						
	nf=1			nf=3		
T/N	50	100	150	50	100	150
50	0.837	0.651	0.575	0.823	0.615	0.512
100	0.988	0.842	0.705	1.023	0.767	0.553
150	1.169	0.863	0.759	0.863	0.667	0.543
Mean Group estimator $\frac{RMSE(V^N)}{RMSE(V^T)}$						
50	0.873	0.665	0.609	0.889	0.652	0.526
100	1.163	0.838	0.731	1.070	0.807	0.549
150	1.288	0.928	0.755	0.885	0.662	0.489

Table 2: Results for Model with Factors						
Pooled estimator 10^3*RMSE($V^{N,T}$)						
	nf=1			nf=3		
T/N	50	100	150	50	100	150
50	1.161	0.560	0.371	1.077	0.527	0.345
100	0.515	0.255	0.167	0.509	0.247	0.157
150	0.334	0.160	0.106	0.324	0.158	0.102
Mean Group estimator 10^3*RMSE($V^{N,T}$)						
50	1.602	0.773	0.511	1.463	0.706	0.459
100	0.701	0.339	0.225	0.666	0.319	0.202
150	0.446	0.213	0.139	0.418	0.202	0.129

Table 3: Results for Model without Factors			
Pooled estimator $\frac{RMSE(V^N)}{RMSE(V^T)}$			
T/N	50	100	150
50	0.822	0.638	0.590
100	0.949	0.813	0.701
150	1.107	0.845	0.738
Mean Group estimator $\frac{RMSE(V^N)}{RMSE(V^T)}$			
50	0.842	0.678	0.610
100	1.112	0.809	0.735
150	1.177	0.903	0.732

Table 4: Results for Model without Factors			
Pooled estimator 10^3*RMSE($V^{N,T}$)			
T/N	50	100	150
50	1.157	0.564	0.372
100	0.509	0.253	0.167
150	0.324	0.159	0.105
Mean Group estimator 10^3*RMSE($V^{N,T}$)			
50	1.611	0.781	0.513
100	0.698	0.339	0.225
150	0.434	0.213	0.138

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm