DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Panel Data with Cross-Sectional Dependence Characterized by a Multi-Level Factor Structure

Carlos Vladimir Rodríguez-Caballero

CREATES Research Paper 2016-31

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Panel Data with Cross-Sectional Dependence Characterized by a Multi-Level Factor Structure

Carlos Vladimir Rodríguez-Caballero*
Aarhus University, Department of Economics and Business Economics,
CREATES†

October 31, 2016

Abstract

A panel data model with a multi-level cross-sectional dependence is proposed. The factor structure is driven by top-level common factors as well as non-pervasive factors. I propose a simple method to filter out the full factor structure that overcomes limitations in standard procedures which may mix up both levels of unobservable factors and may hamper the identification of the model. The model covers both stationary and non-stationary cases and takes into account other relevant features that make the model well suited to the analysis of many types of time series frequently addressed in macroeconomics and finance. The model makes it possible to examine the time series and cross-sectional dynamics of variables allowing for a rich fractional cointegration analysis. A Monte Carlo simulation is conducted to examine the finite sample features of the suggested procedure. Findings indicate that the methodology proposed works well in a wide variety of data generation processes and has much lower biases than the alternative estimation methods either in the I(0) or I(d) cases.

Keywords: Cross-section dependence; Multi-level factor models; Large panels; Long memory; Fractional cointegration; Common correlated effects.
JEL Classification: C12, C22, C33.

*Corresponding author. E-mail: vrodriguez@econ.au.dk. Fuglesangs Allé 4, 8210 Aarhus V, Denmark

†Acknowledgments. Some of this paper was written while I was visiting the Department of Statistics at University of Padua as part of the Young Investigator Training Program (YITP) prize awarded by the Association of Foundations of Banking Origin (ACRI) and the Italian Econometric Association. I wish to give thanks to Massimiliano Caporin for hosting me and sharing some interesting ideas and comments. Furthermore, I am grateful to Niels Haldrup, Carlos Velasco, Eric Hillebrand, Yunus Emre Ergemen, and Giovanni Bonaccolto for most helpful comments on a previous version of this paper. I am also acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

1 Introduction

Panel data models are used in economics and finance to analyze complex systems and phenomena that control for individual heterogeneity, allow for cross-sectional dependence, and identify and estimate effects that are not detectable in pure cross section or time series frameworks. Recent studies have focused on testing non-stationarity in presence of cross-sectionally correlated errors. Phillips and Sul (2003) and Moon and Perron (2004) allow for different factor structures to test for unit roots in panels but the cross-sectional dependence is considered only as a nuisance. In contrast, the 'Panel Analysis of Nonstationarity in Idiosyncratic and Common components' (PANIC) approach proposed by Bai and Ng (2004) considers such dependence as an object of interest and makes the framework very suitable to study unit roots, common trends, and common cycles in large dimensional panels.

The presence of cross-sectional dependence can noticeably complicate statistical inference in a panel data model. In cases of 'full' dependence among cross-sectional units, or denoted as 'strong dependence' by Pesaran and Tosetti (2011), estimators that ignore such a dependence could be inconsistent no matter how large the cross section dimension N is for finite time dimension T, see e.g. Hsiao and Tahmiscioglu (2008), Phillips and Sul (2007). There has been an increasing interest in dealing with cross-sectionally correlated errors in panel data sets. It is well-known that when the cross section dimension (N) is small and the time series dimension (T) is large, the standard approach of treating the model as a system of seemingly unrelated regression equations (SURE) and then estimate this system as generalized least squares (GLS) may be applicable. However, in macroeconomics and finance, panel data sets are generally presented in the form of a large cross-sectional dimension and with errors that are typically correlated with the regressors making the basic approach inappropriate.

Based on factor models that have been extensively studied in recent years from works of Bai (2003) and Bai and Ng (2002), the use of factor-augmented regressions has recently become very popular in the literature. The main idea is that cross-sectional dependence in panel data sets is driven by a small number of unobservable common factors that can be included as additional regressors. A common approach to deal with this unobservable structure in a panel data model is to use estimates of the factors to augment the model. Pesaran (2006) proposes to use cross-sectional averages of the observables as good proxies for the unobservable common factors. He refers to these estimators as the Common Correlated Effects (CCE). In contrast, Bai (2009) suggests to estimate the factor structure with principal components analysis (PCA). Westerlund and Urbain (2015) formally compare both of these approaches.

In contrast with the PCA approach, CCE methods do not require a prior knowledge of the number of unobserved common factors. This makes the estimation procedure simpler than the PCA. For this reason, the CCE approach has gained attention in the literature in recent years. Pesaran and Tosetti (2011) study

asymptotic properties of the CCE estimates when disturbances are generated by a spatial process. Kapetanios et al. (2011) and Ergemen and Velasco (2015) study the performance of CCE estimators when I(1) and I(d) processes, respectively, are considered in the common factors. Furthermore, Harding and Lamarche (2011) and Ergemen (2016) introduce endogeneity between the observable variables.

The baseline framework in this paper is a panel data model which can be divided in R blocks whose full cross-sectional dependence is characterized by two orthogonal levels, see e.g. Wang (2010), Choi et al. (2016), and Breitung and Eickmeier (2016). In the first (top) level, I define pervasive factors that drive the cross-sectional dependence between blocks, while in the second (sub) level, I define the block-specific factors that drive the cross-sectional dependence within-blocks. I follow ideas embodied by Pesaran (2006) to get CCE estimators but with a simple extended procedure that is executed in two separate steps. In the first step, the cross-sectional averages within-blocks are used to proxy the sub-level cross-sectional dependence. Then, after filtering out the block-specific factors from the panel, the cross-sectional averages between-blocks are used to project out the top-level factor from the specification. CCE estimators, under slope heterogeneity or homogeneity assumptions, are then obtained as discussed in Pesaran (2006).

In this paper, I also cover the case with long-range dependence extending the results of a couple of very recent proposals. First, I extend the results provided by Ergemen and Velasco (2015) who do not consider more than one block and do not allow a multi-level factor structure. Second, the approach of Ergemen and Rodríguez-Caballero (2016) is extended by adding explanatory variables in the model. The fractionally integrated model proposed in this paper allows to exhibit long-range dependence without restrictions on both levels of unobservable common factors being either stationary or nonstationary processes as in Ergemen and Rodríguez-Caballero (2016). Then, the model does not restrict the common factors to the I(1) case as Kapetanios et al. (2011). Furthermore, innovations of regressors and regressand are allowed to be fractionally integrated. Thereby, the model can be useful for a wide range of empirical applications where the variables exhibit long-range dependence on non-integer orders. To estimate the model, I follow the ideas mentioned before to filter out the full factor structure but using (fractionally) differenced cross-sectional averages as proposed by Ergemen and Velasco (2015). Then, estimation of the residual memory parameters of the model is based on a conditional-sum-of-square (CSS) criterion function of the residuals.

The framework studied in this paper differs from Pesaran (2006) since I assume that cross-sectional dependence is driven by a multi-level factor structure that is characterized by unobservable pervasive top-level common factors as well as unobservable block-specific factors which characterize the between- and within-block cross-sectional dependence in the data. Even though the model makes use of multi-level factor structures, it differs from some dynamic multi-level factors models, such as Wang (2010), Diebold et al. (2008), Kose et al. (2008), and Choi et al. (2016), in the sense that I use them merely to control for the cross-sectional dependence. Moreover, dynamic factor models do not consider the presence of

explanatory variables.

The study of panel data models in blocks with a factor structure has not been considered until very recently. On the one hand, Ando and Bai (2016b) consider a grouped panel data model in which cross-section units are divided into several groups with its own factor structure which is different in nature to the factor structure proposed in this paper in the sense that they do not consider a pervasive factor between different groups. On the other hand, Ando and Bai (2016a) propose a model where cross-section units are classified by clustering techniques in unknown grouped factors whose structure is closely similar to that proposed in this paper. Both proposals do not focus on the CCE method of Pesaran (2006) and consider only the I(0) case in contrast to the present paper which also allows for long-range dependence.

The framework of this paper can be applied whenever a panel of data can be organized into blocks. These blocks can be formed naturally for instance by dividing the panel data sets by some economic sectors or countries as in many macroeconomic studies, or by using some prior information or by implementing multivariate statistical procedures as clustering or recursive partitioning, see e.g. Bonhomme and Manresa (2015). The block factor structure approach would provide an easy way to allow for cross-sectional covariations that are not sufficiently pervasive to be treated as common factors in contrast with standard procedures. A large number of papers have applied multi-level factor models to study government bond yield data (Diebold et al. (2008)), international business cycle comovements (Kose et al. (2003)), and national and regional factors in housing construction (Stock and Watson (2009b)).

The paper is organized as follows. The next section introduces the model and the factor structure that characterizes the cross-sectional dependence in the panel for the I(0) case and the necessary model assumptions are discussed. Section 3 explains the modeling strategy to control the cross-sectional dependence. Section 4 presents the estimation procedure and the respective asymptotic analysis. Section 5 details the model with long-range dependence, the filtering method as well as the estimation procedure and the asymptotic theory. Section 6 briefly discusses some extensions. Section 7 presents the finite-sample properties of both models based on an extensive Monte Carlo designs, and finally, Section 8 concludes the paper.

Throughout the paper, M stands for a finite positive constant, $\|A\| = (trace(A'A))^{1/2}$ for a matrix A, A^- denotes the generalized inverse for a matrix A, $rk(A)$ denotes the rank of A, $(N,T)_j$ denotes the joint cross-section and time series asymptotics, $\rightarrow_{q.m.}$ denotes convergence in quadratic mean, $\rightarrow_p$ denotes convergence in probability, and $\rightarrow_d$ denotes convergence in distribution. All mathematical proofs are presented in the appendix.

2 The model

Consider a panel data model composed by R blocks of data. Such blocks are frequently referred to as regions in the literature. A multi-level factor structure drives the cross-sectional dependence in the model and is characterized by pervasive top-level factors affecting all the blocks in the panel and block-specific factors that affect only a specific block of the panel. Such factors are also denominated as global and regional factors in the literature.

Let $y_{r,it}$ be the observation on the i^{th} cross section unit at time t in the block r for $i = 1, \ldots, N$, $t = 1, \ldots, T$, $r = 1, \ldots, R$, and suppose that it is generated according to the following linear heterogeneous panel data model

$$\begin{aligned} y_{r,it} &= \alpha'_{r,i}\mathbf{d}_{r,t} + \beta'_{r,i}\mathbf{x}_{r,it} + e_{r,it}, \\ e_{r,it} &= \mu'_{r,i}G_t + \lambda'_{r,i}F_{r,t} + \epsilon_{r,it}, \end{aligned} \tag{1}$$

where $\mathbf{d}_{r,t}$ is a $R \cdot N \times 1$ vector of observed common effects (including deterministics such as intercepts, trends, or seasonal dummies) in the block r, $\mathbf{x}_{r,it}$ is a $k \times 1$ vector of observed individual-specific regressors on the i^{th} cross section unit at time t in the block r, G_t is the $\mathbf{r}_G \times 1$ vector of unobserved top-level common effects or global factors, $F_{r,t}$ is the $\mathbf{r}_F \times 1$ vector of unobserved block-specific common effects or regional factors, and $\epsilon_{r,it}$ are the individual-specific idiosyncratic errors that are independent of G_t, $F_{r,t}$, and $X_{r,it}$. I assume the number of blocks, R, to be fixed because it is more reasonable for practical purposes and is much more tractable for an asymptotic analysis. It is possible to consider that the number of cross section units vary among blocks at the cost of complicating notation. So it is assumed that the cross section dimension of the panel is $R \times N$. R could be much more smaller than N in empirical applications, however it is not required to have a specific rate between them.

If the multi-level factor structure given by G_t and $F_{r,t}$ is also independent of $\mathbf{x}_{r,it}$, (1) is a simple panel data regression model with exogenous regressors and can be estimated consistently and efficiently using GLS based on the multi-level factor structure. However, in general, the unobserved pervasive and block-specific factors can be correlated with $(\mathbf{d}_{r,t}, \mathbf{x}_{r,it})$, then consistency will be lost.

To allow for this possibility, I adopt the following specification for the individual specific regressors

$$\mathbf{x}_{r,it} = A'_{r,i}\mathbf{d}_{r,t} + M'_{r,i}G_t + \Lambda'_{r,i}F_{r,t} + \mathbf{v}_{r,it}, \tag{2}$$

where $A_{r,i}$, $M_{r,i}$, and $\Lambda_{r,i}$ are $N \times k$, $r_G \times k$, and $r_f \times k$ matrices with fixed components in the specific block r, $\mathbf{v}_{r,it}$ are the specific components of $x_{r,it}$ distributed independently of the factor structure and across i and r. In this section, it is assumed that all innovations are stationary, however, more general processes can be allowed in $\mathbf{x}_{r,it}$ and $y_{r,it}$ by including long memory, unit roots or deterministic trends in $\mathbf{d}_{r,t}$ or in the multi-level factor structure. I discuss these cases in the following sections. In this section, I only focus on the case where $\mathbf{d}_{r,t}$, G_t, and $F_{r,t}$ are covariance stationary.

Following the literature on multi-level factor models, see e.g. Wang (2010),Choi et al. (2016), and Breitung and Eickmeier (2016), the factor structure incorporated in models (1) and (2) imposes a block of zero restrictions on the associated matrix of factor loadings so that the factor structure on the model can be represented in vector form as

$$\begin{pmatrix} e_{1,\cdot t} \\ \vdots \\ e_{R,\cdot t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 & \boldsymbol{\lambda}_1 & 0 & \cdots & 0 \\ \boldsymbol{\mu}_2 & 0 & \boldsymbol{\lambda}_2 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ \boldsymbol{\mu}_R & 0 & 0 & \cdots & \boldsymbol{\lambda}_R \end{pmatrix} \begin{pmatrix} G_t \\ F_{1,t} \\ \vdots \\ F_{R,t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,\cdot t} \\ \vdots \\ \epsilon_{R,\cdot t} \end{pmatrix},$$

or more compactly as

$$e_t = \Gamma^* \, \Phi_t^* + \epsilon_t, \tag{3}$$

where $\Phi_t^* = \left(G_t', F_{1,t}', \ldots, F_{R,t}'\right)'$. Then, the factor structure considered in this paper is the sum of a pervasive common component, a non-pervasive common component, and an idiosyncratic component. Common components of the block r are driven by the respective $\mathbf{r}_G$ and $\mathbf{r}_F$ vectors of common factors, which are loaded with possibly different coefficients and lags in the case of a dynamic setup.

The intuition behind this factor structure is that the top-level factor component would capture common movements between blocks of data whilst the block-specific component would capture only common movements within the specific block. This framework differs from standard factor structures in two ways; first, in the block of zero restrictions in Γ^* defined in (3), and second, in the number of factors that grows with the number of blocks contrary to standard factor models where the number of factors is always fixed. Generally, when adding new series by considering new block of data in 3, the factor space will be expanded since new blocks of data would bring new block-specific shocks into the system. However, since the number of blocks at R is fixed, the number of factors is also fixed, which is necessary to achieve asymptotic results.

The importance of this kind of factor structure arises naturally in many different applications. For example when analyzing international business cycles, a multi-level factor model could be used to separate the country-specific comovements from the world-wide comovements.

Combining (1) and (2) the model can be re-written for the specific block r as

$$\underbrace{\mathbf{z}_{r,it}}_{(k+1)\times 1} = \underbrace{\mathbf{B}_{r,i}'}_{(k+1)\times N} \underbrace{\mathbf{d}_{r,t}}_{N\times 1} + \underbrace{\mathbf{C}_{r,i}'}_{(k+1)\times r_G} \underbrace{G_t}_{r_G\times 1} + \underbrace{\mathbf{D}_{r,i}'}_{(k+1)\times r_{F_r}} \underbrace{F_{r,t}}_{r_{F_r}\times 1} + \underbrace{\mathbf{u}_{r,it}}_{(k+1)\times 1}, \tag{4}$$

where

$$\mathbf{z}_{r,it} = \begin{pmatrix} y_{r,it} \\ \mathbf{x}_{r,it} \end{pmatrix}, \qquad \mathbf{B}_{r,i} = (\alpha_{r,i}\ A_{r,i}) \begin{pmatrix} 1 & 0 \\ \beta_{r,i} & \mathbf{I}_k \end{pmatrix},$$

$$\mathbf{C}_{r,i} = (\mu_{r,i}\ M_{r,i}) \begin{pmatrix} 1 & 0 \\ \beta_{r,i} & \mathbf{I}_k \end{pmatrix}, \qquad \mathbf{D}_{r,i} = (\lambda_{r,i}\ \Lambda_{r,i}) \begin{pmatrix} 1 & 0 \\ \beta_{r,i} & \mathbf{I}_k \end{pmatrix},$$

$$\mathbf{u}_{r,it} = \begin{pmatrix} \epsilon_{r,it} + \beta'_{r,i}\ \mathbf{v}_{r,it} \\ \mathbf{v}_{r,it} \end{pmatrix},$$

$\mathbf{I}_k$ is an identity matrix of order k, and the rank of matrices $\mathbf{C}_{r,i}$ and $\mathbf{D}_{r,i}$ are determined by the rank of the $\mathbf{r}_G \times (k+1)$ and $\mathbf{r}_{F_r} \times (k+1)$ matrices of the unobserved top-level and block-specific factor loadings, respectively. However, in a general way, the rank of the multi-level factor structure is determined by $rk(\Gamma^*)$ which is $\mathbf{r}_G + \sum_{r=1}^{R} \mathbf{r}_{F_r}$, i.e., the number of pervasive factors plus the sum of the number of block-specific factors in each region r.

This setup is similar to that proposed in Pesaran (2006) and Bai (2009) but I extent the analysis to large panels, which can be composed of several blocks of data. In this sense, I consider cross-sectional dependence in the panel not only to be due to unobservable common factors that affect to all cross section units at the same time, but also by some block-specific factors that drive a cross-sectional dependence only in that specific block without affecting the remaining blocks. Hence, as special cases, the model simplifies to that proposed by Pesaran (2006) and Bai (2009): i) When there is only one block ($R = 1$), i.e. only one country or only one economic sector. ii) When there are no block-specific factors, i.e. when $\sum_{r=1}^{R} \mathbf{r}_{F_r} = 0$. In these cases, a pervasive top-level common factor would completely drive the cross-sectional dependence in the panel even if the panel data is composed by blocks of data. Naturally, the present framework also renders a variety of panel data models as special cases as discussed in Pesaran (2006).

The main interest lies more on the estimation of $\beta_{r,i}$ than the estimation of the common component specified by (3) as in the case of Bai (2009) and Greenaway McGrevy et al. (2012). With M denoting a generic positive constant to indicate finiteness, the assumptions of the model are as follows:

Assumption A. Observed Common Effects:
The $R \cdot N \times 1$ vector of observed common effects $\mathbf{d}_{r,t}$ is covariance stationary with absolute summable autocovariances, distributed independently of the individual-specific errors $\epsilon_{r,it}$ and $\mathbf{v}_{r,it}$. Each $\mathbf{d}_r$ is orthogonal to G.

Assumption B. Unobserved Common Factors:

B_1 Block-specific factors $F_{r,t}$ are covariance stationary such that $E||F_{r,t}||^4 \leq M < \infty$ with $T^{-1} \sum_{t=1}^{T} F_{r,t} F'_{r,t} \xrightarrow{p} \Sigma_{F_\mathbf{r}}$ for some $\mathbf{r}_F \times \mathbf{r}_F$ positive definite matrix $\Sigma_{F_\mathbf{r}}$ $\forall r = 1, \ldots, R$.

B_2 The pervasive top-level factor G_t is covariance stationary such that $E||G_t||^4 \leq M < \infty$ with $T^{-1}\sum_{t=1}^{T} G_t G_t' \xrightarrow{p} \Sigma_G$ for some $\mathbf{r}_G \times \mathbf{r}_G$ positive definite matrix Σ_G.

B_3 Define $H_t = \left[G_t', F_{r,t}'\right]'$. For a fixed $\mathbf{r}$, assume that $T^{-1}\sum_{t=1}^{T} H_t H_t' \xrightarrow{p} \sum_H$ for some positive-definite matrix $\sum_H$ with rank $\mathbf{r}_G + \mathbf{r}_1 + \cdots + \mathbf{r}_R$.

B_4 Factors have zero mean, and $\sum_{t=1}^{T} G_t F_{r,t}' = 0$ for $r = 1, \ldots, R$.

Assumption C. Individual-Specific Errors:

C_1 The idiosyncratic shocks, $\epsilon_{r,it}$, $r = 1, \ldots, R$, $i = 1, \ldots, N$, $t = 1 \ldots, T$, are independently across r, i, and t with zero mean and variance σ_i^2, and have a finite fourth-moment.

C_2 The idiosyncratic shocks, $\mathbf{v}_{r,it}$, $r = 1, \ldots, R$, $i = 1, \ldots, N$, $t = 1 \ldots, T$, are independently across r, i, and t with zero mean and variance $\Sigma_i > 0$, and $\sup_{r,it} E||\mathbf{v}_{r,it}||^4 < \infty$.

C_3 Furthermore, $\epsilon_{r,it}$ as well as $\mathbf{v}_{r,jt'}$ are distributed independently for all r, i, j, t, and t'. For each r and i, $\epsilon_{r,it}$ and $\mathbf{v}_{r,it}$ could follow linear stationary processes with absolute summable autocovariances.

Assumption D. Factor loadings:
The unobserved factor loadings $\lambda_{r,i}$, $\mu_{r,i}$, $\Lambda_{r,i}$, and $M_{r,i}$ are independently and identically distributed across r, i, and of the individual specific errors $\epsilon_{r,it}$ and $\mathbf{v}_{r,jt}$, the common observable factors $\mathbf{d}_{r,t}$ and the unobserved common factors $(G_r, F_{r,t})$ for all r, i, j, and t with fixed means λ, μ, Λ, and M, and finite variances. In particular,

D_1 $\lambda_{r,i}$ is either deterministic such that $||\lambda_{r,i}|| \leq M < \infty$, or it is stochastic such that $E||\lambda_{r,i}||^4 \leq M < \infty$. In the latter case, $N_r^{-1}\Lambda_r'\Lambda_r \xrightarrow{p} \Sigma_{\Lambda_r} > 0$ for an $\mathbf{r}_F \times \mathbf{r}_F$ non-random matrix Σ_{Λ_r} for all $r = 1, \ldots, R$ with a generic positive constant M.

D_2 $\mu_{r,i}$ is either deterministic such that $||\mu_{r,i}|| \leq M$, or it is stochastic such that $E||\mu_{r,i}||^4 \leq M < \infty$ with $N_r^{-1}\mu_r'\mu_r \xrightarrow{p} \Sigma_{\mu_r} > 0$ for an $\mathbf{r}_G \times \mathbf{r}_G$ non-random matrix Σ_{μ_r} for all $r = 1, \ldots, R$.

Assumption E. Random Slope Coefficients:
The slope coefficients $\beta_{r,i}$ follow the random coefficient model

$$\beta_{r,i} = \beta + \nu_{r,i} \sim IID\left(\mathbf{0}, \mathbf{\Omega_v}\right), \quad \text{for } i = 1, 2, \ldots, N \text{ and } r = 1, 2, \ldots, R,$$

where $||\beta|| < K$, $||\mathbf{\Omega_v}|| < K$, $\mathbf{\Omega_v}$ is $k \times k$ symmetric nonnegative definite matrix, and the random deviations $\mathbf{v_{r,i}}$ are distributed independently of $\lambda_{r,j}$, $\mu_{r,j}$, $\Lambda_{r,j}$, $M_{r,j}$, $\epsilon_{r,jt}$, $\nu_{r,jt}$, $\mathbf{d}_{r,t}$, $F_{r,t}$, and G_t for all r, i, j, and t.

Assumption F. Identification of $\beta_{r,i}$:
Identification of the slope coefficients are given by a two-step procedure of cross-sectional averages. In the first step, consider the cross-sectional averages of the individual-specific variables $\mathbf{z}_{r,it}$ in each one of R blocks separately. Define by $\bar{\mathbf{z}}_{r,it} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{z}_{r,jt}$, and let

$$\overline{\mathbf{W}}_r \quad = \quad \mathbf{I}_T - \overline{H}_r \left(\overline{H}_r' \overline{H}_r\right)^{-} \overline{H}_r', \tag{5}$$

and

$$\mathbf{W}_{\mathbb{F}_r} \quad = \quad \mathbf{I}_T - \mathbb{F}_r \left(\mathbb{F}_r' \mathbb{F}_r\right)^{-} \mathbb{F}_r', \tag{6}$$

where $\overline{H}_r = \left(\mathbf{D}_r, \overline{\mathbf{Z}}_r\right)$, with $\mathbf{D}_r = (\mathbf{d}_{r,1}, \ldots, \mathbf{d}_{r,T})'$ a $T \times N$ matrix on observed common factors, and $\overline{\mathbf{Z}}_r = (\bar{\mathbf{z}}_{r,1}, \ldots, \bar{\mathbf{z}}_{r,T})'$ is the $T \times (k+1)$ matrix of time observations on the cross-sectional averages for each region r and $\mathbb{F}_r = (\mathbf{D}_r, F_r)$ where $F_r = (F_{r,1}, F_{r,2}, \ldots, F_{r,t})'$ is $T \times \mathbf{r}_{F_r}$ data matrices on unobserved block-specific factors, respectively.

For the second step,

$$\underbrace{\mathbf{Z}^*}_{(k+1)\cdot R\cdot N\times T} = \big[\underbrace{(\mathbf{Z}_1}_{(k+1)\cdot N\times T} \underbrace{\overline{\mathbf{W}}_1)'}_{T\times T}, \left(\mathbf{Z}_2 \overline{\mathbf{W}}_2\right)', \ldots, \left(\mathbf{Z}_R \overline{\mathbf{W}}_R\right)' \big]' \tag{7}$$

and consider the cross-sectional average of the complete panel and let

$$\overline{\mathbf{W}}^* \quad = \quad \mathbf{I}_T - \overline{H}^* \left(\overline{H}^{*'} \overline{H}^*\right)^{-} \overline{H}^{*'}, \tag{8}$$

$$\overline{\mathbf{W}}_G \quad = \quad \mathbf{I}_T - \mathbf{G} \left(\mathbf{G}' \mathbf{G}\right)^{-} \mathbf{G}', \tag{9}$$

where $\overline{H}^* = \overline{\mathbf{Z}^*}$, with $\overline{\mathbf{Z}^*} = (\bar{\mathbf{z}}_1^*, \ldots, \bar{\mathbf{z}}_T^*)'$ is the $T \times (k+1)$ matrix of observations on the cross-sectional averages. To simplify notation, I denote $\overline{\mathbf{W}}_{r,t} G_t'$ as G_t^*, then $\mathbf{G}^* = (\mathbf{G}_1^*, \mathbf{G}_2^*, \ldots, \mathbf{G}_T^*)'$ is $T \times \mathbf{r}_G$ data matrix on unobserved top-level factors.

F_1 Identification of $\beta_{\mathbf{r},\mathbf{i}}$:
The $k \times k$ matrix $\hat{\mathbf{\Psi}}_{r,iT} = \left(\frac{\mathbf{X}_{r,i}^{*'} \overline{\mathbf{W}}^* \mathbf{X}_{r,i}^*}{T}\right)$ and $\hat{\mathbf{\Psi}}_{r,iG} = \left(\frac{\mathbf{X}_{r,i}^{*'} \mathbf{W}_G \mathbf{X}_{r,i}^*}{T}\right)$ are nonsingular, and $\hat{\mathbf{\Psi}}_{r,iT}$ $\hat{\mathbf{\Psi}}_{r,iG}$ have finite second-order moments for all r and i.

Most of the assumptions are based on those provided by Pesaran (2006) and Wang (2010). Assumption A describes the structure of the observed common factors. Different observed common effects, $\mathbf{d}_{r,t}$, are allowed for each block r in (4) making this approach suitable to consider some specific characteristics for each block involved in the system. Although it is possible to include deterministic trends in $\mathbf{d}_{r,t}$, appropriate scaling should need to be considered in the variables,

see Pesaran (2006). For the sake of simplicity, global observed factors are not considered, however these can be included in some empirical studies.

The Assumptions B_1-B_2 describe the structure of the unobservable common factors $F_{r,t}$ and G_t defined earlier and impose standard moment conditions. At this point, I follow Wang (2010) and Choi et al. (2016) to impose an I(0) stationarity assumption in both levels of unobservable common factors. I will relax this restrictive assumption in Section (5) by allowing for I(d) processes in $F_{r,t}$ and G_t following ideas proposed in Ergemen and Rodríguez-Caballero (2016). The rank condition in Assumption B_3 implies that different factors are not perfectly correlated. Assumption B_4 rules out any possibility of correlation between the pervasive and block-specific factors meaning that factors at different levels among blocks do not contain information of each other. Note that the usual normalization conditions to identify such factors are discarded because we do not need to estimate them. Identification conditions following assumptions in Bai (2003), extended by Wang (2010) and Choi et al. (2016) for the multi-level case, can be considered in the framework discussed in Bai (2009). This possibility deviates too much from the main target of this paper and is left for future research.

The usual structure for the individual-specific errors is provided in Assumption C. Assumption D_1 ensures that each block-specific factor $F_{r,jt}$ has a nontrivial contribution to the variance of $y_{r,t}$, $j = 1, \ldots, r_F$ whereas Assumption D_2 ensures that $G_{m,t}$ has nontrivial contribution to the variance of y_t, $m = 1, \ldots, r$. The latter means that G_t pervades all variables whereas $F_{r,jt}$ only act as an impact within region r. Assumption D indicates that factor loadings are independent to each other. Westerlund and Urbain (2013) show inconsistency in the pool version of CCE when the factor loadings are correlated.

Assumption E is also standard and extends Assumption 4 in Pesaran (2006) only by including blocks. It is worth mentioning that Assumption E can be easily relaxed allowing for $\beta_{r,i} = \beta_r + \nu_{r,i} \sim IID\left(\mathbf{0}, \mathbf{\Omega_v}\right)$ implying that the slope β can freely vary among regions but being the same within the specific block r.

Assumption F details the identification strategy of the slope coefficients. Identification is carried out by an extension of the methodology discussed in Pesaran (2006). I detail the strategy in the next section. I consider simple averages to simplify the exposition, nevertheless cross-section weighted averages satisfying some regular granularity conditions can be considered instead, see e.g. Bailey et al. (2015), and Chudik and Pesaran (2015).

3 Strategy to control the impact of the cross-sectional multi-level dependence

Pesaran (2006) suggests the Common Correlated Effects (CCE) estimation procedure that consists of filtering out the cross-sectional dependence in

$$y_{it} = \mathbf{x}_{it}'\beta + \mathbf{b}_i'\mathbf{f}_t + u_{it}, \tag{10}$$

by using the cross section averages of y_{it} and $\mathbf{x}_{it}$ as suitable proxies for the unobserved factors leading to

$$\bar{\mathbf{b}}'\mathbf{f}_t \approx \bar{y}_t - \bar{\mathbf{x}}_t'\beta.$$

This means that β can be consistently estimated by augmenting the pooled regression of y_{it} on $\mathbf{x}_{it}$ and their cross section averages.

The main difference between the proposed setup in this paper and that of Pesaran (2006) relies on the factor structure suggested in this paper in (3) in contrast to the standard in the literature as in (10). Pesaran's CCE approach may not be enough to filter out the full factor space and the consistency of conventional estimators can be affected since the extent of cross-sectional dependence may still be correlated with the regressors. To see why, consider applying Pesaran's CCE approach for the full panel $\mathbf{z}$ in (4). Intuitively $\bar{\mathbf{z}}$ could proxy the top-level factor G_t or a type of mixing-factor structure but it would not be able to filter out the block-specific cross section dependence introduced by $F_{r,t}$, which would still affect the estimation of $\beta_{r,i}$. Therefore, a simple cross-sectional average of $\mathbf{z}$ may not be a good proxy for the cross-sectional dependence under the setup proposed in this paper. Alternatively, another strategy could be to use only the block-specific cross-sectional averages separately, i.e. block-by-block averages. However, since G_t is driving the cross section dependence between blocks, the latter strategy would not be capturing the top-level dependence.

I propose an extended CCE procedure to filter out the full factor space involved in a model whose cross-sectional dependence is driven by a multi-level factor structure as in (4). I shall refer to such methodology as the Multi-Level Common Correlated Effect Estimators (MLCCE) and it consists of two steps in the case of the baseline model.

In the first step, consider separately each of the $\mathbf{z}_{r,it}$ vectors belonging to each block $r = 1, \ldots, R$. Note that the cross-sectional dependence in block r is only driven by the block-specific observable common factors, $\mathbf{d}_{r,t}$ as well as the unobservable common factors $F_{r,t}$. The pervasive common factors G_t do not play a role hitherto since none of the remaining blocks are considered. In this sense, $\mu_{r,i} = 0$ and $M_{r,i} = \mathbf{0}$ lead to $\mathbf{C}_{r,i} = \mathbf{0}$ for all $i = 1, \ldots, N$. Then, to simplify notation, (4) is re-written as

$$\mathbf{z}_{r,it} = \mathbf{B}_{r,i}'\mathbf{d}_{r,t} + \mathbf{D}_{r,i}'F_{r,t} + \mathbf{u}_{r,it}. \tag{11}$$

Here I follow Pesaran (2006) to discuss why the cross-sectional averages of the observables of the respective block r, $\bar{\mathbf{z}}_{r,it}$, can work as a proxy variable for these block-specific factors. Consider the cross section averages on (11), and recall that r only denotes the specific block,

$$\bar{\mathbf{z}}_{r,t} = \bar{\mathbf{B}}_r'\mathbf{d}_{r,t} + \bar{\mathbf{D}}_r'F_{r,t} + \bar{\mathbf{u}}_{r,t}. \tag{12}$$

Assuming

$$\text{rk}\left(\bar{\mathbf{D}}\right) = \mathbf{r}_{F_r} < k + 1 \;\; \forall N, \tag{13}$$

then

$$F_{r,t} = \left(\bar{\mathbf{D}}\bar{\mathbf{D}}'\right)^{-1}\bar{\mathbf{D}}\left(\bar{\mathbf{z}}_{r,t} - \bar{\mathbf{B}}_r'\bar{\mathbf{d}}_{r,t} - \bar{\mathbf{u}}_{r,t}\right).$$

Lemma 1 in Pesaran (2006) shows that $\bar{\mathbf{u}}_{r,t} \overset{q.m.}{\to} \mathbf{0}$ as $N \to \infty$, for each t, which implies

$$F_{r,t} - \left(\mathbf{D}\mathbf{D}'\right)^{-1}\mathbf{D}\left(\bar{\mathbf{z}}_{r,t} - \bar{\mathbf{B}}_r'\bar{\mathbf{d}}_{r,t}\right) \overset{q.m.}{\to} 0, \text{ as } N \to \infty,$$

where

$$\mathbf{D} = \lim_{N\to\infty}\left(\bar{\mathbf{D}}\right) = \tilde{\mathbf{\Lambda}}\begin{pmatrix} 1 & 0 \\ \beta_{r,i} & \mathbf{I}_k \end{pmatrix},$$

with $\tilde{\mathbf{\Lambda}} = E\left(\lambda_{r,i}, \Lambda_{r,i}\right) = \left(\lambda_{r,i}, \Lambda_{r,i}\right)$, and $\beta_r = E\left(\beta_{r,i}\right)$.

Therefore, the block-specific factors, $F_{r,t}$, can be approximated separately by a linear combination in the respective block r of observed common factors, $\mathbf{d_r}$, the cross-sectional averages of the dependent variable, $\bar{y}_{r,t}$, and those of the individual-specific regressors, $\bar{\mathbf{x}}_{r,t}$. Then, let $\mathbf{Z}^*$ as in (7),

$$G^* = \left(G\,\overline{\mathbf{W}}_r\right)', \tag{14}$$

and $\mathbf{U}^* = \left[\left(\mathbf{U}_1\overline{\mathbf{W}}_1\right)', \ldots, \left(\mathbf{U}_R\overline{\mathbf{W}}_R\right)'\right]'$, where $\overline{\mathbf{W}}_r$ is defined by (5), then

$$\mathbf{z}_{it}^* = \mathbf{C}_i'G_t^* + \mathbf{u}_{it}^*, \tag{15}$$

after partialing out the effects of the block-specific factors from all the blocks $r = 1, \ldots, R$ by using the orthogonal projection matrix $\overline{\mathbf{W}}_r$ in each block separately. The cross-section averages of 15 will be useful to obtain some asymptotic results, and they are given by

$$\bar{\mathbf{z}}_t^* = \bar{\mathbf{C}}'G_t^* + \bar{\mathbf{u}}_t^*. \tag{16}$$

In the second step, the same reasoning suggests that $\overline{H}^* = \overline{\mathbf{Z}^*}$ is an observable proxy for the pervasive top-level unobserved factor G_t.

Under the rank condition in (13) and Assumptions A-C the following relevant result for the asymptotic analysis can be obtained

$$\overline{\mathbf{W}}_r F_r \approx \mathbf{W}_{\mathbb{F}_r} F_r = 0. \tag{17}$$

Such a property means that both projection matrices can be used interchangeably to partialing out the block-specific factor in the asymptotics. A similar argument can be followed in the case of the top-level factor G_t after removing the sub-level factors and as long as the assumption

$$\text{rk}\left(\bar{\mathbf{C}}\right) = \mathbf{r}_G < k + 1 \;\; \forall N, \tag{18}$$

is satisfied as well. Note that the current setup is already considering observable common factors for each block separately. In principle, one could consider that this setup is more suitable for practical purposes, however one can include observed global factors as well in the covariates and should be filtering out in the second step of the methodology by enlarging $\overline{\mathbf{W}}^*$ defined by (8) with the cross-sectional averages of such observed global factors.

4 Estimation and asymptotic inference

Estimates of $\beta_{r,i}$ can be obtained running standard factor-augmented panel regressions with these cross-sectional averages depending on the assumption regarding the slope homogeneity. These estimators can be obtained from the mean group and pooled versions of Pesaran (2006). For the sake of brevity, I only discuss the mean group estimators although the analysis can be easily extended to pooled estimators.

The Common Correlated Effects Mean Group (CCEMG) is a simple average of the individual MLCCE estimators, $\hat{\beta}_{r,i}$,

$$\hat{\beta}_{\text{CCEMG}} = R^{-1}N^{-1}\sum_{r=1}^{R}\sum_{i=1}^{N}\hat{\beta}_{r,i}, \tag{19}$$

where

$$\hat{\beta}_{r,i} = \left(\mathbf{X}_{r,i}^{*\prime}\overline{\mathbf{W}}^{*}\mathbf{X}_{r,i}^{*}\right)^{-1}\mathbf{X}_{r,i}^{*\prime}\overline{\mathbf{W}}^{*}y_{r,i}^{*}, \tag{20}$$

with

$$\begin{array}{rcl}\mathbf{X}_{r,i}^{*} & = & \left(\mathbf{x}_{r,i1}\overline{W}_{r,1}, \mathbf{x}_{r,i2}\overline{W}_{r,2}, \ldots, \mathbf{x}_{r,iT}\overline{W}_{r,T},\right), \\ y_{r,i} & = & \left(y_{r,i1}\overline{W}_{r,1}, y_{r,i2}\overline{W}_{r,2}, \ldots, y_{r,iT}\overline{W}_{r,T}\right),\end{array} \tag{21}$$

$\overline{\mathbf{W}}_r$ and $\overline{\mathbf{W}}^{*}$ defined according to (5) and (8), respectively.

The next theorems present the consistency and the associated asymptotic normality of $\hat{\beta}_{r,i}$ and those of $\hat{\beta}_{\text{CCEMG}}$.

Theorem 4.1. *Under Assumptions A-C, and F_1, as $(N,T)_j \to \infty$ and the rank conditions (13)-(18), then $\hat{\beta}_{r,i}$ is a consistent estimator of $\beta_{r,i}$. Furthermore, assuming $\frac{\sqrt{T}}{N} \to 0$ as $(N,T)_j \to \infty$ in the block r, then*

$$\sqrt{T}\left(\hat{\beta}_{r,i} - \beta_{r,i}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Sigma}_{\beta_{r,i}}\right),$$

where $\mathbf{\Sigma}_{\beta_{r,i}} = \sigma_{r,i}^{2}\mathbf{\Sigma}_{r,i\mathbf{v}}^{-1}(0)\mathbf{\Sigma}_{r,i\mathbf{v}}^{-1}(0)$.

As discussed above, the CCEMG estimator is defined as the average of the individual $\hat{\beta}_{r,i}$. Note that it is possible to consider a CCEMG estimator either for the specific block r or for the full panel. I focus on the more general setting.

Theorem 4.2. *Under Assumptions A-E, and F_1, then $\hat{\beta}_{CCEMG}$ is asymptotically unbiased for β for fixed R, and T and as $N \to \infty$. Furthermore, as $(N,T)_j \to \infty$ with R fixed,*

$$\sqrt{R\,N}\left(\hat{\beta}_{CCEMG} - \beta\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Sigma}_{CCEMG}\right),$$

where $\mathbf{\Sigma}_{CCEMG}$ can be consistently estimated non-parametrically by

$$\mathbf{\Sigma}_{CCEMG} = \frac{1}{R-1}\frac{1}{N-1}\sum_{r=1}^{R}\sum_{i=1}^{N}\left(\hat{\beta}_{r,i} - \hat{\beta}_{CCEMG}\right)\left(\hat{\beta}_{r,i} - \hat{\beta}_{CCEMG}\right)^{\prime}. \tag{22}$$

Note that the block-specific and global rank conditions, (13) and (18) respectively, are necessary in the Theorem (4.1) but not in (4.2).

It is well-known that in situations of slope homogeneity, $\beta_{r,i} = \beta$ for all r and i, pooled estimators would help to gain efficiency. Pesaran (2006) proposes the Common Correlated Effects Pooled Estimators (CCEP). Similar asymptotic analysis can be done in this regard under this setup with further regularity conditions. I avoid these details to focus only on the heterogeneity assumption which seems to be more appropriate in macroeconomics and financial applications.

5 The model with long-range dependence

There are some studies on dealing with panel data with cross-sectional dependence under Pesaran's framework where both of observable and unobservable common factors can be nonstationary. Kapetanios et al. (2011) consider the case when unobservable factors follow a unit-root process whereas the regression errors stay as I(0) processes. And very recently, Ergemen and Velasco (2015) consider a fractionally integrated framework.

In this section, I explore how the model specified by (1-2) behaves when the factor structure that drives the cross-sectional dependence and shocks exhibit long-range dependence. In this sense, the approach is more flexible because it is not restricted only to the I(0) or I(1) cases. This fractional approach could be helpful to study the relationships found in a complex system and can be applied to panel data that consist of blocks of data of typical economic time series that have been shown to exhibit long-range dependence such as aggregate ouput, real exchange rates, realized volatility, and electricity prices, to mention a few. In this section, I propose to extent the model discussed in the Section (2) to panels where observable and unobservable factors may follow I(d) processes.

The model I consider is a type-II fractionally integrated panel data model with a multi-level cross-sectional dependence and is given by

$$\begin{aligned} y_{r,it} &= \beta_{r,i}' x_{r,it} + \mu_{r,i}' G_t + \lambda_{r,i}' F_{r,t} + \Delta_t^{-d_{r,i0}} \epsilon_{r,it}, \\ x_{r,it} &= M_{r,i}' G_t + \Lambda_{r,i}' F_{r,t} + \Delta_t^{-\delta_{r,i0}} \mathbf{v}_{r,it}, \end{aligned} \tag{23}$$

where $r = 1, \ldots, R$, $i = 1, \ldots, N$, $t = 1, \ldots, T$ as before. $\mathbf{x}_{r,it}$, G_t, and $F_{r,t}$ define the observed individual-specific regressors, the unobserved top-level common effects, and the unobserved block-specific common effects, respectively. R, k, $\mathbf{r}_G$, $\mathbf{r}_{F_r}$ are assumed to be fixed as before.

In (23), in contrast to the stationary setup previously proposed, long memory in both idiosyncratic terms as well as in both levels of factors are allowed by assuming that

$$G_t = \Delta_t^{-\varrho_0} \omega_t, \qquad F_{r,t} = \Delta_t^{-\vartheta_{r,0}} \nu_{r,t}, \tag{24}$$

where $\omega_t, \nu_{r,t}, \epsilon_{r,it}$, and, $\mathbf{v}_{r,it}$ are zero-mean unobservable white noise sequences and the truncated fractional differencing filter $\Delta_t^{-\zeta}$ allows for the study of both the

stationary case ($\zeta < 1/2$) and the nonstationary case ($\zeta \geq 1/2$) unlike the untruncated filter that does not converge when $\zeta \geq 1/2$, see Davidson and Hashimzade (2009). $\Delta_t^{-\zeta}$ is described as follows. With $\Delta = 1 - L$, $\Delta^{-\zeta}$ has the expansion

$$\Delta^{-\zeta} = \sum_{j=0}^{\infty} \pi_j(-\zeta) L^j, \quad \text{where } \pi_j(-\zeta) = \frac{\Gamma(j+\zeta)}{\Gamma(j+1)\Gamma(\zeta)},$$

for $\zeta > 0$ with $\Gamma(\tau) = \infty$ for $\tau = 0, -1, \ldots$, but $\Gamma(0)/\Gamma(0) = 1$. $\Delta_t^{-\zeta}$ truncates the latter expansion to $\Delta_t^{-\zeta} = \sum_{j=0}^{t} \pi_j(-\zeta) L^j$.

Note that (23) does not incorporate fixed or observed common effects, $\mathbf{d}_{r,t}$, as in (1-2) to focus only on the main difference with respect to the stationary setup. The framework of the fractional integrated model is in nature the same as that proposed by Ergemen and Velasco (2015), however the present approach is slightly different in two aspects. First, I extend the analysis considering R blocks of panels and not only one. Second, I consider that the cross-sectional dependence in the R blocks is introduced by the multi-level structure specified by (3) extending their approach that considers only pervasive common factors.

The Assumptions (A-F) are maintained in the fractional setup. The pervasive and block-specific factors in (24) still have finite fourth-order moments and positive definite covariance matrices and are orthogonal to each other. Assumptions (C-D) are held identical. Identification provided by Assumption (F) is held after fractional differencing as discussed later. Nevertheless, it is necessary to add some conditions regarding the fractional memory parameters in (23) to extend the previous analysis. Such conditions are

Assumption G. Fractional integration parameters:

G_1 Idiosyncratic shocks ($\epsilon_{r,it}$): $d_{r,i0}$ takes values on the compact set $\mathcal{D} = [\underline{d}_{r,i0}, \overline{d}_{r,i0}]$ with $0 \leq \underline{d}_{r,i0} < \overline{d}_{r,i0} < 3/2$.

G_2 Idiosyncratic shocks ($\mathbf{v}_{r,it}$): $\max_{r,i} \delta_{r,i0} < 3/2$.

G_3 Top-level factor: $\varrho_0 < 3/2$.

G_4 Block-specific factor: $\vartheta_{r,0} < 3/2$.

Furthermore the following conditions are required. Let $d_{\max} = \max_{r,i} d_{r,i0}$, $\vartheta_{\max} = \max_{r,i} \vartheta_{r,i0}$, and $\delta_{\max} = \max_{r,i} \delta_{r,i0}$, then

G_5 $\max\{d_{\max}, \delta_{\max}, \vartheta_{\max}, \varrho\} - \underline{d} < 1/2$, and

G_6 $\max\{d_{\max}, \delta_{\max}\} < 5/4$.

Assumptions (G_1-G_4) are based on those provided by Ergemen and Velasco (2015) and impose standard restrictions in fractional integration on the range of allowed values for memory parameters relaxing the usual I(0) and I(1) restrictions imposed commonly in the literature. Although the range of the fractional

parameters covers slightly beyond the unit-root, for most of the applications it is frequently enough to consider the fractional memory until the $d \leq 1$. The residual memory estimates $(\hat{d}_{r,i})$ are only implicitly defined and entail optimization over $\Theta = \mathcal{D} \times \Xi$, where Ξ is a compact subset of $\Re^p$ and $\mathcal{D} = [\underline{d}_{r,i0}, \overline{d}_{r,i0}]$ with $0 < \underline{d}_{r,i0} < \overline{d}_{r,i0} < 3/2$. Even if a large range of values of $d \in \mathcal{D}$ are covered, there are some necessary requirements on the interplay between the fractional memories of the unobservable common factors and the idiosyncratic shocks (Assumptions (G_5 and G_6)). These assumptions are necessary for the asymptotic analysis of CSS estimates as I show in detail later.

In this setup, the persistence in the observed $y_{r,it}$ is absorbed by several channels and can be set by $\max(d_{r,i0}, \delta_{r,i0}, \varrho_0, \vartheta_{r,0})$ in each r and i. The model guarantees a fractionally cointegrating relationship in the cross section unit i within the region r when $\max(\delta_{r,i0}, \varrho_0, \vartheta_{r,0}) > d_{r,i0}$. It will be desirable that such a cointegration relationship may be given by cases where $\delta_{r,i0} > d_{r,i0}$, being a consequence of $\delta_{r,i0} > \max(\varrho_0, \vartheta_{r,0})$. This is because the underlying idea in the methodology is to filter out the factor structure that is driving the cross-sectional dependence in $y_{r,it}$ in order to establish causal relationships between $y_{r,it}$ and $x_{r,it}$. So, even if one is able to project out the factor structure in cases when $\delta_{r,i0} < \max(\varrho_0, \vartheta_{r,0})$, the cointegration analysis will be cumbersome since that would indicate that there exists a cointegration relationship between $y_{r,it}$ and G_t, for instance. For practical economic purposes or for forecasting for instance, what would be desirable is that covariates exhibit more persistence than unobservable common factors in order to ensure cointegration among observable variables.

As pointed out before, the main interest is not the identification of the multilevel structure but on the estimation of $\beta_{r,i}$. Note that in the setups proposed in this paper, the factor structure in (23) only controls for the cross-sectional dependence in contrast to Ergemen and Rodríguez-Caballero (2016) who focus on the identification of global and regional unobservable factors allowing for fractional cointegration between $y_{r,it}$ and G_t or $F_{r,t}$, however their setup do not consider explanatory variables as in the present paper.

I follow the estimation methodology of Ergemen and Velasco (2015) for each $\beta_{r,i}$. The underlying reasoning is to estimate (23) after fractional differencing by a fractional memory parameter denoted by $d^*_{r,i}$ approximately at the level of $d_{r,i0}$. In other words, when differencing by $d^*_{r,i}$ one gets

$$\Delta_t^{d^*_{r,i}} y_{r,it} = \beta'_{r,i} \Delta_t^{d^*_{r,i}} x_{r,it} + \mu'_{r,i} \Delta_t^{d^*_{r,i}} G_t + \lambda'_{r,i} \Delta_t^{d^*_{r,i}} F_{r,t} + \Delta_t^{d^*_{r,i}} \Delta_t^{-d_{r,i0}} \epsilon_{r,it}, \quad (25)$$

and naturally $\Delta_t^{d^*_{r,i} - d_{r,i0}} \epsilon_{r,it} \approx I(0)$ when $d^*_{r,i}$ is well calibrated ($d^*_{r,i} \approx d_{r,i0}$).

The full factor structure is filtered out by using (5) and (8) with the fractionally differenced cross-sectional averages following the methodology explained in Section(3). Define the projection matrices (5) and (8) as

$$\overline{\mathbf{W}}_r \left(d^*_{r,i}\right) = \mathbf{I}_T - \overline{H}_r \left(d^*_{r,i}\right) \left(\overline{H}'_r \left(d^*_{r,i}\right) \overline{H}_r \left(d^*_{r,i}\right)\right)^{-} \overline{H}'_r \left(d^*_{r,i}\right), \quad (26)$$

$$\overline{\mathbf{W}}^{*}\left(d_{r,i}^{*}\right)=\mathbf{I}_{T}-\overline{H}^{*}\left(d_{r,i}^{*}\right)\left(\overline{H}^{*^{\prime}}\left(d_{r,i}^{*}\right)\overline{H}^{*}\left(d_{r,i}^{*}\right)\right)^{-}\overline{H}^{*^{\prime}}\left(d_{r,i}^{*}\right), \tag{27}$$

where

$$\overline{H}_{r}(d_{r,i}^{*})=\overline{\mathbf{Z}}(d_{r,i}^{*})\quad\text{with }\overline{\mathbf{Z}}(d_{r,i}^{*})=\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\Delta_{t}^{d_{r,i}^{*}}\mathbf{Z}(d_{r,i}^{*}),\text{ and}$$

$$\overline{H}^{*}(d_{r,i}^{*})=\overline{\mathbf{Z}}^{*}(d_{r,i}^{*})\quad\text{with }\overline{\mathbf{Z}}^{*}(d_{r,i}^{*})=\tfrac{1}{R}\tfrac{1}{N}\textstyle\sum_{r=1}^{R}\sum_{i=1}^{N}\Delta_{t}^{d_{r,i}^{*}}\mathbf{Z}^{*}(d_{r,i}^{*}).$$

As before, the projection matrices (26) and (27) can be used to partial out both the block-specific differenced factor, firstly, and the global differenced factor, secondly, in the asymptotics as discussed before.

The CCEMG estimator is then defined as

$$\hat{\beta}_{\text{CCEMG}}\left(d_{r,i}^{*}\right)=R^{-1}N^{-1}\sum_{r=1}^{R}\sum_{i=1}^{N}\hat{\beta}_{r,i}\left(d_{r,i}^{*}\right), \tag{28}$$

with

$$\hat{\beta}_{r,i}\left(d_{r,i}^{*}\right)=\left(\mathcal{X}_{r,i}^{*\prime}\overline{\mathbf{W}}^{*}\left(d_{r,i}^{*}\right)\mathcal{X}_{r,i}^{*}\right)^{-1}\mathcal{X}_{r,i}^{*\prime}\overline{\mathbf{W}}^{*}\left(d_{r,i}^{*}\right)y_{r,i}^{*}, \tag{29}$$

where $\mathcal{X}_{r,i}^{*}=\Delta_{t}^{d_{r,i}^{*}}X_{r,i}^{*\prime}$, and $y_{r,i}^{*}=\Delta_{t}^{d_{r,i}^{*}}y_{r,i}^{*\prime}$ with $\mathcal{X}_{r,i}^{*}$, and $y_{r,i}^{*}$ defined as before in (21).

Consistency of $\hat{\beta}_{r,i}(d_{r,i}^{*})$ depends on the selection of $d_{r,i}^{*}$. The idea behind the estimation procedure is that all the variables in (25) are asymptotically stationary. Consequently, taking $d_{r,i}^{*}=1$ would be enough in cases when all these variables are close to the unit root as in the PANIC model of Bai and Ng (2004). Furthermore, all detrended variables will be asymptotically stationary when taking $d_{r,i}^{*}\geq 1$ whereas $\delta_{r,i}+d_{r,i0}-2d_{r,i}^{*}<1$ no matter the values on $\delta_{r,i}$ or $d_{r,i0}$.

Ergemen and Velasco (2015) use the latter condition to achieve the consistency of $\hat{\beta}_{r,i}(d_{r,i}^{*})$ no matter the divergence rates of N or T. Following them, it is also possible to guarantee asymptotic normality through $d_{r,i}^{*}\geq 1$ and the following assumption

Assumption H. Restriction of fractional integration parameters for asymptotic normality:

H_1 $\delta_{r,i}+d_{r,i0}-2d_{r,i}^{*}<1/2$,

H_2 $\max(\delta_{\max},d_{\max})<11/8$, and

H_3 $\max(\max(\vartheta_r,\varrho)+\delta_{\max},\max(\vartheta_r,\varrho)+d_{\max},\delta_{\max}+d_{\max})<11/4$.

The next theorem summarizes this discussion.

Theorem 5.1. *Define the pervasive and block-specific factors as in (24) with fractional parameters as in G_3 and G_4, respectively, and ω_t and $\nu_{r,t}$ are white noise processes with finite fourth-moments. Orthogonality between both factors still applies, Assumptions (B_3-B_4). The idiosyncratic shocks are defined in the Assumptions C_1-C_3 with fractional parameters defined by the Assumptions (G_1, G_2). Then choosing $d^*_{r,i} \geq 1$ and under the rank conditions (13)-(18), as $(N,T)_j \to \infty$,*

$$\hat{\beta}_{r,i}(d^*_{r,i}) \xrightarrow{p} \beta_{r,i}. \tag{30}$$

Furthermore, assuming $\frac{\sqrt{T}}{N} \to 0$ as $(N,T)_j \to \infty$ in the block r, and Assumption H and $d^*_{r,i} \geq 1$, then

$$\sqrt{T}\left(\hat{\beta}_{r,i}(d^*_{r,i}) - \beta_{r,i}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Sigma}_{\beta_{r,i}}\right), \tag{31}$$

and with R fixed,

$$\sqrt{RN}\left(\hat{\beta}_{CCEMG}(d^*_{r,i}) - \beta\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Sigma}_{CCEMG}\right), \tag{32}$$

where $\mathbf{\Sigma}_{\beta_{r,i}} = \sigma^2_{r,i}\mathbf{\Sigma}^{-1}_{r,i\mathbf{v}}(0)\mathbf{\Sigma}^{-1}_{r,i\mathbf{v}}(0)$, and $\mathbf{\Sigma}_{CCEMG}$ can be consistently estimated non-parametrically by (22) as before.

To estimate the fractional memory parameters $d_{r,i}$, a conditional-sum-of-square (CSS) estimation based on the regression residuals is adopted. Let $\hat{d}_{r,i}$ denote the estimate of the unknown true fractional integration parameter $d_{r,i0}$ and is given by

$$\hat{d}_{r,i} = \underset{d_{r,i}\in\mathcal{D}}{\operatorname{argmin}} L^*_{r,i,T}\left(d_{r,i}\right), \tag{33}$$

where $\mathcal{D} = [\underline{d}_{r,i0}, \overline{d}_{r,i0}] \subseteq (0, 3/2)$, and

$$L^*_{r,i,T}\left(d_{r,i}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\Delta_t^{d_{r,i}}\hat{\epsilon}_{r,it}\right)^2, \tag{34}$$

with

$$\hat{\epsilon}_{r,it} = \tilde{y}_{r,it}(d^*_{r,i}) - \hat{\beta}'_{r,i}(d^*_{r,i})\tilde{\mathbf{X}}_{r,i}(d^*_{r,i}),$$

where $\tilde{y}_{r,it}(d^*_{r,i}) = y^*_{r,i}\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)$, $\tilde{\mathbf{X}}_{r,i}(d^*_{r,i}) = \mathcal{X}^*_{r,i}\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)$ with $y^*_{r,i}, \mathcal{X}^*_{r,i}$, and $\overline{\mathbf{W}}^*(d^*_{r,i})$ defined as before, and $\hat{\beta}'_{r,i}(d^*_{r,i})$ is given by (29).

In the next theorem the asymptotic results for residual memory estimates are established.

Theorem 5.2. *Under conditions of Theorem (5.1) and Assumptions (G_5 and G_6), $d_{r,i} \in Int(\mathcal{D})$ as $(N,T)_j \to \infty$, $\hat{d}_{r,i} \xrightarrow{p} d_{r,i0}$, and*

$$T^{1/2}\left(\hat{d}_{r,i} - d_{r,i0}\right) \xrightarrow{d} N\left(0, 6/\pi^2\right).$$

Note that the $\sqrt{T}$-consistency of the memory estimate is also guaranteed under the conditions of Theorem (5.2).

6 Monte Carlo analysis

In this section I examine the finite-sample properties of the proposed procedure to investigate the performance of the models specified by (1-2), and (23). In the first set of Monte Carlo studies, (Experiments 1-4), I compare the performance of the CCEMG estimator given by (19) and that proposed in Pesaran (2006). I also include a couple of robustness checks to analyze the performance of the I(0) model in cases when i) ranks conditions (13 and 18) are violated (Experiment 3), and ii) cases when there are structural breaks in the means of the top-level and block-specific factors (Experiment 4). In the second set, (Experiments 5-8), I investigate the CCEMG estimator given by (28) and the performance of the residual memories estimators $\hat{d}_{r,i}$ under several cases including fractional cointegration and non-cointegration schemes. I compare the present methodology with that proposed by Ergemen and Velasco (2015). In both sets of Monte Carlo simulations I include the CCEP estimator and cases when $\beta_{r,i} = \beta$ or $\beta_{r,i} \neq \beta \; \forall \; r = 1, \ldots, R$ and $i = 1, \ldots, N$. I report summaries of the estimators in terms of averages biases and root mean square errors in all cases. All results are based on 1000 replications.

6.1 Designs for I(0) cases

6.1.1 Experiment 1

For this Monte Carlo design I use the following data-generating process (DGP):

$$\begin{array}{rcl} y_{r,it} & = & \beta_{r,i}' x_{r,it} + \mu_{\mathbf{y}_{r,i}}' G_t + \lambda_{\mathbf{y}_{r,i}}' F_{r,t} + \epsilon_{r,it}, \\ x_{r,it} & = & \mu_{\mathbf{x}_{r,i}}' G_t + \lambda_{\mathbf{x}_{r,i}}' F_{r,t} + \upsilon_{r,it}, \end{array} \tag{35}$$

for $r = 1, 2$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$ with $(N, T) \in \{20, 50, 100, 500, 2000\}$. One top-level factor is taken to characterize the cross-sectional dependence between blocks and one block-specific factor in each block. $G_t, F_{1,t},$ and $F_{2,t}$ are generated as independent stationary AR(1) processes with zero means and unit variances. Autoregression coefficients in all the factors are fixed at 0.5. Top-level and block-specifics loadings are generated as $N(1, 1)$. $(\epsilon_{r,it}, \upsilon_{r,it}) \sim IIDN(0, \Sigma)$ with the covariance matrix Σ equals to the identity matrix. I consider slope homogeneity and slope heterogeneity by fixing $\beta_{r,i} = 1$ and $\beta_{r,i} = 1 + w_{r,i}$ with $w_{r,i} \sim N(0, 0.5)$, respectively. I also consider the CCEP estimator after filtering out the full structure following the procedure proposed in this paper as in the case of CCEMG estimator. I include the CCEMG and CCEP estimators following the methodology of Pesaran (2006) only for the case of slope homogeneity. The results are reported in Table 1.

Table 1: Small-sample properties of CCEMG and CCEP in the case of Experiment 1 (Full ranks + one regressor). Bias and RMSE are multiplied by 100 in the report.

		Slope homogeneity								Slope heterogeneity			
		Pesaran (2006)				MLCCE				MLCCE			
		CCEMG		CCEP		CCEMG		CCEP		CCEMG		CCEP	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
20	20	18.32	18.46	18.70	18.94	0.05	5.34	-0.06	6.31	0.02	6.34	0.31	6.37
	50	18.65	18.73	19.22	19.40	-0.09	4.57	-0.15	5.83	-0.14	5.43	0.01	5.89
	100	18.69	18.73	19.51	19.63	0.08	4.23	0.02	5.56	0.13	5.02	0.26	5.61
	500	18.49	18.51	19.14	19.26	0.07	4.25	-0.07	5.54	-0.08	5.05	-0.01	5.59
	2000	18.57	18.62	19.41	19.54	0.09	4.42	0.23	5.83	-0.19	5.25	-0.27	5.89
50	20	17.47	17.49	17.65	17.72	-0.02	3.30	0.10	3.82	-0.04	3.92	0.00	3.86
	50	18.26	18.26	18.82	18.82	-0.01	3.02	0.07	3.75	0.18	3.59	0.17	3.79
	100	18.29	18.29	19.04	19.07	0.11	2.81	0.33	3.70	-0.13	3.34	-0.20	3.74
	500	18.52	18.52	19.08	19.09	0.21	2.60	0.25	3.50	-0.17	3.09	-0.22	3.53
	2000	18.34	18.34	19.03	19.04	-0.09	2.65	-0.08	3.51	0.06	3.15	0.12	3.54
100	20	17.98	17.98	18.25	18.25	0.06	2.35	0.11	2.69	-0.08	2.79	-0.06	2.72
	50	18.44	18.44	18.83	18.83	-0.09	2.03	-0.07	2.60	-0.12	2.41	-0.23	2.62
	100	18.49	18.49	19.12	19.12	0.01	1.92	0.03	2.48	0.02	2.28	0.06	2.50
	500	18.46	18.46	19.17	19.17	-0.03	1.85	-0.05	2.51	0.05	2.20	0.07	2.53
	2000	18.39	18.39	19.12	19.12	0.03	1.88	0.04	2.41	-0.01	2.23	0.12	2.43
500	20	18.04	18.04	18.38	18.38	-0.01	1.11	-0.02	1.26	-0.01	1.32	-0.05	1.27
	50	18.33	18.33	18.88	18.88	-0.01	0.92	-0.06	1.17	0.03	1.09	-0.02	1.18
	100	18.45	18.45	19.00	19.00	-0.01	0.87	0.00	1.12	0.02	1.03	-0.03	1.13
	500	18.59	18.59	19.19	19.19	-0.01	0.85	0.02	1.12	0.01	1.01	0.05	1.13
	2000	18.54	18.54	19.14	19.14	-0.02	0.82	-0.05	1.08	0.01	0.97	-0.01	1.09
2000	20	18.03	18.03	18.34	18.34	-0.02	0.54	-0.03	0.61	0.02	0.64	0.04	0.62
	50	17.90	17.90	18.34	18.34	-0.02	0.45	-0.03	0.57	0.01	0.53	0.01	0.58
	100	18.68	18.68	19.23	19.23	0.01	0.43	-0.01	0.55	0.03	0.51	0.01	0.56
	500	18.68	18.68	19.27	19.27	0.01	0.42	-0.01	0.55	0.02	0.50	0.01	0.56
	2000	18.38	18.38	18.96	18.96	-0.02	0.40	-0.05	0.54	-0.03	0.47	-0.07	0.55

Notes: The DGP is $y_{r,it} = \beta'_{r,i}x_{r,it} + \mu_{\mathbf{y}'_{r,i}}G_t + \lambda_{\mathbf{y}'_{r,i}}F_{r,t} + \epsilon_{r,it}$. Regressors are generated by $x_{r,it} = \mu_{\mathbf{x}'_{r,i}}G_t + \lambda_{\mathbf{x}'_{r,i}}F_{r,t} + v_{r,it}$. $r = 1, 2$ and i and $t \in (20, 50, 100, 500, 2000)$. $\epsilon_{r,it} \sim IIDN(0,1)$ and $v_{r,it} \sim IIDN(0,1)$. Only one top-level factor and only one block-specific factor in each block are considered. $G_t = 0.5G_{t-1} + e_t^G$ with $e_t^G \sim IIDN(0,1)$. $F_{r,t} = 0.5F_{r,t-1} + e_t^{F_r}$ with $e_t^{F_r} \sim IIDN(0,1)$. Slope homogeneity is defined as $\beta_{r,i} = 1$ while slope heterogeneity as $\beta_{r,i} = 1 + w_{r,i}$ with $w_{r,i} \sim N(0, 0.5)$. CCEMG and CCEP estimators in the first two columns correspond to those proposed in Pesaran (2006). CCEMG and CCEP estimators in the columns titled MLCCE correspond to those proposed in this paper in (20). Rank conditions (13) and (18) are fulfilled. All experiments are based on 1000 replications.

6.1.2 Experiment 2

In this experiment I modify the DGP (35). Now I consider the case of four blocks, $r = 1, \ldots, 4$, two top-level and two block-specific factors in each block. I also add another co-variate in the model, $z_{r,it}$. Factors $G_{1_t}, G_{2_t}, F_{r_1,t}, F_{r_2,t}$, loading factors and idiosyncratic shocks $\epsilon_{r,it}, v_{r,it_x}, v_{r,it_z}$ are generated according to Experiment 1. In this experiment, slope homogeneity and slope heterogeneity are examined in both co-variates by fixing $\beta^x_{r,i} = 1$ and $\beta^x_{r,i} = 1 + w^x_{r,i}$ with $w^x_{r,i} \sim N(0, 0.5)$ for $x_{r,i}$ and $\beta^z_{r,i} = 10 + w^z_{r,i}$ with $w^z_{r,i} \sim N(0, 0.5)$ for $z_{r,i}$. Note that both rank conditions (13 and 18) are still satisfied. Only CCEMG estimators are reported. The results are reported in Table 2.

Table 2: Small-sample properties of CCEMG in the case of experiment 2 (Full ranks + two regressors). Bias and RMSE are multiplied by 100 in the report.

		Slope homogeneity				Slope heterogeneity			
		β^x_{CCEMG}		β^z_{CCEMG}		β^x_{CCEMG}		β^z_{CCEMG}	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
20	20	0.23	4.16	-0.05	4.22	0.13	6.72	-0.13	6.71
	50	-0.23	3.54	-0.10	3.60	0.01	5.72	0.35	5.73
	100	0.33	3.49	0.08	3.58	-0.01	5.64	-0.06	5.69
	500	0.03	3.47	-0.21	3.43	0.04	5.61	-0.21	5.46
50	20	-0.08	2.67	0.09	2.56	0.01	4.31	-0.01	4.07
	50	0.11	2.38	0.04	2.34	0.08	3.84	0.02	3.72
	100	0.02	2.26	0.05	2.16	0.05	3.65	0.05	3.44
	500	0.09	2.12	0.06	2.14	-0.08	3.42	-0.03	3.40
100	20	-0.12	1.84	0.00	1.87	0.12	2.97	0.10	2.97
	50	0.09	1.65	0.02	1.61	0.09	2.67	0.04	2.56
	100	-0.10	1.54	0.02	1.56	0.12	2.49	0.05	2.48
	500	0.09	1.48	0.01	1.57	0.12	2.39	0.03	2.50
500	20	-0.01	0.81	0.04	0.85	-0.01	1.31	-0.01	1.35
	50	0.04	0.74	0.00	0.70	0.07	1.20	-0.02	1.11
	100	-0.08	0.69	-0.03	0.71	-0.02	1.11	0.02	1.13
	500	-0.02	0.68	0.03	0.66	0.00	1.10	-0.01	1.05

Notes: The DGP is the same as that of Table 1, except that now $r = 1, \ldots, 4$ blocks are considered. Two top-level and two block-specific factors in each blocks are included. In this experiment the slope heterogeneity for the regressor $x_{r,it}$ is the same than before but that of $z_{r,it}$ is defined as $\beta^z_{r,i} = 10 + w^z_{r,i}$ with $w^z_{r,i} \sim N(0, 0.5)$. Rank conditions (13) and (18) are fulfilled. All experiments are based on 1000 replications.

6.1.3 Experiment 3

I provide evidence of the effects of violations in both rank conditions (13 and 18) from Experiment 1. I consider DGP (35), but an extra top-level common component $\mu_{\mathbf{2y}'_{r,i}} G_{t_2}$ and an extra block-specific common components $\lambda_{\mathbf{2y}'_{r,i}} F_{r,t_2}$ are added to the right-hand side of the model. The new common components follow the same characteristic of those in the Experiment 1. In this experiment, note that for the block $r = 1, 2$, Rank $\left(\bar{\mathbf{D}}\right) = \mathbf{r}_{F_r} = j > 2$ and Rank $\left(\bar{\mathbf{C}}\right) = \mathbf{r}_G = j > 2$ with $j = (3, 4)$ violating the rank conditions (13) and (18), respectively. The results are reported in Table 3.

Table 3: Small-sample properties of CCEMG in the case of Experiment 3 (Number of factors m=(3,4) exceed the number of regressors and regressand (k=2)). Bias and RMSE of CCEMG estimators are multiplied by 100 in the report.

		Slope homogeneity				Slope heterogeneity			
		$\mathbf{r}_G = \mathbf{r}_{F_r} = 3$		$\mathbf{r}_G = \mathbf{r}_{F_r} = 4$		$\mathbf{r}_G = \mathbf{r}_{F_r} = 3$		$\mathbf{r}_G = \mathbf{r}_{F_r} = 4$	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
20	20	0.10	6.79	0.11	4.85	-0.05	7.05	-0.25	6.82
	50	-0.42	6.10	-0.28	4.36	-0.14	6.33	0.09	6.12
	100	0.52	5.68	0.27	4.06	0.12	5.89	0.29	5.70
	500	0.04	5.32	-0.01	3.80	0.16	5.52	0.18	5.34
50	20	0.11	4.14	0.11	2.96	-0.22	4.30	-0.12	4.16
	50	0.11	3.65	-0.17	2.61	-0.09	3.79	-0.08	3.66
	100	0.01	3.61	0.07	2.58	-0.01	3.75	-0.12	3.62
	500	0.32	3.48	-0.16	2.49	-0.31	3.61	-0.22	3.49
100	20	-0.27	2.94	0.02	2.10	-0.01	3.05	-0.03	2.95
	50	-0.01	2.73	0.07	1.95	-0.08	2.83	0.12	2.74
	100	-0.02	2.66	0.00	1.90	0.18	2.76	0.03	2.67
	500	0.06	2.53	0.08	1.81	-0.09	2.63	0.06	2.54
500	20	0.10	1.36	-0.08	0.97	0.03	1.41	-0.03	1.37
	50	-0.02	1.15	-0.02	0.82	-0.01	1.19	-0.06	1.15
	100	0.00	1.16	0.07	0.83	-0.03	1.20	0.02	1.16
	500	-0.10	1.13	0.01	0.81	0.03	1.17	-0.06	1.13

Notes: The DGP is the same as that of Table 2, except that now extra block-specific and top-level factors are included to violate rank conditions (13) and (18). These new common factors follow the same DGP as that of Table 1. All experiments are based on 1000 replications.

6.1.4 Experiment 4

Following Stock and Watson (2009a) who suggest that structural breaks in the means of the common factors do not hamper the consistency of the CCE type estimators, I analyze the performance of the model (35) in presence of structural breaks in the means of i) block-specific factors, ii) pervasive top-level factors, and iii) in both. I consider the DGP (35), but now I generate the top-level and block-specific factors as follows:

1. $F_{r,t}$ generated as Experiment 1. With $\mathbb{G}_t$ generated as G_t in Experiment 1, I define the top-level factor as $G_t = \mathbb{G}_t$ for $t < [T/2]$ and $G_t = 5 + \mathbb{G}_t$ for $t \geq [T/2]$.

2. G_t generated as Experiment 1. With $\mathbb{F}_{r,t}$ generated as $F_{r,t}$ in Experiment 1, I define the block-specific factors as $F_{r,t} = \mathbb{F}_{r,t}$ for $t < [3T/4]$ and $F_{r,t} = 10 + \mathbb{F}_{r,t}$ for $t \geq [3T/4]$ for $r = 1, 2$.

3. With $\mathbb{G}_t$, and $\mathbb{F}_{r,t}$ generated as above and G_t, and $F_{r,t}$ as in Experiment 1.

Only CCEMG estimators are reported. The results are reported in Table 4.

6.2 Designs for I(d) cases

For this set of Monte Carlo simulations I use the following DGP:

$$\begin{aligned} y_{r,it} &= \beta'_{r,i} x_{r,it} + \mu'_{r,i} G_t + \lambda'_{r,i} F_{r,t} + \Delta_t^{-d_{r,i0}} \epsilon_{r,it}, \\ x_{r,it} &= \mu'_{r,i} G_t + \lambda'_{r,i} F_{r,t} + \Delta_t^{-\delta_{r,i0}} v_{r,it}, \end{aligned} \tag{36}$$

for $r = 1, 2$, $i = 1, \ldots, N \in \{20, 100, 200\}$ and $t = 1, \ldots, T \in \{150, 500, 2500\}$. One pervasive top-level factor and one block-specific factor in each region are considered for simplicity although more factors are possible as in the first sets of experiments (6.1). Both levels of factors and all the idiosyncratic terms are independently generated by ARFIMA(1,d*,0) processes where d* corresponds to $d_{r,i}$, $\delta_{r,i}$, ϑ_r or ϱ as appropriate. Autoregressive parameters are 0.5 in all cases. $\epsilon_{r,it} \overset{iid}{\sim} N(0, 2\phi)$ and $v_{r,it} \overset{iid}{\sim} N(0, 2\phi)$ are generated independently with ϕ controlling the signal-to-noise-ratio with $\phi = \{5, 2, 0.5\}$, corresponding to low, medium, and high signal-to-noise-ratios. All factor loadings are generated as $N(1, 1)$. I only consider slope homogeneity by fixing $\beta_{r,i} = 1$ but slope heterogeneity can be also investigated as before. I also consider the CCEP estimator after filtering out the full structure following the procedure proposed in this paper as in the case of CCEMG estimator. I include the CCEMG and CCEP estimators following the methodology of Ergemen and Velasco (2015) for comparison. I also present the average of the estimated residual integration orders by the CSS procedure as proposed before. For projection of estimated pervasive and block-specific factors based on prewhitened cross-sectional averages, I take $d^* = 1$. All results are based on 1000 replications of the model. The experiments focus mainly on the performance of model

Table 4: Small-sample properties of ECMG in the case of Experiment 4 (one break in the means of factors). Bias and RMSE of CCEMG estimators are multiplied by 100 in the report.

		Slope homogeneity						Slope heterogeneity					
		Break in G_t		Break in F_{rt}		Break in G_t and F_{rt}		Break in G_t		Break in F_{rt}		Break in G_t and F_{rt}	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
20	20	-0.170	6.130	-0.110	6.440	0.220	10.520	-0.140	6.988	0.230	6.895	0.720	9.357
	50	0.100	6.160	0.200	5.720	0.580	10.670	-0.440	7.022	0.160	6.124	-0.440	9.490
	100	-0.260	5.870	-0.030	5.610	0.960	9.560	-0.010	6.692	-0.030	6.006	-0.300	8.503
	500	0.090	5.560	0.030	5.650	-0.560	10.100	-0.070	6.338	-0.640	6.049	0.060	8.983
50	20	0.070	4.270	0.270	4.100	-0.290	6.410	-0.100	4.868	-0.030	4.390	0.070	5.701
	50	0.290	3.820	-0.410	3.550	0.120	6.200	-0.030	4.355	0.030	3.801	0.110	5.515
	100	-0.130	3.570	0.090	3.630	0.010	6.270	-0.080	4.070	0.260	3.886	-0.160	5.577
	500	0.000	3.380	0.180	3.340	0.300	6.100	-0.050	3.853	0.080	3.576	0.180	5.426
100	20	0.020	2.820	0.060	2.840	-0.220	4.380	0.020	3.215	0.080	3.041	0.070	3.896
	50	0.080	2.500	0.110	2.520	-0.060	4.260	-0.020	2.850	0.000	2.698	0.260	3.789
	100	0.100	2.670	-0.040	2.400	-0.030	4.330	0.100	3.044	0.160	2.570	-0.100	3.851
	500	0.060	2.470	0.110	2.440	0.170	4.400	-0.010	2.816	0.180	2.612	0.110	3.914
500	20	0.010	1.240	0.040	1.250	-0.010	2.010	0.120	1.414	0.110	1.338	-0.060	1.788
	50	0.060	1.150	0.070	1.120	0.080	1.950	-0.020	1.311	0.060	1.199	0.000	1.734
	100	0.000	1.110	0.060	1.100	-0.050	1.970	-0.050	1.265	0.000	1.178	0.020	1.752
	500	0.070	1.100	-0.090	1.070	0.080	2.010	-0.010	1.254	0.000	1.146	0.010	1.788

Notes: The DGP is the same as that of Table 1, except that now it is considered a break in the mean of i) block-specific factors, ii) pervasive top-level factors, and iii) in both. The top-level and block-specific factors are generated as $G_t = \mathbb{G}_t$ for $t < [T/2]$ and $G_t = 5 + \mathbb{G}_t$ for $t \geq [T/2]$, and $F_{r,t} = \mathbb{F}_{r,t}$ for $t < [3T/4]$ and $F_{r,t} = 10 + \mathbb{F}_{r,t}$ for $t \geq [3T/4]$ for $r = 1, 2$, as the case may be. Rank conditions (13) and (18) are fulfilled. All experiments are based on 1000 replications.

(36) when fractional parameters allow for cointegrating (Experiments $5\mathbb{A}-\mathbb{D}$) and non-cointegrating relationships (Experiment $6\mathbb{A}$). Note that in all experiments the multi-level factor structure that drives the cross-sectional dependence in the observables has long-range dependence. Experiments are summarized as follows:

1. Experiment $5\mathbb{A}$. Both levels of common factors are stationary. The pervasive top-level factor is slightly more persistent ($\varrho = 0.4$) than the block-specific factors ($\vartheta = 0.2$). $v_{r,it}$ follows a driftless I(1) process while the memory of the residuals of model (36) are $d_{r,it} < 0.75$. Fractional cointegration is guaranteed since $d_{r,it} < \delta_{r,it}$.

2. Experiment $5\mathbb{B}$. The only difference to the Experiment $5\mathbb{A}$ is that the top-level factor is now non-stationary with ($\varrho = 0.6$) but the block-specific factors are still stationary with ($\vartheta = 0.4$).

3. Experiment $5\mathbb{C}$. Persistences of G_t and $v_{r,it}$ are now more similar with $\varrho = 0.6$ and $\delta_{r,it} = 0.8$. The main difference with respect to the last two experiments is that the residual integration orders are now stationary with $d_{r,it} = 0.4$. Note that $d_{r,it} < \delta_{r,it}$ as well as $d_{r,it} < \varrho_t$.

4. Experiment $5\mathbb{D}$. The residual integration orders ($d_{r,it} = 0.2$) are now smaller whereas the persistence of the top-level factor is the greater with $\varrho = 0.6$.

5. Experiment $6\mathbb{A}$. The impact of non-cointegration in model (36) is analyzed assuming that $d_{r,it} = 1$ is greater than $\delta_{r,it} = 0.7$, $\varrho_t = 0.6$, and $\vartheta_{r,it} = 0.4$.

   The results are reported in Table 5.

6.2.1 Results

The results of Experiments 1-4 are reported in Tables 1-4, respectively. As can be seen from Table 1, the CCEMG and CCEP estimators of Pesaran (2006) are substantially biased, performing very poorly even when the cross section dimension increases considerably. In contrast, the CCEMG (and CCEP) estimators provided by (19) perform well with biases practically equal to zero in all cases independently of size distortions between N and T. The RMSE falls steadily as N and/or T increase. The proposed methodology performs well, in both homogeneous and heterogeneous slope cases. These conclusions also apply when the number of regressors are increased as seen from Table 2. The performance of β^x_{CCEMG} and β^z_{CCEMG} is practically identical. The biases of the CCEMG estimators are generally higher than those in Table 1 but with smaller RMSE.

Tables 3 and 4 report the results of the Monte Carlo simulation carried out as robustness checks. The CCEMG estimators perform well irrespective of whether rank conditions (13 and 18) are satisfied as can be seen from Table 3. Despite the number of block-specific factors or the number of top-level factors exceeds the number of regressors and regressand, the RMSEs of the CCEMG estimators

Table 5: Small-sample properties of the Experiment 5. ($d^* = 1$). Bias and RMSE of CCEMG and CCEP estimators are multiplied by 100 in the report.

			Ergemen and Velasco (2015)						MLCCE					
			CCEMG		CCEP		Res.Mem.Bias		CCEMG		CCEP		Res.Mem.Bias	
N	T	ϕ	Bias	RMSE	Bias	RMSE	CCEMG	CCEP	Bias	RMSE	Bias	RMSE	CCEMG	CCEP
Experiment 5$\mathbb{A}$: $\vartheta_{r,0} = 0.2$ $\varrho_0 = 0.4$ $\delta_{r,i0} = 1$ $d_{r,i0} = 0.75$ (Cointegration)														
20	150	0.5	30.25	30.25	30.60	30.60	0.070	0.070	0.19	3.88	0.32	4.87	0.018	0.019
20	150	2	14.96	14.96	15.33	15.33	0.020	0.020	0.04	1.99	0.01	2.21	0.007	0.007
20	150	5	7.60	7.61	7.42	7.49	-0.004	-0.003	0.16	1.29	0.16	1.34	0.005	0.005
100	500	0.5	30.32	30.32	30.81	30.81	0.081	0.081	0.01	1.57	0.01	2.09	0.021	0.021
100	500	2	15.12	15.12	15.67	15.67	0.031	0.032	-0.03	0.79	-0.05	0.94	0.006	0.006
100	500	5	8.02	8.02	8.02	8.03	0.010	0.010	0.03	0.48	0.04	0.52	0.002	0.002
200	500	0.5	30.70	30.70	31.23	31.23	0.083	0.083	-0.11	1.14	-0.14	1.47	0.020	0.020
200	500	2	15.19	15.19	15.73	15.73	0.032	0.033	0.02	0.55	0.05	0.67	0.006	0.006
200	500	5	7.94	7.94	7.97	7.97	0.010	0.010	0.01	0.29	0.00	0.32	0.002	0.002
200	2500	0.5	31.13	31.13	31.74	31.74	0.090	0.090	0.02	1.22	0.00	1.59	0.021	0.021
200	2500	2	15.27	15.27	15.92	15.92	0.039	0.040	0.04	0.56	0.04	0.66	0.006	0.006
200	2500	5	7.94	7.94	7.97	7.97	0.017	0.017	0.02	0.28	0.02	0.31	0.002	0.002
Experiment 5$\mathbb{B}$: $\vartheta_{r,0} = 0.4$ $\varrho_0 = 0.6$ $\delta_{r,i0} = 1$ $d_{r,i0} = 0.75$ (Cointegration)														
20	150	0.5	31.25	31.25	31.62	31.62	0.155	0.155	-0.52	3.79	-0.94	4.74	0.046	0.046
20	150	2	15.83	15.83	16.21	16.21	0.068	0.068	-0.01	1.92	0.00	2.19	0.023	0.023
20	150	5	8.49	8.50	8.49	8.50	0.027	0.027	-0.03	1.28	-0.03	1.33	0.013	0.013
100	500	0.5	31.09	31.09	31.45	31.45	0.174	0.174	0.06	1.69	0.05	2.18	0.045	0.045
100	500	2	16.17	16.17	16.71	16.71	0.088	0.088	-0.02	0.83	-0.04	0.98	0.023	0.023
100	500	5	8.63	8.63	8.66	8.66	0.043	0.043	0.00	0.46	0.00	0.50	0.012	0.012
200	500	0.5	31.08	31.08	31.32	31.32	0.174	0.174	0.11	1.21	0.12	1.60	0.044	0.044
200	500	2	15.97	15.97	16.51	16.51	0.088	0.088	0.02	0.55	0.00	0.65	0.022	0.022
200	500	5	8.61	8.61	8.62	8.62	0.043	0.043	0.00	0.33	0.00	0.36	0.011	0.011
200	2500	0.5	31.32	31.32	31.66	31.66	0.185	0.185	0.02	1.18	-0.07	1.56	0.045	0.045
200	2500	2	16.30	16.30	16.98	16.98	0.097	0.097	0.03	0.53	0.04	0.63	0.022	0.022
200	2500	5	8.61	8.61	8.68	8.68	0.052	0.052	-0.03	0.27	-0.03	0.30	0.012	0.012

(*continued on next page*).

Table 5 (*continued*)

			Ergemen and Velasco (2015)						MLCCE					
			CCEMG		CCEP		Res.Mem.Bias		CCEMG		CCEP		Res.Mem.Bias	
N	T	ϕ	Bias	RMSE	Bias	RMSE	CCEMG	CCEP	Bias	RMSE	Bias	RMSE	CCEMG	CCEP
Experiment $5\mathbb{C}$: $\vartheta_{r,0}=0.2$ $\varrho_0=0.6$ $\delta_{r,i0}=0.8$ $d_{r,i0}=0.4$ (Cointegration)														
20	150	0.5	34.71	34.71	35.18	35.18	0.356	0.356	-0.03	3.79	0.12	4.98	0.150	0.151
20	150	2	19.74	19.74	21.11	21.11	0.227	0.227	0.06	2.13	0.11	2.45	0.108	0.108
20	150	5	11.07	11.07	11.8	11.8	0.146	0.146	-0.01	1.40	-0.02	1.47	0.079	0.079
100	500	0.5	34.63	34.63	34.82	34.82	0.365	0.365	0.06	1.61	0.06	2.08	0.143	0.143
100	500	2	19.67	19.67	21.22	21.22	0.234	0.234	0.03	0.87	0.00	1.05	0.097	0.097
100	500	5	11.30	11.30	12.12	12.12	0.152	0.152	0.05	0.47	0.06	0.51	0.067	0.067
200	500	0.5	34.65	34.65	34.71	34.71	0.327	0.327	0.02	1.16	-0.02	1.55	0.141	0.141
200	500	2	19.67	19.67	21.18	21.18	0.232	0.232	0.01	0.55	-0.01	0.67	0.097	0.097
200	500	5	11.31	11.31	12.14	12.14	0.151	0.151	0.02	0.32	0.02	0.35	0.066	0.066
200	2500	0.5	34.63	34.63	34.82	34.82	0.368	0.368	0.03	1.24	0.04	1.70	0.147	0.147
200	2500	2	19.84	19.84	21.38	21.38	0.236	0.237	0.00	0.57	-0.01	0.67	0.096	0.096
200	2500	5	11.31	11.31	12.15	12.15	0.154	0.155	0.00	0.29	0.00	0.33	0.065	0.065
Experiment $5\mathbb{D}$: $\vartheta_{r,0}=0.4$ $\varrho_0=0.8$ $\delta_{r,i0}=0.6$ $d_{r,i0}=0.2$ (Cointegration)														
100	500	0.5	34.92	34.92	35.16	35.16	0.632	0.632	-0.02	1.83	-0.08	2.46	0.329	0.329
100	500	2	21.96	21.96	23.68	23.68	0.472	0.470	-0.06	1.11	-0.07	1.40	0.269	0.269
100	500	5	15.04	15.04	16.35	16.35	0.367	0.366	0.00	0.71	0.00	0.84	0.213	0.213
Experiment $6\mathbb{A}$: $\vartheta_{r,0}=0.4$ $\varrho_0=0.6$ $\delta_{r,i0}=0.7$ $d_{r,i0}=1$ (No cointegration)														
100	500	0.5	32.73	32.73	33.32	33.32	-0.013	-0.014	-0.02	1.72	-0.03	2.36	-0.022	-0.022
100	500	2	17.87	17.87	19.18	19.18	-0.053	-0.053	-0.08	0.79	-0.10	0.94	-0.010	-0.010
100	500	5	10.20	10.20	10.89	10.89	-0.067	-0.067	0.05	0.42	0.05	0.46	-0.004	-0.004
200	500	0.5	32.43	32.43	33.03	33.03	-0.042	-0.043	-0.10	1.57	-0.15	2.13	-0.029	-0.029
200	500	2	17.79	17.79	19.12	19.12	-0.064	-0.064	0.06	0.82	0.09	0.97	-0.015	-0.015
200	500	5	9.79	9.79	10.42	10.42	-0.072	-0.072	0.04	0.49	0.04	0.53	-0.009	-0.009

Notes: The DGP is $y_{r,it}=\beta'_{r,i}x_{r,it}+\mu'_{r,i}G_t+\lambda'_{r,i}F_{r,t}+\Delta_t^{-d_{r,i0}}\epsilon_{r,it}$. Regressors are generated by $x_{r,it}=\mu'_{r,i}G_t+\lambda'_{r,i}F_{r,t}+\Delta_t^{-\delta_{r,i0}}v_{r,it}$. $r=1,2$ and $i\in(20,100,200)$ and $t\in(150,500,2500)$. $\epsilon_{r,it}\overset{iid}{\sim}N(0,2\phi)$ and $v_{r,it}\overset{iid}{\sim}N(0,2\phi)$ are generated independently with ϕ controlling the signal-to-noise-ratio with $\phi=\{5,2,0.5\}$. Only one top-level factor and only one block-specific factor in each block are considered. $G_t=0.5G_{t-1}+\Delta_t^{-\varrho}e_t^G$ with $e_t^G\sim IIDN(0,1)$. $F_{r,t}=0.5F_{r,t-1}+\Delta_t^{-\vartheta_r}e_t^{F_r}$ with $e_t^{F_r}\sim IIDN(0,1)$. Only slope homogeneity is considered by $\beta_{r,i}=1$. CCEMG, CCEP and Residual memory estimates in the first three columns correspond to those proposed in Ergemen and Velasco (2015). Remaning estimators correspond to the MLCCE method proposed in this paper. CCEMG proposed in (29) and the residual memory estimated by (34). Rank conditions (13) and (18) are fulfilled. All experiments are based on 1000 replications.

decrease as N or T increases confirming consistency in rank-deficient case. These results are in line with those in Kapetanios et al. (2011) who discuss the same experiment considering I(1) processes in the common factors. For the sake of brevity I do not incorporate a rank-deficient case in the second set of simulations but the conclusions are the same.

In presence of rank-deficiency, the CCEMG estimators have slightly larger RMSEs than those in the experiments where rank conditions are satisfied (See Tables 1 and 2). A surprising finding is that the RMSEs with $\mathbf{r}_G = \mathbf{r}_{F_r} = 3$ are in general larger than those with $\mathbf{r}_G = \mathbf{r}_{F_r} = 4$ even when the number of regressors (k) is the same. Table 4 shows the performance of model 35 when the block-specific factors or/and the top-level factor are subject to mean shifts. It is clear that even if bias and RMSEs of ECMG estimators are larger than those of Table 1, they vanish as N and/or T increases. The findings are still consistent with those of Stock and Watson (2009a) even when I consider breaks in the multi-level factor structure. Finally, there is not a clear conclusion if the break in the mean of G_t impacts more in the estimation of ECMG than the break in the mean of $F_{r,t}$, however it seems that in cases when both levels of common factors present a break in the respective means, the impact over ECMG estimators are larger than those cases with individual breaks.

The second set of experiments focuses on the model with long-range dependence (36) and results are reported in Table 5. In the designs I consider that the top-level factor is more persistent than the block-specific factors. For the sake of brevity I do not study the case where $\vartheta_{r,i0} > \varrho_0$ but the results do not differ much. As can be seen from both tables, the CCEMG and CCEP estimators when applying the procedure of Ergemen and Velasco (2015) are substantially biased. Such biases are considerably reduced only when the noise dominates the signal (rows with $\phi = 5$) and do not change as N and/or T increase. The CCEMG estimators provided by (28) (and the respective CCEP) perform very well. I find the same effect in the RMSEs in the signal-to-noise ratio analysis. These conclusions are basically identical in Experiments $5\mathbb{A} - 5\mathbb{D}$. As seen from Table $6\mathbb{A}$, $\hat{\beta}_{r,i}$ can be estimated consistently even in the absence of cointegration which is in line with the findings in Ergemen and Velasco (2015) and using a similar approach as in Kapetanios et al. (2011). Note that Theorem 5.1 does not require cointegration for the consistency of $\hat{\beta}_{r,i}$.

The estimated residual integration orders by the CSS procedure following the steps proposed in this paper are slightly less biased than the counterparts of Ergemen and Velasco (2015) in Experiment $5\mathbb{A}$ and considerably less biased in Experiments $5\mathbb{B}$ and $5\mathbb{C}$. Furthermore, the higher the signal-to-noise ratio, the larger the biases of the estimated residual integration orders regardless of the experiment in question. These impacts are slightly more pronounced using Ergemen and Velasco (2015) than in the methodology proposed in this paper. Furthermore, the smaller the distance between $d_{r,i0}$ and ϱ_0, and between ϱ_0 and $\delta_{r,i0}$, the larger the biases in the estimated residual integration orders. This result is most pronounced in Experiment $5\mathbb{D}$ where $d_{r,i0}$ is much smaller than the other fractional

parameters. Finally, in case of no cointegration (Experiment $6\mathbb{A}$), the estimated residual integration orders perform well.

Evidence with respect to model (36) suggests that the factor structure should be much less persistent than the regressors and the regressand in order to improve the estimation of $\hat{d}_{r,it}$ via CSS, although such a condition does not impact the consistency of $\hat{\beta}_{r,i}$ via CCEMG or CCEP.

7 Concluding remarks

This paper provides a simple procedure for estimation of a couple of large panel data models, which are composed of several but fixed blocks of data. Both models are subject to cross-sectional error dependence that is driven by top-level and block-specific factors. In the first setup, I study the case where observables, factors, and regression errors are I(0) processes. In the second setup, I extend the study to include long-range dependence in the observables, factors, and disturbances without restrictions that can be either stationary or non-stationary processes.

I have proposed a simple methodology to completely filter out the cross-sectional dependence involved in the panel data models considered. This methodology is carried out in two steps. The first uses sample means of data separately in each block whereas the second uses the complete sample means of the new projected variables. In the fractional integrated setup, the methodology proposed follows the same steps but considers fractionally differenced data instead.

The focus of this paper is the estimation of $\beta_{r,i}$ for which I use the so-called Common Correlated Effect methods. Furthermore, only in the setup with long-range dependence, the memory estimation is also relevant and consists of conditional-sum-of-squares of defactored (fractionally differenced) variables. I have established asymptotic results of the estimation methods.

From an extensive Monte Carlo study, I illustrate that the methodology proposed works well in relatively small sample sizes even when considering some variations in the proposed specification.

The methodology proposed can be extended at least in the following directions. i) Fixed effects and deterministic trends can be easily incorporated in the fractional integrated model. The properties of estimators in such models have already been studied in the literature. ii) It might be interesting to analyze if further point of views regarding the multi-level factor structure can be useful in panel data models to control a different kind of cross-sectional dependence. iii) A possible extension of this paper is to study panel data sets whose cross-sectional dependence can be composed of more than two-levels. iv) A direct extension of the paper is to relax the assumption of independence between the idiosyncratic shocks in the models proposed.

Appendix

A Proof of Theorem (4.1)

First, note that both observable and unobservable block-specific factors are filtered out from (1 and 2) using the block-specific projection matrix $\overline{\mathbf{W}}_r$ as explained before in (17) as long as the rank condition (13) is satisfied. Second, by Frisch-Waugh-Lovell Theorem, $\hat{\beta}_{r,i}$ is consistently estimated from

$$\begin{aligned}\overline{\mathbf{W}}_{r,t}\, y_{r,it} &= \beta'_{r,i}\overline{\mathbf{W}}_{r,t}\, \mathbf{x}_{r,it} + \mu'_{r,i}\overline{\mathbf{W}}_{r,t}\, G_t + \overline{\mathbf{W}}_{r,t}\, \epsilon_{r,it},\\ \overline{\mathbf{W}}_{r,t}\, \mathbf{x}_{r,it} &= M'_{r,i}\overline{\mathbf{W}}_{r,t}\, G_t + \overline{\mathbf{W}}_{r,t}\, \mathbf{v}_{r,it}.\end{aligned}$$

To simplify notation, I write the last expressions as

$$y^*_{r,it} = \beta'_{r,i}\, \mathbf{x}^*_{r,it} + \mu'_{r,i}\, G^*_t + \epsilon^*_{r,it}, \tag{37}$$

$$\mathbf{x}^*_{r,it} = M'_{r,i}\, G^*_t + \mathbf{v}^*_{r,it}. \tag{38}$$

To clarify the exposition of the proof, recall that r and i denote the specific block and cross-section unit treated. $y^*_{r,it}$ and $\mathbf{x}^*_{r,it}$ are R $N \times 1$ and R $N \times k$ observable data matrices. I conduct the asymptotic analysis for a separate block r, but naturally it is valid for each $r = 1, \ldots, R$. Moreover, recall that $F_{r,t}$ denotes the block-specific factor of the region r whose loadings are null in the remaining blocks while the loading of G_t, the top-level factor, are non null regardless of the block treated. Then, taking in mind that r denotes the specific block treated, and for each i and $t = 1, 2, \ldots, T$, write (37) in matrix notation as

$$\mathbf{y}^*_{r,i} = \mathbf{X}^*_{r,i}\beta_{r,i} + \mathbf{G}^*\mu_{r,i} + \epsilon^*_{r,i}, \tag{39}$$

where $\epsilon^*_{r,i} = \left(\epsilon^*_{r,i1}, \epsilon^*_{r,i2}, \ldots, \epsilon^*_{r,iT}\right)'$ and $\mathbf{G}^* = (\mathbf{G}^*_1, \mathbf{G}^*_2, \ldots, \mathbf{G}^*_T)'$, with G^*_t defined in (14).

The main interest relies on $\hat{\beta}_{r,i}$ which was defined previously in (20), recalling

$$\hat{\beta}_{r,i} = \left(\mathbf{X}^{*\prime}_{r,i}\overline{\mathbf{W}}^*\mathbf{X}^*_{r,i}\right)^{-1}\mathbf{X}^{*\prime}_{r,i}\overline{\mathbf{W}}^* y^*_{r,i}.$$

Then, combining (20) and (39)

$$\begin{aligned}\hat{\beta}_{r,i} - \beta_{r,i} &= \left(\frac{\mathbf{X}^{*\prime}_{r,i}\overline{\mathbf{W}}^*\mathbf{X}^*_{r,i}}{T}\right)^{-1}\left(\frac{\mathbf{X}^{*\prime}_{r,i}\overline{\mathbf{W}}^*\mathbf{G}^*}{T}\right)\mu_{r,i}\\ &\quad + \left(\frac{\mathbf{X}^{*\prime}_{r,i}\overline{\mathbf{W}}^*\mathbf{X}^*_{r,i}}{T}\right)^{-1}\left(\frac{\mathbf{X}^{*\prime}_{r,i}\overline{\mathbf{W}}^*\epsilon^*_{r,i}}{T}\right),\end{aligned} \tag{40}$$

which implies that $\hat{\beta}_{r,i}$ depends on the top-level factor $\mathbf{G}^*$ as well as the cross-sectional averages computed in the first step of the procedure through $\overline{\mathbf{W}}^*$ defined in (8).

I first analyze the component $\frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}^{*}\mathbf{G}^{*}}{T}$ in (40). See that

$$\frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}^{*}\mathbf{G}^{*}}{T}=\frac{\mathbf{X}_{r,i}^{*'}\left[\mathbf{I}_{T}-\overline{H}^{*}\left(\overline{H}^{*'}\overline{H}^{*}\right)^{-}\overline{H}^{*'}\right]\mathbf{G}^{*}}{T}, \tag{41}$$

from which the analysis can now separate in four components, $\frac{\mathbf{X}_{r,i}^{*'}\mathbf{G}^{*}}{T}$, $\frac{\mathbf{X}_{r,i}^{*'}\overline{H}^{*}}{T}$, $\frac{\overline{H}^{*'}\overline{H}^{*}}{T}$, and $\frac{\overline{H}^{*'}\mathbf{G}^{*}}{T}$.

In the component $\frac{\mathbf{X}_{r,i}^{*'}\mathbf{G}^{*}}{T}$, first note that from (7 and 14) and due to $\overline{\mathbf{W}}_{r}$ is idempotent, so

$$\frac{\mathbf{X}_{r,i}^{*'}\mathbf{G}^{*}}{T}=\frac{\mathbf{X}_{r,i}^{'}\overline{\mathbf{W}}_{r}\mathbf{G}}{T}=\frac{\mathbf{X}_{r,i}^{'}\left[\mathbf{I}_{T}-\overline{H}_{r}\left(\overline{H}_{r}^{\prime}\overline{H}_{r}\right)^{-}\overline{H}_{r}^{\prime}\right]\mathbf{G}}{T}. \tag{42}$$

Recall now that $\overline{H}_{r}=\left(\mathbf{D}_{r},\overline{\mathbf{Z}}_{r}\right)$. I write (12) in a matrix notation and note that

$$\overline{H}_{r}=\mathbb{F}_{r}\bar{\mathbf{P}}_{r}+\bar{\mathbf{U}}_{r}^{+}, \tag{43}$$

where

$$\mathbb{F}_{r}=\left(\mathbf{D}_{r},\mathbf{F}_{r}\right),\quad \bar{\mathbf{P}}_{r}=\begin{pmatrix}\mathbb{I}_{N} & \bar{\mathbf{B}}_{r}\\ \mathbf{0} & \bar{\mathbf{D}}_{r}\end{pmatrix},\quad \bar{\mathbf{U}}^{+}=\left(\mathbf{0},\bar{\mathbf{U}}_{r}\right),$$

with $\mathbf{D}_{r}=\left(\mathbf{d}_{r,1},\ldots,\mathbf{d}_{r,T}\right)^{'}$, $\mathbf{F}_{r}=\left(\mathbf{F}_{r,1},\ldots,\mathbf{F}_{r,T}\right)^{'}$, and $\bar{\mathbf{U}}_{r}=\left(\bar{\mathbf{u}}_{r,1},\ldots,\bar{\mathbf{u}}_{r,T}\right)^{'}$.

Then[1],

$$\overline{H}_{r}^{\prime}\mathbf{G}=O_{p}\left(\frac{1}{\sqrt{NT}}\right)=o_{p}(1) \tag{44}$$

due to the top-level factor is orthogonal to the observable and unobservable block-specific factors and, reasoning as in Pesaran (2006), since

$$\frac{\mathbf{G}^{\prime}\bar{\mathbf{U}}_{r}}{T}=O_{p}\left(\frac{1}{\sqrt{NT}}\right), \tag{45}$$

from which one gets

$$\frac{\mathbf{X}_{r,i}^{*'}\mathbf{G}^{*}}{T}=\frac{\mathbf{X}_{r,i}^{'}\mathbf{G}}{T}=O_{p}(1). \tag{46}$$

To examine the component $\frac{\mathbf{X}_{r,i}^{*'}\overline{H}^{*}}{T}$, note that from (15) in matrix notation,

$$\mathbf{X}_{r,i}^{*}=\mathbf{G}^{*}\mathbf{M}_{r,i}+\mathbf{V}_{r,i}^{*}, \tag{47}$$

$$\overline{H}^{*}=\mathbf{G}^{*}\bar{\mathbf{C}}+\bar{\mathbf{U}}^{*}, \tag{48}$$

[1]Proofs of convergence rates presented here are proved in Appendix A in Pesaran (2006).

where $\bar{\mathbf{U}}^* = (\bar{\mathbf{u}}_1^*, \ldots, \bar{\mathbf{u}}_T^*)$ and $\mathbf{V}_{r,i}^* = (\mathbf{v}_{r,i1}, \ldots, \mathbf{v}_{r,it})$. Then

$$\begin{aligned} \mathbf{X}_{r,i}^{*'}\overline{H}^* &= \mathbf{X}_{r,i}^{*'}\mathbf{G}^*\bar{\mathbf{C}} + \mathbf{X}_{r,i}^{*'}\bar{\mathbf{U}}_{(r)}^* \\ &= \mathbf{X}_{r,i}^{'}\overline{\mathbf{W}}_r\mathbf{G}\bar{\mathbf{C}} + \mathbf{X}_{r,i}^{'}\overline{\mathbf{W}}_r\bar{\mathbf{U}}_r, \end{aligned} \tag{49}$$

with $\bar{\mathbf{U}}_{(r)}^*$ indicating $\bar{\mathbf{U}}^*$ in the specific block r..

I now analyze both terms of the right-hand side of (49). First, from (46), I have that

$$\frac{\mathbf{X}_{r,i}^{*'}\mathbf{G}^*\bar{\mathbf{C}}}{T} = \frac{\mathbf{X}_{r,i}^{'}\mathbf{G}}{T}\bar{\mathbf{C}} = O_p(1).$$

Second,

$$\frac{\mathbf{X}_{r,i}^{'}\overline{\mathbf{W}}_r\bar{\mathbf{U}}_r}{T} = \frac{\mathbf{X}_{r,i}^{'}\left[\mathbf{I}_T - \overline{H}_r\left(\overline{H}_r'\overline{H}_r\right)^{-}\overline{H}_r'\right]\bar{\mathbf{U}}_r}{T},$$

from which, reasoning as in Pesaran (2006),

$$\frac{\mathbf{X}_{r,i}^{'}\bar{\mathbf{U}}_r}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{50}$$

$$\frac{\mathbf{X}_{r,i}^{'}\bar{H}_r}{T} = \left(\frac{\mathbf{X}_{r,i}^{'}\mathbb{F}_r}{T}\right)\bar{\mathbf{P}}_r + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{51}$$

$$\frac{\bar{H}_r'\bar{H}_r}{T} = \bar{\mathbf{P}}_r'\left(\frac{\mathbb{F}_r'\mathbb{F}_r}{T}\right)\bar{\mathbf{P}}_r + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{52}$$

$$\frac{\bar{H}_r'\bar{\mathbf{U}}_r}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right). \tag{53}$$

(53) is obtained immediately considering (43) and Assumptions (A and B_1) regarding that $F_{r,t}, d_{r,t}$ and $\bar{\mathbf{u}}_t$ are independently distributed covariance stationary processes. $\bar{\mathbf{U}}_{(r)}$ indicates $\bar{\mathbf{U}}$ in the specific block r. Then,

$$\frac{\mathbf{X}_{r,i}^{'}\overline{\mathbf{W}}_r\bar{\mathbf{U}}_r}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Hence, it is obtained that

$$\frac{\mathbf{X}_{r,i}^{*'}\overline{H}^*}{T} = \frac{\mathbf{X}_{r,i}^{'}\mathbf{G}}{T}\bar{\mathbf{C}} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{54}$$

Now I focus on the component $\frac{\overline{H}^{*'}\overline{H}^*}{T}$. Using (48), it can be separated in $\frac{\mathbf{G}^{*'}\mathbf{G}^*}{T}$, $\frac{\mathbf{G}^{*'}\bar{\mathbf{U}}^*}{T}$, and $\frac{\bar{\mathbf{U}}^{*'}\bar{\mathbf{U}}^*}{T}$. Re-expressing these terms as before and using (44),

(52), and (53), it is easily seen that

$$\frac{\mathbf{G}^{*'}\mathbf{G}^{*}}{T} = \frac{\mathbf{G}'\overline{\mathbf{W}}_r\mathbf{G}}{T} = \frac{\mathbf{G}'\mathbf{G}}{T} + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{55}$$

$$\frac{\mathbf{G}^{*'}\bar{\mathbf{U}}^{*}}{T} = \frac{\mathbf{G}'\overline{\mathbf{W}}_r\bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{56}$$

$$\frac{\bar{\mathbf{U}}^{*'}\bar{\mathbf{U}}^{*}}{T} = \frac{\bar{\mathbf{U}}'\overline{\mathbf{W}}_r\bar{\mathbf{U}}}{T} = O_p\left(\frac{1}{N^2}\right) + O_p\left(\frac{1}{N^{\frac{3}{2}}T^{\frac{1}{2}}}\right) + O_p\left(\frac{1}{NT}\right), \tag{57}$$

from which,

$$\frac{\overline{H}^{*'}\overline{H}^{*}}{T} = \bar{\mathbf{C}}'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right)\bar{\mathbf{C}} + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N^2}\right). \tag{58}$$

And the last component, $\frac{\overline{H}^{*'}\mathbf{G}^{*}}{T}$, is directly obtained using (55) and (56),

$$\frac{\overline{H}^{*'}\mathbf{G}^{*}}{T} = \bar{\mathbf{C}}'\left(\frac{\mathbf{G}'\mathbf{G}}{T}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{59}$$

Finally, using (46), (54), (58), (59), from (41) and following Pesaran (2006), it is easy to see that

$$\frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}^{*}\mathbf{G}^{*}}{T} = \frac{\mathbf{X}_{r,i}'\overline{\mathbf{W}}_G\,\mathbf{G}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),$$

and, since $\overline{\mathbf{W}}_G\,\mathbf{G} = \mathbf{0}$,

$$\frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}^{*}\mathbf{G}^{*}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{60}$$

Furthermore, the same steps would lead to get

$$\frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}^{*}\mathbf{X}_{r,i}^{*}}{T} = \frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}_G\,\mathbf{X}_{r,i}^{*}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{61}$$

$$\frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}^{*}\epsilon_{r,i}^{*}}{T} = \frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}_G\,\epsilon_{r,i}^{*}}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{62}$$

Using the results (60-62) in 40, then

$$\hat{\beta}_{r,i} - \beta_{r,i} = \left(\frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}_G\,\mathbf{X}_{r,i}^{*}}{T}\right)^{-1}\left(\frac{\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}_G\,\epsilon_{r,i}^{*}}{T}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{63}$$

where $T^{-1}\left(\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}_G\,\mathbf{X}_{r,i}^{*}\right) = O_p(1)$ and converge in probability to a positive definite matrix, $\Sigma_{\beta_{r,i}}$, under Assumption (F_1). Since $\epsilon_{r,i}^{*}$ is independently distributed of $\mathbf{X}_{r,i}^{*}$, $\mathbf{G}$ and $\mathbb{F}_r$, then $T^{-1}\left(\mathbf{X}_{r,i}^{*'}\overline{\mathbf{W}}_G\,\epsilon_{r,i}^{*}\right) \xrightarrow{p} 0$. From which, the desired property

$$\hat{\beta}_{r,i} - \beta_{r,i} = o_p(1)$$

is obtained.

Furthermore, multiplying (63) by $\sqrt{T}$ and assuming that $\frac{\sqrt{T}}{N} \to 0$, it is obtained that

$$\sqrt{T}\left(\hat{\beta}_{r,i} - \beta_{r,i}\right) \xrightarrow{d} N\left(\mathbf{0}, \Sigma_{\beta_{r,i}}\right),$$

□

B Proof of Theorem (4.2)

Reasoning as in Pesaran (2006), under Assumption E and using (40), it can be easily obtained that

$$\sqrt{R\,N}\left(\hat{\beta}_{\text{CCEMG}} - \beta\right) = \frac{1}{\sqrt{R}}\frac{1}{\sqrt{N}}\sum_{r=1}^{R}\sum_{i=1}^{N}\nu_{r,i} + o_p(1),$$

then, as $(N,T) \to \infty$ with R fixed ($R < N$)

$$\sqrt{R\,N}\left(\hat{\beta}_{\text{CCEMG}} - \beta\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Sigma}_{CCEMG}\right),$$

□

C Proof of Theorem (5.1)

Considering Assumptions (G_1-G_4), take $d^*_{r,i} \geq 1$ in order to establish stationarity in all the components involved in (23). Then, following the same strategy as in the proof of Theorem (4.1), under rank condition (18), the unobservable block-specific factors are filtered from (23) by using the projection matrix $\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)$. From which, I get

$$\begin{aligned}\hat{\beta}_{r,i}\left(d^*_{r,i}\right) - \beta_{r,i} \;=\; & \underbrace{\left(\frac{\mathcal{X}^{*'}_{r,i}\,\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)\mathcal{X}^*_{r,i}}{T}\right)^{-1}}_{\mathcal{H}_1}\underbrace{\left(\frac{\mathcal{X}^{*'}_{r,i}\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)\,G^*}{T}\right)\mu_{r,i}}_{\mathcal{H}_2} \\ & + \left(\frac{\mathcal{X}^{*'}_{r,i}\,\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)\mathcal{X}^*_{r,i}}{T}\right)^{-1}\underbrace{\left(\frac{\mathcal{X}^{*'}_{r,i}\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)\Delta_t^{d^*_{r,i}}\epsilon^*_{r,i}}{T}\right)}_{\mathcal{H}_3}. \quad (64)\end{aligned}$$

First, to examine the component $\mathcal{H}_1$, first see that from 23,

$$\mathcal{X}^*_{r,i} = M'_{r,i}\,\Delta_t^{d^*_{r,i}}\,G^*_t + \Delta_t^{d^*_{r,i}-\delta_{r,i0}}\mathbf{v}^*_{r,it},$$

after filtering the block-specific factors. Hence, it can be addressed a similar reasoning as in (61) to get

$$\mathcal{X}^*_{r,i}\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)\mathcal{X}^{*'}_{r,i} = \mathcal{X}^*_{r,i}\overline{\mathbf{W}}_G\left(d^*_{r,i}\right)\mathcal{X}^{*'}_{r,i} + o_p(1),$$

from which $T^{-1}\mathcal{X}^*_{r,i}\overline{\mathbf{W}}_G\left(d^*_{r,i}\right)\mathcal{X}^{*'}_{r,i} \xrightarrow{p} \Sigma_{\beta_{r,i}}$. Note that $T^{-1}\left(\Delta_t^{d^*_{r,i}-\delta_{r,i0}}\mathbf{v}_{r,it}\ \Delta_t^{d^*_{r,i}-\delta_{r,i0}}\mathbf{v}'_{r,it}\right) \xrightarrow{p} \Sigma_{r,i\mathbf{v}}$.

Second, the proof that $\mathcal{H}_2 = o_p(1)$ is similar to that proof of Theorem 4 in Ergemen and Velasco (2015) following the same strategy it is used to get (60). In short,

$$\frac{\mathcal{X}^{*'}_{r,i}\overline{\mathbf{W}}^*\left(d^*_{r,i}\right)G^*}{T} = \frac{\mathcal{X}^{*'}_{r,i}\left[\mathbf{I}_T - \overline{\mathcal{H}}^*\left(\overline{\mathcal{H}}^{*'}\overline{\mathcal{H}}^*\right)^{-}\overline{\mathcal{H}}^{*'}\right]G^*}{T}, \tag{65}$$

then, it can be separated into four components which are examined as before.

For the component, $\mathcal{X}^{*'}_{r,i}G^*/T$, it can be easily obtained that

$$\frac{\mathcal{X}^{*'}_{r,i}G^*}{T} = \frac{\mathcal{X}'_{r,i}G}{T} + o_p(1),$$

following the steps of a fractionally differenced version of 42. Note that $G\left(d^*_{r,i}\right)$ is orthogonal to $F\left(d^*_{r,i}\right)$ and, adjusting Ergemen and Velasco (2015), it is obtained that $G'\bar{U}_r = O_p\left(\frac{1}{\sqrt{NT}}\right)$ as in (45).

To prove that the approximation of the remaining components in (65),

$$\mathcal{X}^{*'}_{r,i}\overline{\mathcal{H}}^*/T, \qquad \overline{\mathcal{H}}^{*'}\overline{\mathcal{H}}^*/T, \qquad \overline{\mathcal{H}}^*G^*/T,$$

are negligible, it is necessary to apply the same strategy discussed to get (54), (58), and (59), respectively, but using appropriated convergence rates obtained in Ergemen and Velasco (2015). Therefore, since $\overline{\mathbf{W}}^*_G\left(d^*_{r,i}\right)G^* = 0$,

$$\mathcal{H}_2 = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) = o_p(1).$$

Finally, since $\overline{\mathbf{W}}^*_G\mathcal{X}^*_{r,i} = \Delta_t^{d^*_{r,i}-\delta_{r,i0}}\mathbf{v}^*_{r,it}$ and due to the independence of $\mathbf{v}^*_{r,it}$ and $\epsilon^*_{r,i}$, then $\mathcal{H}_3 \xrightarrow{p} 0$. Consequently,

$$\hat{\beta}_{r,i}\left(d^*_{r,i}\right) - \beta_{r,i} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{66}$$

and the proof of consistency in (30) is now complete.

As discussed, the asymptotic distribution (31) is directly obtained assuming $\frac{\sqrt{T}}{N} \to 0$ as $(N,T)_j \to \infty$ while the asymptotic distribution (32) needs the random coefficient model of Assumption E. □

D Proof of Theorem (5.2)

I closely follow Ergemen and Velasco (2015) to prove the consistency of the fractional integration parameter estimate. Note that observable variables can be written

as

$$\begin{aligned}
\tilde{y}_{r,i}(d^*_{r,i}) &= \Delta_t^{d^*_{r,i}} y^*_{r,i} \overline{\mathbf{W}}^* \left(d^*_{r,i}\right) = \Delta_t^{d^*_{r,i}} y_{r,i} \overline{\mathbf{W}}_r \left(d^*_{r,i}\right) \overline{\mathbf{W}}^* \left(d^*_{r,i}\right) \\
&= \Delta_t^{d^*_{r,i}} y_{r,i} \overline{\mathbf{W}}_r \left(d^*_{r,i}\right) \overline{\mathbf{W}}^* \left(d^*_{r,i}\right), \\
&= \Delta_t^{d^*_{r,i}} y_{r,i} \left[\mathbf{I}_T - \overline{\mathcal{H}}_r \left(\overline{\mathcal{H}}_r^{'} \overline{\mathcal{H}}_r\right)^{-} \overline{\mathcal{H}}_r^{'}\right] \left[\mathbf{I}_T - \overline{\mathcal{H}}^* \left(\overline{\mathcal{H}}^{*'} \overline{\mathcal{H}}^*\right)^{-} \overline{\mathcal{H}}^{*'}\right], \\
&= \Delta_t^{d^*_{r,i}} y_{r,i} - \Delta_t^{d^*_{r,i}} y_{r,i} \left[\overline{\mathcal{H}}_r \left(\overline{\mathcal{H}}_r^{'} \overline{\mathcal{H}}_r\right)^{-} \overline{\mathcal{H}}_r^{'} + \overline{\mathcal{H}}^* \left(\overline{\mathcal{H}}^{*'} \overline{\mathcal{H}}^*\right)^{-} \overline{\mathcal{H}}^{*'}\right], \\
&\quad + \Delta_t^{d^*_{r,i}} y_{r,i} \left[\overline{\mathcal{H}}_r \left(\overline{\mathcal{H}}_r^{'} \overline{\mathcal{H}}_r\right)^{-} \overline{\mathcal{H}}_r^{'}\right] \left[\overline{\mathcal{H}}^* \left(\overline{\mathcal{H}}^{*'} \overline{\mathcal{H}}^*\right)^{-} \overline{\mathcal{H}}^{*'}\right].
\end{aligned}$$

Now, when prefiltering both variables by $\Delta_t^{d_{r,i}-d^*_{r,i}}$, it is obtained that

$$\begin{aligned}
\Delta_t^{d_{r,i}-d^*_{r,i}} \tilde{y}_{r,i}(d^*_{r,i}) &= \Delta_t^{d_{r,i}} y_{r,i} - \Delta_t^{d^*_{r,i}} y_{r,i} \left[\overline{\mathcal{H}}_r^{+} \left(\overline{\mathcal{H}}_r^{'} \overline{\mathcal{H}}_r\right)^{-} \overline{\mathcal{H}}_r^{'} + \overline{\mathcal{H}}^{+*} \left(\overline{\mathcal{H}}^{*'} \overline{\mathcal{H}}^*\right)^{-} \overline{\mathcal{H}}^{*'}\right], \\
&\quad \Delta_t^{d^*_{r,i}} y_{r,i} \left[\overline{\mathcal{H}}_r^{+} \left(\overline{\mathcal{H}}_r^{'} \overline{\mathcal{H}}_r\right)^{-} \overline{\mathcal{H}}_r^{'}\right] \left[\overline{\mathcal{H}}^{+*} \left(\overline{\mathcal{H}}^{*'} \overline{\mathcal{H}}^*\right)^{-} \overline{\mathcal{H}}^{*'}\right],
\end{aligned}$$

$\overline{\mathcal{H}}_r^{+} = \Delta_t^{d_{r,i}-d^*_{r,i}} \overline{\mathcal{H}}_r$ and $\overline{\mathcal{H}}^{*+} = \Delta_t^{d_{r,i}-d^*_{r,i}} \overline{\mathcal{H}}^*$.

Let $\mathcal{W}_{F_r} = \mathcal{F}_r \left(\mathcal{F}_r^{'} \mathcal{F}_r\right)^{-} \mathcal{F}_r^{+'}$ and $\mathcal{W}_G = \mathcal{G} \left(\mathcal{G}^{'} \mathcal{G}\right)^{-} \mathcal{G}^{+'}$, where $\mathcal{F}_r = \Delta_t^{d^*_{r,i}} F_r$, $\mathcal{G} = \Delta_t^{d^*_{r,i}} G$, $\mathcal{F}_r^{+} = \Delta_t^{d_{r,i}-d^*_{r,i}} \mathcal{F}_r$, and $\mathcal{G}^{+} = \Delta_t^{d_{r,i}-d^*_{r,i}} \mathcal{G}$. Moreover, denote $\overline{\mathbb{H}}^* = \overline{\mathcal{H}}^* \left(\overline{\mathcal{H}}^{*'} \overline{\mathcal{H}}^*\right)^{-} \overline{\mathcal{H}}^{*+'}$ and $\overline{\mathbb{H}}_r = \overline{\mathcal{H}}_r \left(\overline{\mathcal{H}}_r^{'} \overline{\mathcal{H}}_r\right)^{-} \overline{\mathcal{H}}_r^{+'}$. Then,

$$\begin{aligned}
\Delta_t^{d_{r,i}-d^*_{r,i}} \tilde{y}_{r,i}(d^*_{r,i}) &= \Delta_t^{d_{r,i}} y_{r,i} - \Delta_t^{d^*_{r,i}} y_{r,i} \left[\mathcal{W}_{F_r} + \mathcal{W}_G\right] + \Delta_t^{d^*_{r,i}} y_{r,i} \left[\mathcal{W}_{F_r} + \mathcal{W}_G \right. \\
&\quad \left. - \left[\overline{\mathbb{H}}^* + \overline{\mathbb{H}}_r\right]\right] + \Delta_t^{d^*_{r,i}} y_{r,i} \left[\overline{\mathbb{H}}^{*'} \overline{\mathbb{H}}_r\right].
\end{aligned}$$

The same for $\tilde{x}_{r,i}(d^*_{r,i})$ replacing $y_{r,i}$ by $x_{r,i}$ everywhere. Now, with the corresponding terms,

$$\begin{aligned}
\Delta_t^{d_{r,i}-d^*_{r,i}} \tilde{y}_{r,i}(d^*_{r,i}) &= \Delta_t^{d_{r,i}-d_{r,i0}} \epsilon_{r,i} + \beta_{r,i0}^{'} \Delta_t^{d_{r,i}-\delta_{r,i}} \mathbf{v}_{r,i} \\
&\quad - \left[\Delta_t^{d^*_{r,i}-d_{r,i0}} \epsilon_{r,i} + \beta_{r,i0}^{'} \Delta_t^{d^*_{r,i}-\delta_{r,i}} \mathbf{v}_{r,i}\right] \left[\mathcal{W}_{F_r} + \mathcal{W}_G\right] \\
&\quad + \left[\Delta_t^{d^*_{r,i}-d_{r,i0}} \epsilon_{r,i} + \beta_{r,i0}^{'} \Delta_t^{d^*_{r,i}-\delta_{r,i}} \mathbf{v}_{r,i} + \left(\beta_{r,i0}^{'} M_{r,i}^{'} + \mu_{r,i}^{'}\right) \mathcal{G}\right. \\
&\quad \left. + \left(\beta_{r,i0}^{'} \Lambda_{r,i}^{'} + \lambda_{r,i}^{'}\right) \mathcal{F}_r\right] \left[\mathcal{W}_{F_r} + \mathcal{W}_G - \left[\overline{\mathbb{H}}^* + \overline{\mathbb{H}}_r\right]\right] \\
&\quad + \left[\Delta_t^{d^*_{r,i}-d_{r,i0}} \epsilon_{r,i} + \beta_{r,i0}^{'} \Delta_t^{d^*_{r,i}-\delta_{r,i}} \mathbf{v}_{r,i}\right] \left[\overline{\mathbb{H}}^{*'} \overline{\mathbb{H}}_r\right],
\end{aligned}$$

and

$$\begin{aligned}\Delta_t^{d_{r,i}-d^*_{r,i}}\tilde{x}_{r,i}(d^*_{r,i}) &= \Delta_t^{d_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i} - \left[\Delta_t^{d^*_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i}\right][\mathcal{W}_{F_r}+\mathcal{W}_G] \\ &+ \left[\Delta_t^{d^*_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i} + \left(M'_{r,i}\mathcal{G}+\Lambda'_{r,i}\mathcal{F}_r\right)\right][\mathcal{W}_{F_r}+\mathcal{W}_G \\ &- \left[\overline{\mathbb{H}}^* + \overline{\mathbb{H}}_r\right]] + \left[\Delta_t^{d^*_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i}\right]\left[\overline{\mathbb{H}}^{*'}\overline{\mathbb{H}}_r\right].\end{aligned}$$

Hence, when appropriately accommodating the terms presented above, it is possible to write the residuals in the CSS (34) as

$$\Delta_t^{d_{r,i}-d^*_{r,i}}\left[\tilde{y}_{r,it}(d^*_{r,i}) - \hat{\beta}'_{r,i}(d^*_{r,i})\tilde{\mathbf{X}}_{r,i}(d^*_{r,i})\right] = \epsilon_i^1(d^*_{r,i})+\epsilon_i^2(d^*_{r,i})+\epsilon_i^3(d^*_{r,i})+\epsilon_i^4(d^*_{r,i}),$$

where

$$\begin{aligned}\epsilon_i^1(d^*_{r,i}) &= \Delta_t^{d_{r,i}-d_{r,i0}}\epsilon_{r,i} - \Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon_{r,i}\left[\mathcal{W}_{F_r}+\mathcal{W}_G\right], \\ \epsilon_i^2(d^*_{r,i}) &= -\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^*_{r,i})\right)'\left[\Delta_t^{d_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i} + \Delta_t^{d^*_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i}\left(\mathcal{W}_{F_r}+\mathcal{W}_G\right)\right], \\ \epsilon_i^3(d^*_{r,i}) &= \left\{\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^*_{r,i})\right)'\Delta_t^{d^*_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i} + \left[\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^*_{r,i})\right)'M'_{r,i}+\mu'_{r,i}\right]\mathcal{G}\right. \\ &+ \left.\left[\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^*_{r,i})\right)'\Lambda'_{r,i}+\lambda'_{r,i}\right]\mathcal{F}_r + \Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon_{r,i}\right\} \\ &\times\left\{\mathcal{W}_{F_r}+\mathcal{W}_G-\overline{\mathbb{H}}^*-\overline{\mathbb{H}}_r\right\}, \\ \epsilon_i^4(d^*_{r,i}) &= \overline{\mathbb{H}}^{*'}\overline{\mathbb{H}}_r\left[\Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon_{r,i} + \left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^*_{r,i})\right)'\mathbf{v}_{r,i}\right].\end{aligned}$$

Now it is necessary to study the contribution of each product $\epsilon_i^j(d^*_{r,i})\epsilon_i^k(d^*_{r,i})'$, $j,k=1,2,3,4$ to fully examine the properties of the CSS (34).

$$\begin{aligned}\frac{1}{T}\epsilon_i^1(d^*_{r,i})\epsilon_i^1(d^*_{r,i})' &= \underbrace{\frac{1}{T}\left[\Delta_t^{d_{r,i}-d_{r,i0}}\epsilon_{r,i}\Delta_t^{d_{r,i}-d_{r,i0}}\epsilon'_{r,i}\right]}_{\mathcal{P}1} \\ &\underbrace{+\frac{1}{T}\left[\left(\Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon_{r,i}\left[\mathcal{W}_{F_r}+\mathcal{W}_G\right]\right)\left(\Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon_{r,i}\left[\mathcal{W}_{F_r}+\mathcal{W}_G\right]\right)'\right]}_{\mathcal{P}2} \\ &\underbrace{-\frac{2}{T}\left[\Delta_t^{d_{r,i}-d_{r,i0}}\epsilon_{r,i}\left(\Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon_{r,i}\left[\mathcal{W}_{F_r}+\mathcal{W}_G\right]\right)'\right]}_{\mathcal{P}3}\end{aligned}$$

Theorem 1 in Ergemen and Velasco (2015) proves that the $\mathcal{P}1$ converges uniformly in $\mathcal{D}$ and is minimized for $d_{r,i}=d_{r,i0}$. Similar treatment for a general case can be found in the Theorem 2.1 in Hualde and Robinson (2011).

Since the fractionally differenced top-level and block-specific factors are orthogonal to each other, $\mathcal{P}2$ can be re-written as

$$\frac{1}{T}\left[\Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon_{r,i}\mathcal{W}_{F_r}\mathcal{W}'_{F_r}\Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon'_{r,i}\right]+\frac{1}{T}\left[\Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon_{r,i}\mathcal{W}_{G}\mathcal{W}'_{G}\Delta_t^{d^*_{r,i}-d_{r,i0}}\epsilon'_{r,i}\right].$$

Then, inasmuch as $\max(\varrho,\vartheta_{max})-\underline{d_{r,i}}<1/2$, one has that

$$\begin{aligned}
&\frac{\mathcal{F}_r\mathcal{F}'_r}{T}\xrightarrow{p}\Sigma_{\mathcal{F}_r},\quad \frac{\mathcal{G}\mathcal{G}'}{T}\xrightarrow{p}\Sigma_{\mathcal{G}},\\
&\frac{\mathcal{F}_r^{+}\mathcal{F}_r^{+\prime}}{T}=O_p\left(1+T^{2(\vartheta_r-\underline{d_{r,i}})-1}\right)=O_p(1),\\
&\frac{\mathcal{G}^{+}\mathcal{G}^{+\prime}}{T}=O_p\left(1+T^{2(\varrho-\underline{d_{r,i}})-1}\right)=O_p(1).
\end{aligned}$$

Finally, since

$$\begin{aligned}
\frac{\Delta_t^{d_{r,i0}-d^*_{r,i}}\epsilon_{r,i}\mathcal{F}'_r}{T} &= O_p\left(T^{-1/2}+T^{d_{r,i0}+\vartheta_r-2d^*_{r,i}-1}\right)=o_p(1),\\
\frac{\Delta_t^{d_{r,i0}-d^*_{r,i}}\epsilon_{r,i}\mathcal{G}'}{T} &= O_p\left(T^{-1/2}+T^{d_{r,i0}+\varrho-2d^*_{r,i}-1}\right)=o_p(1),
\end{aligned}$$

from which $\mathcal{P}2=o_p(1)$. The same applies with $\mathcal{P}3$.

Moreover, under the same assumption of orthogonality between both levels of factors and following the last steps in Theorem 7 in Ergemen and Velasco (2015), one has that

$$\frac{1}{T}\epsilon_i^2(d^*_{r,i})\epsilon_i^2(d^*_{r,i})'=o_p(1),\qquad \frac{1}{T}\epsilon_i^3(d^*_{r,i})\epsilon_i^3(d^*_{r,i})'=o_p(1),$$

as well as the cross-terms $\epsilon_i^j(d^*_{r,i})\epsilon_i^k(d^*_{r,i})'$, $j,k=1,2,3,4$ with $j\neq k$ are $o_p(1)$. Note that $\frac{1}{T}\epsilon_i^4(d^*_{r,i})\epsilon_i^4(d^*_{r,i})'=o_p(1)$ due to the properties of $\overline{\mathbb{H}}^{*'}$ and $\overline{\mathbb{H}}_r$, which are the same as in (44) and (53).

To proof the asymptotic normality of $\hat{d}_{r,i}$, the $\sqrt{T}$-normalized score evaluated at the true value, $d_{r,i0}$ is examined. It is given by

$$\sqrt{T}\left.\frac{\partial L^{*}_{r,i,T}(d_{r,i})}{\partial(d_{r,i})}\right|_{d_{r,i}=d_{r,i0}} = \frac{2}{\sqrt{T}}\Big\{\epsilon_{r,i} - \Delta_t^{d^{*}_{r,i}-d_{r,i0}}\epsilon_{r,i}\left[\mathcal{W}_{F_r}+\mathcal{W}_G\right]$$
$$-\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'\left[\Delta_t^{d_{r,i0}-\delta_{r,i}}\mathbf{v}_{r,i}+\Delta_t^{d^{*}_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i}\left(\mathcal{W}_{F_r}+\mathcal{W}_G\right)\right]$$
$$+\left[\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'\Delta_t^{d^{*}_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i}+\left[\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'M'_{r,i}+\mu'_{r,i}\right]\mathcal{G}\right.$$
$$\left.+\left[\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'\Lambda'_{r,i}+\lambda'_{r,i}\right]\mathcal{F}_r+\Delta_t^{d^{*}_{r,i}-d_{r,i0}}\epsilon_{r,i}\right]\left[\mathcal{W}_{F_r}+\mathcal{W}_G-\overline{\mathbb{H}}^{*}\right.$$
$$\left.-\overline{\mathbb{H}}_r\right]+\overline{\mathbb{H}}^{*'}\overline{\mathbb{H}}_r\left[\Delta_t^{d^{*}_{r,i}-d_{r,i0}}\epsilon_{r,i}+\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'\mathbf{v}_{r,i}\right]\Big\}\Big\{\left(\log\Delta_t\right)\epsilon_{r,i}$$
$$-\Delta_t^{d^{*}_{r,i}-d_{r,i0}}\epsilon_{r,i}\left[\dot{\mathcal{W}}_{F_r}+\dot{\mathcal{W}}_G\right]-\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'\left[\Delta_t^{d_{r,i0}-\delta_{r,i}}\mathbf{v}_{r,i}\right.$$
$$\left.+\Delta_t^{d^{*}_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i}\left(\dot{\mathcal{W}}_{F_r}+\dot{\mathcal{W}}_G\right)\right]+\left[\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'\Delta_t^{d^{*}_{r,i}-\delta_{r,i}}\mathbf{v}_{r,i}\right.$$
$$+\left[\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'M'_{r,i}+\mu'_{r,i}\right]\mathcal{G}+\left[\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'\Lambda'_{r,i}+\lambda'_{r,i}\right]\mathcal{F}_r$$
$$\left.+\Delta_t^{d^{*}_{r,i}-d_{r,i0}}\epsilon_{r,i}\right]\left[\dot{\mathcal{W}}_{F_r}+\dot{\mathcal{W}}_G-\dot{\overline{\mathbb{H}}}^{*}-\dot{\overline{\mathbb{H}}}_r\right]+\dot{\overline{\mathbb{H}}}^{*'}\dot{\overline{\mathbb{H}}}_r\left[\Delta_t^{d^{*}_{r,i}-d_{r,i0}}\epsilon_{r,i}\right.$$
$$\left.+\left(\beta_{r,i0}-\hat{\beta}_{r,i}(d^{*}_{r,i})\right)'\mathbf{v}_{r,i}\right]\Big\},$$

where $\dot{\mathcal{W}}_{F_r}=\mathcal{F}_r\left(\mathcal{F}'_r\mathcal{F}_r\right)^{-}\dot{\mathcal{F}}_r^{+\prime}$ with $\dot{\mathcal{F}}_r^{+\prime}=\left(\partial/\partial(d_{r,i})\right)\mathcal{F}_r^{+}$.
$\dot{\mathcal{W}}_G$, $\dot{\overline{\mathbb{H}}}$, and $\dot{\overline{\mathbb{H}}}_r$ are defined similarly.

Taking, R=1 and N=1, as $T\to\infty$, when applying the central limit theorem for martingale difference sequences, see Robinson and Velasco (2015), the term

$$\frac{2}{\sqrt{T}}\epsilon_{r,i}\left(\log\Delta_t\right)\epsilon'_{r,i}\xrightarrow{d}N(0,4\sigma_\epsilon).$$

Using the same assumption of orthogonality between both levels of factors and following Ergemen and Velasco (2015), it can be easily shown that the remaining terms are negligible. Finally, using Theorem 2.2 in Hualde and Robinson (2011), it is also possible to show that the hessian converges uniformly which ends the proof. □

References

Ando, T. and Bai, J. (2016a). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, -(just-accepted).

Ando, T. and Bai, J. (2016b). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31(1):163–191.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, pages 135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, 72(4):1127–1177.

Bailey, N., Kapetanios, G., and Pesaran, M. H. (2015). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Breitung, J. and Eickmeier, S. (2016). Analyzing international business and financial cycles using multi-level factor models: A comparison of alternative approaches. *Dynamic Factor Models (Advances in Econometrics, Volume 35) Emerald Group Publishing Limited*, 35:177–214.

Choi, I., Kim, D., Kim, Y. J., and Kwark, N.-S. (2016). A multilevel factor model: Identification, asymptotic theory and applications. Technical report.

Chudik, A. and Pesaran, M. H. (2015). Large panel data models with cross-sectional dependence: a survey. *The Oxford Handbook on Panel Data*.

Davidson, J. and Hashimzade, N. (2009). Type i and type ii fractional brownian motions: A reconsideration. *Computational Statistics & Data Analysis*, 53(6):2089–2106.

Diebold, F. X., Li, C., and Yue, V. Z. (2008). Global yield curve dynamics and interactions: a dynamic nelson–siegel approach. *Journal of Econometrics*, 146(2):351–363.

Ergemen, Y. E. (2016). System estimation of panel data models under long-range dependence. *CREATES Research Paper, 2016-2*.

Ergemen, Y. E. and Rodríguez-Caballero, C. V. (2016). A dynamic multi-level factor model with long-range dependence. *CREATES Research Paper, 2016-23*.

Ergemen, Y. E. and Velasco, C. (2015). Estimation of fractionally integrated panels with fixed effects and cross-section dependence. *CREATES Research Paper, 2015-35*.

Greenaway McGrevy, R., Han, C., and Sul, D. (2012). Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics*, 169(1):48–53.

Harding, M. and Lamarche, C. (2011). Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. *Economics Letters*, 111(3):197–199.

Hsiao, C. and Tahmiscioglu, A. K. (2008). Estimation of dynamic panel data models with both individual and time-specific effects. *Journal of Statistical Planning and Inference*, 138(9):2698–2721.

Hualde, J. and Robinson, P. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *The Annals of Statistics*, 39(6):3152–3181.

Kapetanios, G., Pesaran, M. H., and Yamagata, T. (2011). Panels with nonstationary multifactor error structures. *Journal of Econometrics*, 160(2):326–348.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors. *The American Economic Review*, 93(4):1216–1239.

Kose, M. A., Otrok, C., and Whiteman, C. H. (2008). Understanding the evolution of world business cycles. *Journal of international Economics*, 75(1):110–130.

Moon, H. R. and Perron, B. (2004). Testing for a unit root in panels with dynamic factors. *Journal of econometrics*, 122(1):81–126.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Phillips, P. C. and Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependence*. *The Econometrics Journal*, 6(1):217–259.

Phillips, P. C. and Sul, D. (2007). Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics*, 137(1):162–188.

Robinson, P. M. and Velasco, C. (2015). Efficient inference on fractionally integrated panel data models with fixed effects. *Journal of Econometrics*, 185(2):435–452.

Stock, J. H. and Watson, M. (2009a). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173:205.

Stock, J. H. and Watson, M. W. (2009b). The evolution of national and regional factors in us housing construction, volatility and time series econometrics: Essays in honour of Robert F. Engle.

Wang, P. (2010). Large dimensional factor models with a multi-level factor structure: Identification, estimation, and inference. *Working paper. Hong Kong University of Science and Technology*.

Westerlund, J. and Urbain, J.-P. (2013). On the estimation and inference in factor-augmented panel regressions with correlated loadings. *Economics Letters*, 119(3):247–250.

Westerlund, J. and Urbain, J.-P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics*, 185(2):372–377.

Research Papers 2016

2016-15: Mikkel Bennedsen, Ulrich Hounyo, Asger Lunde and Mikko S. Pakkanen: The Local Fractional Bootstrap

2016-16: Martin M. Andreasen and Kasper Jørgensen: Explaining Asset Prices with Low Risk Aversion and Low Intertemporal Substitution

2016-17: Robinson Kruse, Christian Leschinski and Michael Will: Comparing Predictive Accuracy under Long Memory - With an Application to Volatility Forecasting

2016-18: Søren Johansen and Bent Nielsen: Tightness of M-estimators for multiple linear regression in time series

2016-19: Tim Bollerslev, Jia Li and Yuan Xue: Volume, Volatility and Public News Announcements

2016-20: Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante: Retrieving Risk-Neutral Densities Embedded in VIX Options: a Non-Structural Approach

2016-21: Mikkel Bennedsen: Semiparametric inference on the fractal index of Gaussian and conditionally Gaussian time series data

2016-22: Søren Johansen and Morten Ørregaard Nielsen: The cointegrated vector autoregressive model with general deterministic terms

2016-23: Yunus Emre Ergemen and Carlos Vladimir Rodríguez-Caballero: A Dynamic Multi-Level Factor Model with Long-Range Dependence

2016-24: Shin Kanaya: Convergence rates of sums of α-mixing triangular arrays: with an application to non-parametric drift function

2016-25: Gustavo Fruet Dias, Marcelo Fernandes and Cristina M. Scherrer: Improving on daily measures of price discovery

2016-26: Martin M. Andreasen, Tom Engsted, Stig V. Møller and Magnus Sander: Bond Market Asymmetries across Recessions and Expansions: New Evidence on Risk Premia

2016-27: Kim Christensen, Ulrich Hounyo and Mark Podolskij: Testing for heteroscedasticity in jumpy and noisy high-frequency data: A resampling approach

2016-28: Kim Christensen, Roel Oomen and Roberto Renò: The Drift Burst Hypothesis

2016-29: Hossein Asgharian, Charlotte Christiansen, Rangan Gupta and Ai Jun Hou: Effects of Economic Policy Uncertainty Shocks on the Long-Run US-UK Stock Market Correlation

2016-30: Morten Ørregaard Nielsen and Sergei S. Shibaev: Forecasting daily political opinion polls using the fractionally cointegrated VAR model

2016-31: Carlos Vladimir Rodríguez-Caballero: Panel Data with Cross-Sectional Dependence Characterized by a Multi-Level Factor Structure