Multi-dimensional Panel Data Modelling in the Presence of Cross Sectional Error Dependence*

George Kapetanios
King's College, London

Camilla Mastromarco
University of Salento

Laura Serlenga
University of Bari

Yongcheol Shin
University of York

February 12, 2017

Abstract

Given the growing availability of the big dataset which contain information on multiple dimensions and following the recent research trend on the multidimensional modelling, we develop the 3D panel data models with three-way error components that allow for strong cross-section dependence (CSD) thorough unobserved heterogeneous global factors, and propose the consistent estimation procedure. We also discuss the extent of CSD in 3D models and provide a diagnostic test for cross-section dependence. We provide the extensions to unbalanced panels and 4D models. The validity of the proposed approach is confirmed by the Monte Carlo simulation results. We also demonstrate the empirical usefulness through the application to the 3D panel gravity model of the intra-EU trade flows.

*Address for correspondence: Y. Shin, A/EC/119 Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, United Kingdom. Email: yongcheol.shin@york.ac.uk. Tel: +44 (0)1904 323757. We are grateful for the insightful comments of the editor, Laszlo Matyas. This work has beneted greatly from the stimulating discussion of delegates at the PANDA Conference at University of York, July 2016 as well as the many detailed comments raised by seminar participants at Universities of Sogang and York. The usual disclaimer applies.

1 Introduction

Given the growing availability of the big dataset which contain information on multiple dimensions, the recent literature on the panel data has focused more on extending the two-way error components models to the multidimensional setting. Balazsi, Matyas and Wansbeek (2015, BMW hereafter) introduce the appropriate within estimators for the most frequently used three-dimensional (3D) fixed effects panel data models, and Balazsi, Baltagi, Matyas and Pus (2016, BBMP hereafter) consider the random effects approach and propose a sequence of optimal GLS estimators. This multi-dimensional approach is expected to become an essential tool for the analysis of complex interconnectedness of the big dataset, and it can be not only applied to the number of bilateral (origin-destination) flows such as trade, FDI, capital or migration flows (e.g. Feenstra, 2004; Bertoli and Fernandez-Huertas Moraga, 2013; Gunnella et al., 2015), but also to a variety of matched dataset which may link the employer-the employee and the pupils-teachers (e.g. Abowd et al., 1999; Kramarz et al., 2008).

However, there has been no study attempting to address an important issue of explicitly controlling cross-sectional error dependence in 3D or higher-dimensional panel data, even though the cross-section dependence (CSD) seems pervasive in 2D panels because it seems rare that the cross-section covariance of the errors is zero (e.g. Pesaran, 2015). Recently, there has been much progress in modelling CSD in 2D panels by two main approaches, the factor-based approach (e.g. Pesaran, 2006; Bai, 2009) and the spatial econometrics techniques (e.g. Behrens et al., 2012; Mastromarco et al., 2016b). Chudik et al. (2011) show that the factor-based models exhibit the strong CSD whilst the spatial-based models can deal with weak CSD only. See also Bailey, Kapetanios and Pesaran (2016) for more general discussions.

Chapter 3 by Le Gallo and Pirotte (2016) reviews the current state-of-art in the analysis of multi-dimensional nested spatial panels, highlighting a range of issues related to the specification, estimation, testing procedures and predictions. Chapter 10 by Baltagi, Egger and Erhardt (2016) provides a survey of empirical issues in the analysis of gravity-model estimation of international trade flows, proceeds with the modelling of the multidimensional stochastic structure, focusing on fixed-effects estimation, and describes how the spatial autocorrelation and spillovers can be introduced into such models. Chapter 12 by Baltagi and Bresson (2016) surveys hedonic housing models and discrete choice models using multi-dimensional panels, also focussing on the spatial econometrics approach.

Following this research trend, we develop the 3D panel data models with strong CSD. In particular, we generalise the multi-dimensional error components specification by modelling residual CSD via unobserved heterogeneous global factors. The multidimensional country-time fixed (CTFE) and random effects (CTRE) estimators proposed by BMW and BBMP fail to remove heterogenous global factors, suggesting that they are biased in the presence of the nonzero correlation between the regressors and unobserved global factors. In this regard, we develop the two-step consistent estimation procedure. First, we follow Pesaran (2006) and augment the 3D model with the cross-section averages of dependent variable and regressors over double cross-section units, which are shown to provide valid proxies for unobserved heterogenous global factors. Next, we apply the 3D-within transformation to the augmented specification and obtain consistent estimators, called the 3D-PCCE estimator. Our approach is the first attempt to accommodate strong CSD in multi-dimensional panels, and expected to make timely contribution to the growing literature.

We discuss the extent of CSD within the 3D panel data models under the three different error components specifications respectively with the CTFE, with the two-way heterogeneous factor, and with both components. We also distinguish between three types of CSD under the hierarchical multi-factor error components specification recently advanced by Kapetanios and Shin (2017). First, the global factor tends to display strong CSD as it influences the (ij) pairwise interactions for $i = 1, ..., N_1$ and $j = 1, ..., N_2$ (of $N_1 N_2$ dimension). Next, the local factors show semi-strong or semi-weak CSD, as they influence origin and destination countries separately (each of N_1 or N_2 dimension). Finally, idiosyncratic errors are characterised with the weak or no CSD.

We then develop a diagnostic test for the null hypothesis of (pairwise) residual cross-section independence or weak dependence in the 3D panels, which is a modified counterpart of an existing CD test in the 2D panels proposed by Pesaran (2015). (**CSD exponent**) Furthermore, we provide a couple of extensions into unbalanced panels and 4D or higher dimensional models.

We have conducted the Monte Carlo studies to investigate the small sample properties of the 3D-PCCE estimators relative to the CTFE estimator. We find strong evidence that the 3D-PCCE estimators perform well when the 3D panel data is subject to the strong CSD through heterogeneous global factors. On the contrary, the CTFE estimator tends to display severe biases and size distortions.

We apply our proposed 3D PCCE estimation techniques, together with the two-way fixed effects and the CTFE estimators, to the dataset over the period 1960-2008 (49 years) for 91 country-pairs amongst 14 EU countries. Based on the CD test results, estimates of CSD exponent, and the predicted signs and statistical significance of the coefficients, we come to a conclusion that the 3D PCCE estimation results are mostly satisfactory and reliable. In particular, when we explicitly control for strong CSD in the 3D panels, we find that the trade effect of currency union is rather modest. This evidence provides strong support for the thesis that the trade increase within the Euro area may reflect a continuation of a long-run historical trend linked to the broader set of EU's economic integration policies.

This Chapter proceeds in 7 Sections. Section 16.2 introduces 3D models with three-way error components that allow for strong cross-section dependence, and develops the consistent estimation procedure. Section 16.3 discusses the nature of CSD in 3D models and provide a diagnostic test for cross-section dependence. Section 16.4 presents the extension to 4D models. Section 16.5 discusses the Monte Carlo simulation results. The empirical results for the gravity model of EU exports flows are presented in Section 16.6, while Section 16.7 concludes.

Throughout the chapter we adopt the following standard notations. $\mathbf{I}_N$ is an $N \times N$ identity matrix, $\mathbf{J}_N$ the $N \times N$ identity matrix of ones, and $\boldsymbol{\iota}_N$ the $N \times 1$ vector of ones, respectively. $\mathbf{M}_A$ projects the $N \times N$ matrix $\mathbf{A}$ into its null-space, i.e., $\mathbf{M}_A = \mathbf{I}_N - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. Finally, $y_{.jt} = N_1^{-1}\sum_{i=1}^{N_1} y_{ijt}$, $y_{i.t} = N_2^{-1}\sum_{j=1}^{N_2} y_{ijt}$ and $y_{ij.} = T^{-1}\sum_{t=1}^{T} y_{ijt}$ denote the average of y over the index i, j and t, respectively, with the definition extending to other quantities such as $y_{..t}$, $y_{.j.}$, $y_{i..}$ and $y_{...}$.

2 3D Models with Cross-sectional Error Dependence

Following Balazsi, Matyas and Wansbeek (2015, BMW hereafter) and Balazsi, Baltagi, Matyas and Pus (2016, BBMP hereafter), we consider the following three-dimensional country-time fixed effects panel data model:

$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \boldsymbol{\gamma}'\mathbf{s}_{it} + \boldsymbol{\delta}'\mathbf{d}_{jt} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\mathbf{z}_{ij} + u_{ijt}, i = 1, ..., N_1, j = 1, ..., N_2, t = 1, ..., T, \tag{1}$$

with the error components:

$$u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \varepsilon_{ijt} \tag{2}$$

where y_{ijt} is the dependent variable observed across three indices (e.g. the import of country j from country i at period t), $\mathbf{x}_{ijt}$, $\mathbf{s}_{it}$, $\mathbf{d}_{jt}$, $\mathbf{q}_t$, $\mathbf{z}_{ij}$ are the $k_x \times 1$, $k_s \times 1$, $k_d \times 1$, $k_q \times 1$, $k_z \times 1$ vectors of covariates covering all possible measurements observed across three indices, and $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\kappa}$, $\boldsymbol{\varphi}$, are the associated vectors of the parameters. The multiple error components in (2) contain bilateral pair-fixed effects (μ_{ij}) as well as origin and destination country-time fixed effects (CTFE), v_{it} and ζ_{jt}, respectively.[1]

To remove all unobserved fixed effects, μ_{ij}, v_{it} and ζ_{jt}, BMW derive the following 3D within transformation:[2]

$$\tilde{y}_{ijt} = y_{ijt} - \bar{y}_{ij.} - y_{.jt} - \bar{y}_{i.t} + \bar{y}_{..t} + \bar{y}_{.j.} + \bar{y}_{i..} - \bar{y}_{...} \tag{3}$$

Applying the 3D within transformation to (1), we can estimate consistently $\boldsymbol{\beta}$ only from the following regression:

$$\tilde{y}_{ijt} = \boldsymbol{\beta}'\tilde{\mathbf{x}}_{ijt} + \tilde{\varepsilon}_{ijt}, \quad i = 1, ..., N_1, \ j = 1, ..., N_2, \ t = 1, ..., T, \tag{4}$$

where $\tilde{\mathbf{x}}_{ijt} = \mathbf{x}_{ijt} - \bar{\mathbf{x}}_{ij.} - \bar{\mathbf{x}}_{.jt} - \bar{\mathbf{x}}_{i.t} + \bar{\mathbf{x}}_{..t} + \bar{\mathbf{x}}_{.j.} + \bar{\mathbf{x}}_{i..} - \bar{\mathbf{x}}_{...}$ and similarly for $\tilde{\varepsilon}_{ijt}$. We write (4) compactly as

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij}\boldsymbol{\beta} + \tilde{\mathbf{E}}_{ij} \tag{5}$$

[1] Notice that the error component specification (2) is proposed by Baltagi et al. (2003).

[2] Baltagi et al. (2015) also derive the same projection by applying Davis' (2002) Lemma twice (see Corollary 1).

where

$$\underset{T\times 1}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij1} \\ \vdots \\ \tilde{y}_{ijT} \end{bmatrix}, \ \underset{T\times k_x}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ij1} \\ \vdots \\ \tilde{\mathbf{x}}'_{ijT} \end{bmatrix}, \ \underset{T\times 1}{\tilde{\mathbf{E}}_{ij}} = \begin{bmatrix} \tilde{\varepsilon}_{ij1} \\ \vdots \\ \tilde{\varepsilon}_{ijT} \end{bmatrix}.$$

The 3D-within estimator of $\boldsymbol{\beta}$ is obtained by

$$\hat{\boldsymbol{\beta}}_W = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \tilde{\mathbf{X}}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \tilde{\mathbf{Y}}_{ij} \right). \tag{6}$$

Then, it follows that, as $(N_1, N_2, T) \to \infty$ (see also BBMP),

$$\sqrt{N_1 N_2 T} \left(\hat{\boldsymbol{\beta}}_W - \boldsymbol{\beta} \right) \overset{a}{\sim} N \left(\mathbf{0}, \sigma_\varepsilon^2 \lim_{(N_1, N_2, T)\to\infty} \left(\frac{1}{N_1 N_2 T} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \tilde{\mathbf{X}}'_{ij} \tilde{\mathbf{X}}_{ij} \right)^{-1} \right).$$

By construction, the within transformation in (3) wipes out all other covariates, $\mathbf{x}_{it}$, $\mathbf{x}_{jt}$, $\mathbf{x}_t$, and $\mathbf{x}_{ij}$ in (1). But, we may be interested in uncovering the effects of those covariates (e.g. the impacts of measured trade costs in the structural gravity model). In order to recover those coefficients, it would be worthwhile to develop an extension of the Hausman-Taylor (1981) estimation, which has been popular in the two-way panel data models even in the presence of cross sectionally correlated errors (e.g. Serlenga and Shin, 2007). Chapter 3 by Balazsi, Bun, Chan and Harris (2016) develops an extended Hausman-Taylor estimator for multi-dimensional panel data models.

BMW also show that the CTFE error components in (2) nests the number of special cases by applying suitable restrictions to (2).[3] Notice, however, that the model (1) with (2) does not address an important issue of cross-sectional error dependence. In the presence of such cross-section dependence (CSD), the 3D-within estimator would be likely to be biased. In this regard, we consider a couple of alternative 3-way error components specifications that can accommodate CSD, and develop the appropriate estimation techniques.

Given that v_{it} and ζ_{jt} are supposed to measure the (local) origin and destination country-time fixed effects, it is natural to add the global factor λ_t to (2):

$$u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \lambda_t + \varepsilon_{ijt}$$

But, the 3D-within transformation, (3) removes λ_t together with μ_{ij}, v_{it} and ζ_{jt}, because λ_t is shown to be proportional to $\sum_{i=1}^{N_1} v_{it}$ or $\sum_{j=1}^{N_2} \zeta_{jt}$ (see also footnote 6 below).

To introduce strong CSD explicitly in the 3D model, (1), we first consider the following error components specification:

$$u_{ijt} = \mu_{ij} + \pi_{ij}\lambda_t + \varepsilon_{ijt}. \tag{7}$$

[3] Baltagi et al. (2003), Baldwin and Taglioni (2006), and Baier and Bergstrand (2007) consider several forms of fixed effects such as $u_{ijt} = \alpha_i + \gamma_j + \lambda_t + \varepsilon_{ijt}$, $u_{ijt} = \mu_{ij} + \lambda_t + \varepsilon_{ijt}$, $u_{ijt} = \zeta_{jt} + \varepsilon_{ijt}$, $u_{ijt} = v_{it} + \varepsilon_{ijt}$, and $u_{ijt} = v_{it} + \zeta_{jt} + \varepsilon_{ijt}$.

This is similar to the two-way heterogeneous factor model considered by Serlenga and Shin (2007). We follow Pesaran (2006) and apply the cross-section averages of (1) and (7) over i and j to obtain:

$$\begin{aligned} \bar{y}_{..t} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\left(\boldsymbol{\beta}'\mathbf{x}_{ijt}+\boldsymbol{\gamma}'\mathbf{s}_{it}+\boldsymbol{\delta}'\mathbf{d}_{jt}+\boldsymbol{\kappa}'\mathbf{q}_t+\boldsymbol{\varphi}'\mathbf{z}_{ij}+\mu_{ij}+\pi_{ij}\lambda_t+\varepsilon_{ijt}\right) \\ &= \boldsymbol{\beta}'\bar{\mathbf{x}}_{..t}+\boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t}+\boldsymbol{\delta}'\bar{\mathbf{d}}_{.t}+\boldsymbol{\kappa}'\mathbf{q}_t+\boldsymbol{\varphi}'\bar{\mathbf{z}}_{..}+\bar{\mu}_{..}+\bar{\pi}_{..}\lambda_t+\bar{\varepsilon}_{..t} \end{aligned} \quad (8)$$

where $\bar{\mathbf{s}}_{.t}=N_1^{-1}\sum_{i=1}^{N_1}\mathbf{s}_{it}$, $\bar{\mathbf{d}}_{.t}=N_2^{-1}\sum_{j=1}^{N_2}\mathbf{d}_{jt}$, $\bar{\mathbf{z}}_{..}=(N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathbf{z}_{ij}$, $\bar{\mu}_{..}=(N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mu_{ij}$ and $\bar{\pi}_{..}=(N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\pi_{ij}$. Hence,

$$\lambda_t=\frac{1}{\bar{\pi}_{..}}\left\{\bar{y}_{..t}-\left(\boldsymbol{\beta}'\bar{\mathbf{x}}_{..t}+\boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t}+\boldsymbol{\delta}'\bar{\mathbf{d}}_{.t}+\boldsymbol{\kappa}'\mathbf{q}_t+\boldsymbol{\varphi}'\bar{\mathbf{z}}_{..}+\bar{\mu}_{..}+\bar{\varepsilon}_{..t}\right)\right\}$$

Using these results we can augment the model (1) with the cross-section averages as follows:

$$y_{ijt}=\boldsymbol{\beta}'\mathbf{x}_{ijt}+\boldsymbol{\gamma}'\mathbf{s}_{it}+\boldsymbol{\delta}'\mathbf{d}_{jt}+\boldsymbol{\psi}_{ij}'\mathbf{f}_t+\tau_{ij}+\mu_{ij}^*+\varepsilon_{ijt}^*, \quad (9)$$

where

$$\boldsymbol{\psi}_{ij}'=\left(\psi_{0ij},\boldsymbol{\psi}_{1ij}',\boldsymbol{\psi}_{2ij}',\boldsymbol{\psi}_{3ij}',\boldsymbol{\psi}_{4ij}'\right)=\left(\frac{\pi_{ij}}{\bar{\pi}_{..}},\frac{-\pi_{ij}\boldsymbol{\beta}'}{\bar{\pi}_{..}},\frac{-\pi_{ij}\boldsymbol{\gamma}'}{\bar{\pi}_{..}},\frac{-\pi_{ij}\boldsymbol{\delta}'}{\bar{\pi}_{..}},\left(1-\frac{\pi_{ij}}{\bar{\pi}_{..}}\right)\boldsymbol{\kappa}'\right)$$

$$\mathbf{f}_t=\left(\bar{y}_{..t},\bar{\mathbf{x}}_{..t}',\bar{\mathbf{s}}_{.t}',\bar{\mathbf{d}}_{.t}',\mathbf{q}_t'\right)' \quad (10)$$

$$\tau_{ij}=\boldsymbol{\varphi}'\mathbf{z}_{ij}-\frac{-\pi_{ij}}{\bar{\pi}_{..}}\boldsymbol{\varphi}'\mathbf{z}_{..},\ \mu_{ij}^*=\mu_{ij}-\frac{\pi_{ij}\mu_{..}}{\bar{\pi}_{..}},\ \varepsilon_{ijt}^*=\varepsilon_{ijt}-\frac{\pi_{ij}}{\bar{\pi}_{..}}\bar{\varepsilon}_{..t}.$$

We write (9) compactly as

$$\begin{aligned} \mathbf{Y}_{ij} &= \mathbf{X}_{ij}\boldsymbol{\beta}+\mathbf{S}_i\boldsymbol{\gamma}+\mathbf{D}_j\boldsymbol{\delta}+\mathbf{F}\boldsymbol{\psi}_{ij}+\tau_{ij}\iota_T+\mu_{ij}^*\iota_T+\mathbf{E}_{ij}^* \\ &= \mathbf{W}_{ij}\boldsymbol{\theta}+\mathbf{H}\boldsymbol{\psi}_{ij}^*+\mathbf{E}_{ij}^*,\ i=1,...,N_1, j=1,...,N_2 \end{aligned} \quad (11)$$

where

$$\underset{T\times 1}{\mathbf{Y}_{ij}}=\begin{bmatrix} y_{ij1} \\ \vdots \\ y_{ijT}\end{bmatrix},\ \underset{T\times k_x}{\mathbf{X}_{ij}}=\begin{bmatrix} \mathbf{x}_{ij1}' \\ \vdots \\ \mathbf{x}_{ijT}'\end{bmatrix},\ \underset{T\times k_s}{\mathbf{S}_i}=\begin{bmatrix} \mathbf{s}_{i1}' \\ \vdots \\ \mathbf{s}_{iT}'\end{bmatrix},$$

$$\underset{T\times k_d}{\mathbf{D}_j}=\begin{bmatrix} \mathbf{d}_{j1}' \\ \vdots \\ \mathbf{d}_{jT}'\end{bmatrix},\ \underset{T\times k_f}{\mathbf{F}}=\begin{bmatrix} \mathbf{f}_1' \\ \vdots \\ \mathbf{f}_T'\end{bmatrix},\ \underset{T\times 1}{\mathbf{E}_{ij}^*}=\begin{bmatrix} \varepsilon_{ij1}^* \\ \vdots \\ \varepsilon_{ijT}^*\end{bmatrix},$$

$\mathbf{W}_{ij}=\left(\mathbf{X}_{ij},\mathbf{S}_i,\mathbf{D}_j\right)$, $\boldsymbol{\theta}=\begin{pmatrix}\boldsymbol{\beta}' & \boldsymbol{\gamma}' & \boldsymbol{\delta}'\end{pmatrix}'$, $\boldsymbol{\psi}_{ij}^*=\left(\boldsymbol{\psi}_{ij}',\left(\tau_{ij}+\mu_{ij}^*\right)\right)'$ and $\mathbf{H}=[\mathbf{F},\boldsymbol{\iota}_T]$. Then, we derive the consistent estimator of $\boldsymbol{\theta}$ (called 3D-PCCE) by[4]

$$\hat{\boldsymbol{\theta}}_{PCCE}=\left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathbf{W}_{ij}'\mathbf{M}_H\mathbf{W}_{ij}\right)^{-1}\left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathbf{W}_{ij}'\mathbf{M}_H\mathbf{Y}_{ij}\right) \quad (12)$$

[4] $\boldsymbol{\kappa}$ and $\boldsymbol{\varphi}$ cannot be identified due to the factor approximations and the within transformation.

where $\mathbf{M}_H = \mathbf{I}_T - \mathbf{H}\left(\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'$. Following Pesaran (2006), it is straightforward to show that as $(N_1, N_2, T) \to \infty$, the PCCE estimator, (12) follows the asymptotic normal distribution (see also Kapetanios and Shin, 2017):

$$\sqrt{N_1 N_2 T}\left(\hat{\boldsymbol{\theta}}_{PCCE} - \boldsymbol{\theta}\right) \overset{a}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_\theta\right),$$

where the (robust) consistent estimator of $\boldsymbol{\Sigma}_\theta$ is given by

$$\hat{\boldsymbol{\Sigma}}_\theta = \frac{1}{N_1 N_2}\mathbf{S}_\theta^{-1}\mathbf{R}_\theta\mathbf{S}_\theta^{-1},$$

$$\mathbf{R}_\theta = \frac{1}{N_1 N_2 - 1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\mathbf{W}'_{ij}\mathbf{M}_H\mathbf{W}_{ij}}{T}\right)\left(\hat{\boldsymbol{\theta}}_{ij} - \hat{\boldsymbol{\theta}}_{MG}\right)\left(\hat{\boldsymbol{\theta}}_{ij} - \hat{\boldsymbol{\theta}}_{MG}\right)'\left(\frac{\mathbf{W}'_{ij}\mathbf{M}_H\mathbf{W}_{ij}}{T}\right),$$

$$\mathbf{S}_\theta = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\mathbf{W}'_{ij}\mathbf{M}_H\mathbf{W}_{ij}}{T}\right),\ \hat{\boldsymbol{\theta}}_{MG} = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\hat{\boldsymbol{\theta}}_{ij},$$

where $\hat{\boldsymbol{\theta}}_{ij}$ is the (ij) pairwise OLS estimator obtained from the individual regression of $\mathbf{Y}_{ij}$ on $(\mathbf{W}_{ij}, \mathbf{H})$ in (11) for $i = 1, ..., N_1$ and $j = 1, ..., N_2$.

Next, we consider the 3D model (1) with the following general error components by combining CTFEs and heterogeneous global factors:

$$u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt}. \tag{13}$$

It is straightforward to show that the 3D-within transformation (3) fails to remove heterogeneous factors $\pi_{ij}\lambda_t$, because it is easily seen that

$$\tilde{u}_{ijt} = \tilde{\pi}_{ij}\tilde{\lambda}_t + \tilde{\varepsilon}_{ijt}$$

where $\tilde{\lambda}_t = \lambda_t - \bar{\lambda}$ with $\bar{\lambda} = T^{-1}\sum_{t=1}^T \lambda_t$ and $\tilde{\pi}_{ij} = \pi_{ij} - \pi_{.j} - \pi_{i.} + \pi_{..}$ with $\pi_{.j} = N_1^{-1}\sum_{i=1}^{N_1}\pi_{ij}$ and $\pi_{i.} = N_2^{-1}\sum_{j=1}^{N_2}\pi_{ij}$.[5] It is clear in the presence of the nonzero correlation between $\mathbf{x}_{ijt}$ and λ_t that the 3D-within estimator of $\boldsymbol{\beta}$ is biased.

We develop the two-step consistent estimation procedure. First, taking the cross-section averages of (1) and (13) over i and j, we have:

$$\bar{y}_{..t} = \boldsymbol{\beta}'\bar{\mathbf{x}}_{..t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\mathbf{z}_{..} + \mu_{..} + \bar{v}_{.t} + \bar{\zeta}_{.t} + \bar{\pi}_{..}\lambda_t + \bar{\varepsilon}_{..t} \tag{14}$$

where $\bar{v}_{.t} = N_1^{-1}\sum_{i=1}^{N_1} v_{it}$, $\bar{\zeta}_{.t} = N_2^{-1}\sum_{j=1}^{N_2}\zeta_{jt}$ and see (8) for other definitions. Hence, we augment the model (1) with the cross-section averages as:

$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \boldsymbol{\gamma}'\mathbf{s}_{it} + \boldsymbol{\delta}'\mathbf{d}_{jt} + \boldsymbol{\psi}'_{ij}\mathbf{f}_t + \tau_{ij} + \mu^*_{ij} + v^*_{ijt} + \zeta^*_{ijt} + \varepsilon^*_{ijt}, \tag{15}$$

where $v^*_{ijt} = v_{it} - \frac{\pi_{ij}\bar{v}_{.t}}{\bar{\pi}_{..}}$, $\zeta^*_{ijt} = \zeta_{jt} - \frac{\pi_{ij}\bar{\zeta}_{.t}}{\bar{\pi}_{..}}$, and see (11) for other definitions. We rewrite (15) as

$$y_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \boldsymbol{\gamma}'\mathbf{s}_{it} + \boldsymbol{\delta}'\mathbf{d}_{jt} + \boldsymbol{\psi}'_{ij}\mathbf{f}_t + \tau_{ij} + \mu^*_{ij} + v_{it} + \zeta_{jt} + \varepsilon^{**}_{ijt}, \tag{16}$$

[5] Unless $\tilde{\pi}_{ij} = 0, \tilde{u}_{ijt} \neq \tilde{\varepsilon}_{ijt}$. This holds only if factor loadings, π_{ij} are homogeneous for all (i, j) pairs.

where $\varepsilon_{ijt}^{**} = \varepsilon_{ijt} - \frac{\pi_{ij}}{\bar{\pi}_{..}}\bar{\varepsilon}_{..t} - \frac{\pi_{ij}\bar{v}_{.t}}{\bar{\pi}_{..}} - \frac{\pi_{ij}\bar{\zeta}_{.t}}{\bar{\pi}_{..}}$. Notice that as $N_1, N_2 \to \infty$, $\varepsilon_{ijt}^{**} \to_p \varepsilon_{ijt}$ since $\bar{v}_{.t} \to_p 0$, $\bar{\zeta}_{.t} \to_p 0$ and $\bar{\varepsilon}_{..t} \to_p 0$. Next, we apply the 3D-within transformation (3) to (16), and obtain:[6]

$$\tilde{y}_{ijt} = \boldsymbol{\beta}'\tilde{\mathbf{x}}_{ijt} + \tilde{\boldsymbol{\psi}}_{ij}'\tilde{\mathbf{f}}_t + \tilde{\varepsilon}_{ijt}^{**}, \tag{17}$$

where $\tilde{\boldsymbol{\psi}}_{ij} = \boldsymbol{\psi}_{ij} - \boldsymbol{\psi}_{.j} - \boldsymbol{\psi}_{j.} + \boldsymbol{\psi}_{..}$, $\tilde{\mathbf{f}}_t = \mathbf{f}_t - \bar{\mathbf{f}}$ with $\bar{\mathbf{f}} = T^{-1}\sum_{t=1}^{T}\mathbf{f}_t$, and $\mathbf{f}_t$ is defined in (10). Rewriting (17) compactly as

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij}\boldsymbol{\beta} + \tilde{\mathbf{F}}\tilde{\boldsymbol{\psi}}_{ij} + \tilde{\mathbf{E}}_{ij}^{**},\ i = 1, ..., N_1, j = 1, ..., N_2 \tag{18}$$

where

$$\underset{T\times 1}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij1} \\ \vdots \\ \tilde{y}_{ijT} \end{bmatrix},\ \underset{T\times k_x}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}_{ij1}' \\ \vdots \\ \tilde{\mathbf{x}}_{ijT}' \end{bmatrix},\ \underset{T\times k_f}{\tilde{\mathbf{F}}} = \begin{bmatrix} \tilde{\mathbf{f}}_1' \\ \vdots \\ \tilde{\mathbf{f}}_T' \end{bmatrix},\ \underset{T\times 1}{\tilde{\mathbf{E}}_{ij}^{**}} = \begin{bmatrix} \tilde{\varepsilon}_{ij1}^{**} \\ \vdots \\ \tilde{\varepsilon}_{ijT}^{**} \end{bmatrix}.$$

Then, the 3D-PCCE estimator of $\boldsymbol{\beta}$ is obtained by

$$\hat{\boldsymbol{\beta}}_{PCCE} = \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}\right)^{-1}\left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{Y}}_{ij}\right) \tag{19}$$

where $\mathbf{M}_{\tilde{F}} = \mathbf{I}_T - \tilde{\mathbf{F}}\left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}}\right)^{-1}\tilde{\mathbf{F}}'$ is the $T \times T$ idempotent matrix. Following Pesaran (2006) and Kapetanios and Shin (2017), it is also straightforward to show that as $(N_1, N_2, T) \to \infty$, the PCCE estimator, (19) follows the asymptotic normal distribution:

$$\sqrt{N_1N_2T}\left(\hat{\boldsymbol{\beta}}_{PCCE} - \boldsymbol{\beta}\right) \overset{a}{\sim} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{\beta}\right),$$

where the (robust) consistent estimator of $\boldsymbol{\Sigma}_{\beta}$ is given by

$$\hat{\boldsymbol{\Sigma}}_{\beta} = \frac{1}{N^2}\mathbf{S}_{\beta}^{-1}\mathbf{R}_{\beta}\mathbf{S}_{\beta}^{-1},$$

$$\mathbf{R}_{\beta} = \frac{1}{N_1N_2 - 1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}}{T}\right)\left(\hat{\boldsymbol{\beta}}_{ij} - \hat{\boldsymbol{\beta}}_{MG}\right)\left(\hat{\boldsymbol{\beta}}_{ij} - \hat{\boldsymbol{\beta}}_{MG}\right)'\left(\frac{\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}}{T}\right),$$

$$\mathbf{S}_{\beta} = \frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left(\frac{\tilde{\mathbf{X}}_{ij}'\mathbf{M}_{\tilde{F}}\tilde{\mathbf{X}}_{ij}}{T}\right),\ \hat{\boldsymbol{\beta}}_{MG} = \frac{1}{N_1N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\hat{\boldsymbol{\beta}}_{ij},$$

[6] It is clear that $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\kappa}$, and $\boldsymbol{\varphi}$ cannot be identified due to the 3D-within transformation and the factor approximation. Define $\boldsymbol{\theta}_{ijt} = \boldsymbol{\psi}_{ij}'\mathbf{f}_t$, then it is straightforward to show that

$$\tilde{\boldsymbol{\theta}}_{ijt} = \boldsymbol{\theta}_{ijt} - (\bar{\boldsymbol{\theta}}_{ij.} + \bar{\boldsymbol{\theta}}_{.jt} + \bar{\boldsymbol{\theta}}_{i.t}) + (\bar{\boldsymbol{\theta}}_{..t} + \bar{\boldsymbol{\theta}}_{i..} + \bar{\boldsymbol{\theta}}_{.j.}) - \bar{\boldsymbol{\theta}}_{...} = (\boldsymbol{\psi}_{ij} - \boldsymbol{\psi}_{.j} - \boldsymbol{\psi}_{j.} + \boldsymbol{\psi}_{..})'(\mathbf{f}_t - \bar{\mathbf{f}}).$$

where $\hat{\boldsymbol{\beta}}_{ij}$ is the (ij) pairwise OLS estimator obtained from the individual regression of $\tilde{\mathbf{Y}}_{ij}$ on $\left(\tilde{\mathbf{X}}_{ij}, \tilde{\mathbf{F}}\right)$ in (18) for $i = 1, ..., N_1$ and $j = 1, ..., N_2$.

We can extend the proposed approach to the 3D panels with heterogeneous slope parameters:

$$y_{ijt} = \boldsymbol{\beta}_{ij}' \mathbf{x}_{ijt} + \boldsymbol{\gamma}_j' \mathbf{s}_{it} + \boldsymbol{\delta}_i' \mathbf{d}_{jt} + \boldsymbol{\kappa}_{ij}' \mathbf{q}_t + \boldsymbol{\varphi}' \mathbf{z}_{ij} + u_{ijt},\ i = 1, ..., N_1, j = 1, ..., N_2, t = 1, ..., T. \tag{20}$$

In this case we can develop the mean group estimators for (2), (7) and (13) in a straightforward manner (e.g. Pesaran, 2006; Kapetanios and Shin, 2017):

$$\hat{\boldsymbol{\beta}}_{W,MG} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(\tilde{\mathbf{X}}_{ij}' \tilde{\mathbf{X}}_{ij}\right)^{-1} \left(\tilde{\mathbf{X}}_{ij}' \mathbf{Y}_{ij}\right)$$

$$\hat{\boldsymbol{\theta}}_{MGCCE} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(\mathbf{W}_{ij}' \mathbf{M}_H \mathbf{W}_{ij}\right)^{-1} \left(\mathbf{W}_{ij}' \mathbf{M}_H \mathbf{Y}_{ij}\right)$$

$$\hat{\boldsymbol{\beta}}_{MGCCE} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(\tilde{\mathbf{X}}_{ij}' \mathbf{M}_{\tilde{F}} \tilde{\mathbf{X}}_{ij}\right)^{-1} \left(\tilde{\mathbf{X}}_{ij}' \mathbf{M}_{\tilde{F}} \tilde{\mathbf{Y}}_{ij}\right)$$

3 Cross-section Dependence (CD) Test

We discuss the extent of cross-section dependence in the 3D panel data models. Following Pesaran (2015) and Bailey et al. (2016, hereafter BKP), we can show that the extent of CSD is captured by non-zero covariance between u_{ijt} and $u_{i'j't}$ for $i \neq i'$ and $j \neq j$, denoted as $\sigma_{ijt,u}$. Here, the extent of CSD involves both N_1 and N_2, and thus relates to the rate at which $\frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sigma_{ijt,u}$ declines with the product, $N_1 N_2$.

First, we consider the 3D model (1) with country-time effects (2). Following BBMP, we make the following random effects assumptions:

$$\mu_{ij} \sim iid\left(0, \sigma_\mu^2\right),\ v_{it} \sim iid\left(0, \sigma_v^2\right),\ \zeta_{jt} \sim iid\left(0, \sigma_\zeta^2\right),\ \varepsilon_{ijt} \sim iid\left(0, \sigma_\varepsilon^2\right)$$

and μ_{ij}, v_{it}, ζ_{jt} and ε_{ijt} are pairwise uncorrelated. Rewrite (2) sequentially as

$$\underset{T\times 1}{\mathbf{u}_{ij}} = \mu_{ij} \boldsymbol{\iota}_T + \mathbf{v}_i + \boldsymbol{\zeta}_j + \boldsymbol{\varepsilon}_{ij},\ i = 1, ..., N_1,\ j = 1, ..., N_2,$$

$$\underset{N_2 T\times 1}{\mathbf{u}_i} = \boldsymbol{\mu}_i \otimes \boldsymbol{\iota}_T + \boldsymbol{\iota}_{N_2} \otimes \mathbf{v}_i + \boldsymbol{\zeta} + \boldsymbol{\varepsilon}_i, i = 1, ..., N_1,$$

$$\underset{N_1 N_2 T\times 1}{\mathbf{u}} = \boldsymbol{\mu} \otimes \boldsymbol{\iota}_T + \mathbf{V} + \boldsymbol{\iota}_{N_1} \otimes \boldsymbol{\zeta} + \boldsymbol{\varepsilon} \tag{21}$$

where

$$\underset{T\times 1}{\mathbf{u}_{ij}} = \begin{bmatrix} u_{ij1} \\ \vdots \\ u_{ijT} \end{bmatrix},\ \underset{T\times 1}{\mathbf{v}_i} = \begin{bmatrix} v_{i1} \\ \vdots \\ v_{iT} \end{bmatrix},\ \underset{T\times 1}{\boldsymbol{\zeta}_j} = \begin{bmatrix} \zeta_{j1} \\ \vdots \\ \zeta_{jT} \end{bmatrix},\ \underset{T\times 1}{\boldsymbol{\varepsilon}_{ij}} = \begin{bmatrix} \varepsilon_{ij1} \\ \vdots \\ \varepsilon_{ijT} \end{bmatrix},$$

$$\underset{N_2T\times 1}{\mathbf{u}_i}=\begin{bmatrix}\mathbf{u}_{i1}\\ \vdots\\ \mathbf{u}_{iN_2}\end{bmatrix},\ \underset{N_2\times 1}{\boldsymbol{\mu}_i}=\begin{bmatrix}\mu_{i1}\\ \vdots\\ \mu_{iN_2}\end{bmatrix},\ \underset{N_2T\times 1}{\boldsymbol{\zeta}}=\begin{bmatrix}\boldsymbol{\zeta}_1\\ \vdots\\ \boldsymbol{\zeta}_{N_2}\end{bmatrix},\ \underset{N_2T\times 1}{\boldsymbol{\varepsilon}_i}=\begin{bmatrix}\boldsymbol{\varepsilon}_{i1}\\ \vdots\\ \boldsymbol{\varepsilon}_{iN_2}\end{bmatrix},$$

$$\underset{N_1N_2T\times 1}{\mathbf{u}}=\begin{bmatrix}\mathbf{u}_{1}\\ \vdots\\ \mathbf{u}_{N_1}\end{bmatrix},\ \underset{N_1N_2\times 1}{\boldsymbol{\mu}}=\begin{bmatrix}\boldsymbol{\mu}_{1}\\ \vdots\\ \boldsymbol{\mu}_{N_1}\end{bmatrix},\ \underset{N_1N_2T\times 1}{\mathbf{V}}=\begin{bmatrix}\boldsymbol{\iota}_{N_2}\otimes\mathbf{v}_{1}\\ \vdots\\ \boldsymbol{\iota}_{N_2}\otimes\mathbf{v}_{N_1}\end{bmatrix},\ \underset{N_1N_2T\times 1}{\boldsymbol{\varepsilon}}=\begin{bmatrix}\boldsymbol{\varepsilon}_{1}\\ \vdots\\ \boldsymbol{\varepsilon}_{N_1}\end{bmatrix} \tag{22}$$

Then, it is easily seen that (see also BBMP):

$$\underset{N_1N_2T\times N_1N_2T}{Cov(\mathbf{u})}=\mathbf{I}_{N_1N_2}\otimes\left(\sigma_\mu^2\mathbf{J}_T\right)+\mathbf{I}_{N_1}\otimes\mathbf{J}_{N_2}\otimes\left(\sigma_v^2\mathbf{I}_T\right)+\mathbf{J}_{N_1}\otimes\left(\sigma_\zeta^2\mathbf{I}_{N_2T}\right)+\sigma_\varepsilon^2\mathbf{I}_{N_1N_2T} \tag{23}$$

Notice that the CTRE model imposes very limited structure of CSD, because for $i\neq i'$ and $j\neq j'$, we have:

$$E[u_{ijt}u_{ij't}]=\sigma_v^2,\ E[u_{ijt}u_{ij't}]=\sigma_\zeta^2 \text{ and } E[u_{ijt}u_{i'j't}]=0. \tag{24}$$

This suggests that the covariance between u_{ijt} and $u_{i'jt}$ is common σ_v^2 for all $i=1,...,N_1$ while the covariance between u_{ijt} and $u_{ij't}$ is common σ_ζ^2 for all $j=1,...,N_2$. Further, it imposes zero covariance between u_{ijt} and $u_{ij't'}$.

Next, we consider the 3D model with the two-way heterogeneous factor specification, (7). In this case it is straightforward to derive:

$$\underset{N_1N_2T\times 1}{\mathbf{u}}=\boldsymbol{\mu}\otimes\boldsymbol{\iota}_T+\boldsymbol{\pi}\otimes\boldsymbol{\lambda}_T+\boldsymbol{\varepsilon} \tag{25}$$

where

$$\underset{N_1N_2\times 1}{\boldsymbol{\pi}}=\begin{bmatrix}\boldsymbol{\pi}_{1}\\ \vdots\\ \boldsymbol{\pi}_{N_1}\end{bmatrix},\ \underset{N_2\times 1}{\boldsymbol{\pi}_i}=\begin{bmatrix}\pi_{i1}\\ \vdots\\ \pi_{iN_2}\end{bmatrix},\ \underset{T\times 1}{\boldsymbol{\lambda}_T}=\begin{bmatrix}\lambda_{1}\\ \vdots\\ \lambda_{T}\end{bmatrix},$$

and see (22) for other definitions. The covariance matrix for $\mathbf{u}$ in (25) is:

$$\underset{N_1N_2T\times N_1N_2T}{Cov(\mathbf{u})}=\mathbf{I}_{N_1N_2}\otimes\left(\sigma_\mu^2\mathbf{J}_T\right)+\left(\boldsymbol{\pi}\boldsymbol{\pi}'\right)\otimes\left(\sigma_\lambda^2\mathbf{I}_T\right)+\sigma_\varepsilon^2\mathbf{I}_{N_1N_2T} \tag{26}$$

Thus, the specification (25) can capture CSD by non-zero covariances between u_{ijt} and $u_{i'j't}$ for $i\neq i$ and $j\neq j$ by

$$E[u_{ijt}u_{ij't}]=\pi_{ij}\pi_{ij'}\sigma_\lambda^2,\ E[u_{ijt}u_{i'jt}]=\pi_{ij}\pi_{i'j}\sigma_\lambda^2,\ E[u_{ijt}u_{i'j't}]=\pi_{ij}\pi_{i'j'}\sigma_\lambda^2. \tag{27}$$

Next, we consider the 3D model with more general error components, (13). Combining the above results, it is straightforward to derive:

$$\underset{N_1N_2T\times 1}{\mathbf{u}}=\boldsymbol{\mu}\otimes\boldsymbol{\iota}_T+\mathbf{V}+\boldsymbol{\iota}_{N_1}\otimes\boldsymbol{\zeta}+\boldsymbol{\pi}\otimes\boldsymbol{\lambda}_T+\boldsymbol{\varepsilon}. \tag{28}$$

Thus, the covariance matrix for $\mathbf{u}$ in (28) is given by

$$\begin{aligned}\underset{N_1N_2T\times N_1N_2T}{Cov(\mathbf{u})} &= \mathbf{I}_{N_1N_2}\otimes\left(\sigma_\mu^2\mathbf{J}_T\right)+\mathbf{I}_{N_1}\otimes\mathbf{J}_{N_2}\otimes\left(\sigma_v^2\mathbf{I}_T\right) \\ &\quad +\mathbf{J}_{N_1}\otimes\left(\sigma_\zeta^2\mathbf{I}_{N_2T}\right)+\left(\boldsymbol{\pi}\boldsymbol{\pi}'\right)\otimes\left(\sigma_\lambda^2\mathbf{I}_T\right)+\sigma_\varepsilon^2\mathbf{I}_{N_1N_2T}\end{aligned} \tag{29}$$

This model can capture CSD by non-zero covariances between u_{ijt} and $u_{i'j't}$ for $i \neq i'$ and $j \neq j'$, given by

$$E[u_{ijt}u_{ij't}] = \pi_{ij}\pi_{ij'}\sigma_\lambda^2 + \sigma_v^2, E[u_{ijt}u_{i'jt}] = \pi_{ij}\pi_{i'j}\sigma_\lambda^2 + \sigma_\xi^2, E[u_{ijt}u_{i'j't}] = \pi_{ij}\pi_{i'j'}\sigma_\lambda^2. \tag{30}$$

Comparing (24), (27) and (30), we find that the CTFE specification in (2) can only accommodate non-zero covariances locally, but it also imposes the same covariance for all $i = 1, ..., N_1$ and $j = 1, ..., N_2$, respectively. Such restrictions are too strong to hold in practice.[7] On the contrary our proposed error components specification (13) can accommodate non-zero covariances both locally and globally.

Notice that v_{it} and ζ_{jt} are related to the local-time factors. In order to examine whether they exhibit weak or strong CSD, we consider the following heterogeneous local factors specifications:

$$v_{it} = v_i\tau_t \text{ and } \zeta_{jt} = \zeta_j\tau_t^*$$

where τ_t and τ_t^* are the origin and the destination-specific local common factors, respectively. Then, (13) can be replaced by

$$u_{ijt} = \mu_{ij} + v_i\tau_t + \zeta_j\tau_t^* + \varepsilon_{ijt}. \tag{31}$$

This specification implies that the exporter, i, reacts heterogeneously to the common import market conditions, τ_t and the importer, j, reacts heterogeneously to the common export market conditions, τ_t^*. Recently, Kapetanios and Shin (2017) propose a more general hierarchical multi-factor error components specification:

$$u_{ijt} = \mu_{ij} + v_i\tau_{it} + \zeta_j\tau_{jt}^* + \pi_{ij}\lambda_t + \varepsilon_{ijt}. \tag{32}$$

Within this model, we can distinguish between three types of CSD: (i) the strong global factor, λ_t which influences the (ij) pairwise interactions (of N_1N_2 dimension); (ii) the semi-strong local factors, τ_{it} and τ_{jt}^*, which influence origin or destination countries separately (each of N_1 or N_2 dimension); and (iii) the weak CSD idiosyncratic errors, ε_{ijt}. We expect that this kind of generalisation would be most natural within the 3D panel data models. Following BBMP, we assume:

$$\mu_{ij} \sim iid\left(0, \sigma_\mu^2\right), \tau_{it} \sim iid\left(0, \sigma_\tau^2\right), \tau_{jt}^* \sim iid\left(0, \sigma_{\tau^*}^2\right), \lambda_t \sim iid\left(0, \sigma_\lambda^2\right), \varepsilon_{ijt} \sim iid\left(0, \sigma_\varepsilon^2\right)$$

and μ_{ij}, τ_{it}, τ_{jt}^*, λ_t and ε_{ijt} are mutually independent.[8] It is clear that the model (32) can capture CSD by non-zero covariances between u_{ijt} and $u_{i'j't}$ for $i \neq i'$ and $j \neq j$, given by

$$E[u_{ijt}u_{ij't}] = v_i^2\sigma_\tau^2 + \pi_{ij}\pi_{ij'}\sigma_\lambda^2,\ E[u_{ijt}u_{i'jt}] = \zeta_j^2\sigma_{\tau^*}^2 + \pi_{ij}\pi_{i'j}\sigma_\lambda^2$$

[7] In the two-way error components with individual effects and time effects, the cross-section correlation is the same for all cross-section pairs. Serlenga and Shin (2007) show that such specification would produce very misleading results in the presence of heterogeneous strong CSD in the 2D panel data model.

[8] This assumption is still more general than the random effects assumptions made in BBMP.

$$E[u_{ijt}u_{i'j't}] = \pi_{ij}\pi_{i'j'}\sigma_\lambda^2. \tag{33}$$

The covariance structure in (33) is clearly more general than (30).

We now develop a diagnostic test for the null hypothesis of residual cross-section independence **and the estimation of the exponent of cross-sectional dependence** in the triple-index panel data models. Those are evaluated using the residuals obtained respectively from (5), (11) and (18), which we denote as $\mathbf{e}_{ij} = (e_{ij1}, ..., e_{ijT})'$. In particular, we have $\mathbf{e}_{ij} = \tilde{\mathbf{Y}}_{ij} - \tilde{\mathbf{X}}_{ij}\hat{\boldsymbol{\beta}}_W$ for the model (5), $\mathbf{e}_{ij} = \mathbf{M}_H\mathbf{Y}_{ij} - \mathbf{M}_H\mathbf{W}_{ij}\hat{\boldsymbol{\theta}}_{PCCE}$ for the model (11), and finally $\mathbf{e}_{ij} = \mathbf{M}_{\bar{F}}\tilde{\mathbf{Y}}_{ij} - \mathbf{M}_{\bar{F}}\tilde{\mathbf{X}}_{ij}\hat{\boldsymbol{\beta}}_{PCCE}$ for the model (18).

The proposed cross-section dependence (CD) test is a modified counterpart of an existing CD test proposed by Pesaran (2015). For convenience, we represent $\mathbf{e}_{ij}$ as the (ij) pair using the single index $n = 1, ..., N_1N_2$, and compute the pair-wise residual correlations between n and n' cross-section units by

$$\hat{\rho}_{nn'} \left(= \rho_{n'n}\right) = \frac{\mathbf{e}_n'\mathbf{e}_{n'}}{\sqrt{\left(\mathbf{e}_n'\mathbf{e}_n\right)\left(\mathbf{e}_{n'}'\mathbf{e}_{n'}\right)}},\ n, n' = 1, ..., N_1N_2 \text{ and } n \neq n'.$$

Then, we construct the CD statistic by

$$CD = \sqrt{\frac{2}{N_1N_2\left(N_1N_2 - 1\right)}} \sum_{n=1}^{N_1N_2-1} \sum_{n'=n+1}^{N_1N_2} \sqrt{T}\hat{\rho}_{nn'} \tag{34}$$

Pesaran derives that the CD test has the limiting $N(0,1)$ distribution under the null hypothesis of cross-sectional error independence, namely $H_0 : \hat{\rho}_{nn'} = 0$ for all $n, n' = 1, ..., N_1N_2$ and $n \neq n'$. Following BKP, Pesaran further shows that the CD statistic, (34) can also be applicable to testing the null hypothesis of weak cross-sectional error dependence.[9] As an extension one can also construct hierarchical CD tests based on 2D sub-dataset out of the 3D dataset.

Following BKP, we introduce the exponent of cross-sectional dependence based on the double cross sectional averages defined as $\bar{u}_{..t} = \left(N_1N_2\right)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} u_{ijt}$. If u_{ijt}'s are cross sectionally correlated across (i,j) pairs, $Var\left(\bar{u}_{..t}\right)$ declines at a rate that is a function of α, where α is defined as

$$\lim_{N_1N_2\to\infty} \left(N_1N_2\right)^{-\alpha}\lambda_{\max}\left(\boldsymbol{\Sigma}_u\right)$$

and $\boldsymbol{\Sigma}_u$ is the $N_1N_2 \times N_1N_2$ covariance matrix of $\mathbf{u}_t = \left(u_{11t}, ..., u_{N_1N_2t}\right)'$ with $\lambda_{\max}\left(\boldsymbol{\Sigma}_u\right)$ denoting the largest egeinlaue. Clearly, $Var\left(\bar{u}_{..t}\right)$ cannot decline at rate faster than $\left(N_1N_2\right)^{-1}$ as well as it cannot decline at rate slower than $\left(N_1N_2\right)^{\alpha-1}$ with $0 \leq \alpha \leq 1$. Given that

$$Var\left(\bar{u}_{..t}\right) \leq \left(N_1N_2\right)^{-1}\lambda_{\max}\left(\boldsymbol{\Sigma}_u\right),$$

[9] Consider the effects of the jth factor on the ith error, and suppose that the factor loadings take nonzero values for M_j out of the N cross-section units. The degree of cross-sectional dependence due to the jth factor can be measured by $\alpha_j = \ln(M_j)/\ln(N)$, and the overall degree of CSD by $\alpha = \max_j \alpha_j$, and $\alpha \in [0,1]$, with 1 indicating the highest CSD. Then, the implicit null is given by $H_0^w : \alpha < \frac{1}{2}$. See BKP for more details.

we find that α defined by $(N_1N_2)^{-1}\lambda_{\max}(\mathbf{\Sigma}_u) = O\left(N^{\alpha-1}\right)$ will provide an upper bound for $Var(\bar{u}_{..t})$. The extent of CSD depends on the nature of factor loadings in the following factor-based errors:

$$u_{ijt} = \pi_{ij}\lambda_t + \varepsilon_{ijt}.$$

If the average of the heterogenous loading parameters, π_{ij}, denoted μ_π, is bounded away from zero, the cross-sectional dependence will be strong, in which case, $(N_1N_2)^{-1}\lambda_{\max}(\mathbf{\Sigma}_u)$ and $Var(\bar{u}_{..t})$ are both $O(1)$, which yields $\alpha = 1$.

Furthermore, for $1/2 < \alpha \leq 1$, BKP propose the following bias-adjusted estimator to consistently estimate α:

$$\mathring{\alpha} = 1 + \frac{1}{2}\frac{\ln\left(\hat{\sigma}^2_{\bar{u}_{..t}}\right)}{\ln(N_1N_2)} - \frac{\ln\left(\hat{\mu}^2_\pi\right)}{2\ln(N_1N_2)} - \frac{\hat{c}_{N_1N_2}}{2\left[N_1N_2\ln\left(N_1N_2\right)\hat{\sigma}^2_{\bar{u}_{..t}}\right]} \quad (35)$$

where $\hat{\sigma}^2_{\bar{u}_{..t}} = T^{-1}\Sigma_{t=1}^T\bar{u}^2_{..t}$, $\hat{c}_{N_1N_2} = \widehat{\bar{\sigma}^2_{N_1N_2}} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\hat{\sigma}^2_{ij}$ and $\hat{\sigma}^2_{ij} = T^{-1}\Sigma_{t=1}^T\hat{\varepsilon}^2_{ijt}$ is the ijth diagonal element of the estimated covariance matrix, $\hat{\Sigma}_{\boldsymbol{\varepsilon}}$ with $\hat{\varepsilon}_{ijt} = u_{ijt} - \hat{\delta}_{ij}\bar{u}_{..t}$ and $\hat{\delta}_{ij}$ is the OLS estimator from the regression of u_{ijt} on $\bar{u}_{..t}$. BKP also discuss that a suitable estimation of μ^2_π can be derived, noticing that μ_π is the mean of the population regression coefficient of u_{ijt} on $\tilde{u}_{..t} = \bar{u}_{..t}/\hat{\sigma}_{\bar{u}_{..t}}$ for units u_{ijt} that have at least one non-zero loading, and those units are selected using Holm's (1979) multiple testing approach.

In the empirical section we apply the above 3D extension of the BKP estimation and testing techniques directly to the residuals e_{ijt} obtained respectively from (5), (11) and (18). We also evaluate the confidence band for the estimated CSD exponent by employing the test statistic defined in (B47) in BKP's Supplementary Appendix VI.

4 Extensions

We provide two extensions of the proposed estimation techniques into unbalanced panels and four dimensional (4D) models. Such extensions would be challenging as they involve several layers of factor specifications.

4.1 Unbalanced Panels

In practice, we may be faced with the unbalanced panel data. Notice, however, that an issue of unbalanced panels or missing data has been almost neglected even in literature on the 2D panels with unobserved factors or interactive effects. Kapetanios and Pesaran (2005) briefly deal with it in their Monte Carlo studies. Bai et al. (2015) investigate the unbalanced 2D panel data model with interactive effects, and propose the functional principal components analysis and the EM algorithm. Via simulation studies they find that the EM-type estimators are consistent for both smooth and stochastic factors, though no asymptotic analysis is provided. For the error components model (2), the 3D within transformation fails to fully eliminate the fixed effects. BMW thus extend the Wansbeek and Kapteyn (1989) approach and derive the complex within transformation, which is computationally quite demanding as it involves an inversion of $NT \times NT$ matrices. Thus, we expect that such extension of our proposed 3D PCCE estimation into unbalanced panels would be more challenging.

We now introduce a vector of selection indicators for each pair (i,j), $\mathbf{s}_{ij} = (s_{ij,1}, ..., s_{ij,T})'$, where $s_{ij,t} = 1$ if time period t for pair (i,j) can be used in estimation. We only use information on units where a full set of data are observed. Therefore, $s_{ij,t} = 1$ if and only if (x_{ijt}, y_{ijt}) is fully observed; otherwise, $s_{ij,t} = 0$. Following Wooldridge (2010), we assume that selection is ignorable conditional on $\left(\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t\right)$:

$$E\left(y_{it}|\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t, \mathbf{s}_i\right) = E\left(y_{it}|\mathbf{x}_{ijt}, \mathbf{s}_{it}, \mathbf{d}_{jt}, \mathbf{q}_t, \mathbf{z}_{ij}, \mu_{ij}, \lambda_t\right).$$

Let $n = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{t=1}^{T} s_{ij,t}$ be the total number of observations. Also define $n_t = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} s_{ij,t}$ and $n_{ij} = \sum_{t=1}^{T} s_{ij,t}$ as the number of cross-section pairs observed for time period t and the number of time periods observed for pair (i,j). Similarly, define $n_i = \sum_{j=1}^{N_2}\sum_{t=1}^{T} s_{ij,t}$, $n_j = \sum_{i=1}^{N_1}\sum_{t=1}^{T} s_{ij,t}$, $n_{it} = \sum_{j=1}^{N_2} s_{ij,t}$ and $n_{jt} = \sum_{i=1}^{N_1} s_{ij,t}$, respectively. To simplify further analysis we maintain the assumption: $\left(\min_i n_i, \min_j n_j, \min_t n_t, \min_{(ij)} n_{ij}\right) \to \infty$ or $\left(\min_t n_t, \min_{(ij)} n_{ij}\right) \to \infty$.[10]

Consider the 3D model (1) with the error components specification (7). We multiply the model by the selection indicator to get:

$$y_{ijt}^s = \boldsymbol{\beta}'\mathbf{x}_{ijt}^s + \boldsymbol{\gamma}'\mathbf{s}_{it}^s + \boldsymbol{\delta}'\mathbf{d}_{jt}^s + \boldsymbol{\kappa}'\mathbf{q}_t^s + \boldsymbol{\varphi}'\mathbf{z}_{ij}^s + \mu_{ij}^s + \pi_{ij}^s\lambda_t + \varepsilon_{ijt}^s, \tag{36}$$

where $y_{ijt} = s_{ij,t}y_{ijt}$, $\mathbf{x}_{ijt}^s = s_{ij,t}\mathbf{x}_{ijt}$, $\mathbf{s}_{it}^s = s_{ij,t}\mathbf{s}_{it}$, $\mathbf{d}_{jt}^s = s_{ij,t}\mathbf{d}_{jt}$, $\mathbf{q}_t^s = s_{ij,t}\mathbf{q}_t$, $\mathbf{z}_{ij}^s = s_{ij,t}\mathbf{z}_{ij}$, $\mu_{ij}^s = s_{ij,t}\mu_{ij}$, $\pi_{ij}^s = s_{ij,t}\pi_{ij}$, and $\varepsilon_{ijt}^s = s_{ij,t}\varepsilon_{ijt}$. Applying the cross-section averages of (36) over i and j, we obtain:

$$\bar{y}_{..t}^s = \boldsymbol{\beta}'\bar{\mathbf{x}}_{..t}^s + \boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t}^s + \boldsymbol{\delta}'\bar{\mathbf{d}}_{.t}^s + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}}_{..t}^s + \bar{\mu}_{..t}^s + \bar{\pi}_{..t}^s\lambda_t + \bar{\varepsilon}_{..t}^s \tag{37}$$

where $\bar{y}_{..t}^s = \frac{1}{n_t}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} s_{ij,t}y_{ijt} = \sum_{i=1}^{N_1} w_{it}\bar{y}_{i.t}^s$ is expressed as a weighted average with $w_{it} = n_{it}/n_t$ and $\bar{y}_{i.t}^s = n_{it}^{-1}\sum_{j=1}^{N_2} s_{ij,t}y_{ijt}$.[11] Similarly for $\bar{\mathbf{x}}_{..t}^s$, $\bar{\mathbf{z}}_{..t}^s$, $\bar{\mu}_{..t}^s$, $\bar{\pi}_{..t}^s$ and $\bar{\varepsilon}_{..t}^s$. Further, $\bar{\mathbf{s}}_{.t}^s = \sum_{i=1}^{N_1} w_{it}\mathbf{s}_{it}$, $\bar{\mathbf{d}}_{.t}^s = \sum_{j=1}^{N_2} w_{jt}\mathbf{d}_{jt}$ with $w_{jt} = n_{jt}/n_t$, and $\bar{\mathbf{q}}_t^s = \mathbf{q}_t$. As $n_t \to \infty$,

$$\bar{\mathbf{z}}_{..t}^s = \bar{\mathbf{z}} + o_p(1), \ \bar{\mu}_{..t}^s = \bar{\mu} + o_p(1), \ \bar{\pi}_{..t}^s = \bar{\pi} + o_p(1) \text{ and } \bar{\varepsilon}_{..t}^s = \bar{\varepsilon}_{..t} + o_p(1) \tag{38}$$

where $\bar{\mathbf{z}} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathbf{z}_{ij} \to_p E(\mathbf{z}_{ij})$, $\bar{\mu} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mu_{ij} \to_p 0$, $\bar{\pi} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\pi_{ij} \to_p E(\pi_{ij}) \neq 0$ and $\bar{\varepsilon}_{..t} = (N_1N_2)^{-1}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\varepsilon_{ijt} \to_p 0$. Using (38), we rewrite (37) as

$$\bar{y}_{..t}^s = \boldsymbol{\beta}'\bar{\mathbf{x}}_{..t}^s + \boldsymbol{\gamma}'\bar{\mathbf{s}}_{.t}^s + \boldsymbol{\delta}'\bar{\mathbf{d}}_{.t}^s + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}} + \bar{\mu} + \bar{\pi}\lambda_t + \bar{\varepsilon}_{..t} + o_p(1)$$

[10] For factor approximation by cross-section averages or the principal components to be valid, we still require $\min_{(ij)} n_{ij} \to \infty$, see Pesaran (2006) and Bai (2009).

[11] $\bar{y}_{..t}^s$ can be expressed as a (column sum) weighted average $\sum_{j=1}^{N_2} w_{jt}\bar{y}_{.jt}^s$ with $w_{jt} = n_{jt}/n_t$ and $\bar{y}_{.jt}^s = n_{jt}^{-1}\sum_{i=1}^{N_1} s_{ij,t}y_{ijt}$.

Hence, λ_t can be approximated by

$$\lambda_t \simeq \frac{1}{\bar{\bar{\pi}}} \left\{ \bar{y}^s_{..t} - \left(\boldsymbol{\beta}' \bar{\mathbf{x}}^s_{..t} + \boldsymbol{\gamma}' \bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}' \bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}' \mathbf{q}_t + \boldsymbol{\varphi}' \bar{\mathbf{z}} + \bar{\mu} + \bar{\varepsilon}_{..t} \right) \right\}.$$

Using these results we can augment the model (36) with the cross-section averages as follows:

$$y^s_{ijt} = \boldsymbol{\beta}' \mathbf{x}^s_{ijt} + \boldsymbol{\gamma}' \mathbf{s}^s_{it} + \boldsymbol{\delta}' \mathbf{d}^s_{jt} + \boldsymbol{\psi}'_{ij} \mathring{\mathbf{f}}^s_t + \tau^s_{ij} + \mu^s_{ij} + \varepsilon^{*s}_{ijt}, \tag{39}$$

where $\tau^s_{ij} = s_{ij,t}\tau_{ij}$, $\varepsilon^{*s}_{ijt} = s_{ij,t}\varepsilon^*_{ijt}$ and

$$\mathring{\mathbf{f}}^s_t = s_{ij,t}\mathbf{f}^s_t \text{ with } \mathbf{f}^s_t = \left(\bar{y}^s_{..t}, \bar{\mathbf{x}}^{s\prime}_{..t}, \bar{\mathbf{s}}^{s\prime}_{.t}, \bar{\mathbf{d}}^{s\prime}_{.t}, \mathbf{q}'_t \right)' \tag{40}$$

Collecting only the n_{ij} observations with $s_{ij,t} = 1$ from (39), we have:

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{S}_{ij}\boldsymbol{\gamma} + \mathbf{D}_{ij}\boldsymbol{\delta} + \mathbf{F}_{ij}\boldsymbol{\psi}_{ij} + \left(\tau_{ij} + \mu_{ij} \right) \boldsymbol{\iota}_{n_{ij}} + \mathbf{E}^*_{ij} = \mathbf{W}_{ij}\boldsymbol{\theta} + \mathbf{H}_{ij}\boldsymbol{\psi}^*_{ij} + \mathbf{E}^*_{ij} \tag{41}$$

where $\mathbf{W}_{ij} = \left(\mathbf{X}_{ij}, \mathbf{S}_{ij}, \mathbf{D}_{ij} \right)$, $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta}' & \boldsymbol{\gamma}' & \boldsymbol{\delta}' \end{pmatrix}'$, $\boldsymbol{\psi}^*_{ij} = \left(\boldsymbol{\psi}'_{ij}, \left(\tau_{ij} + \mu_{ij} \right) \right)'$, $\mathbf{H}_{ij} = \left[\mathbf{F}_{ij}, \boldsymbol{\iota}_{n_{ij}} \right]$ and

$$\underset{n_{ij}\times 1}{\mathbf{Y}_{ij}} = \begin{bmatrix} y_{ij(1)} \\ \vdots \\ y_{ij(n_{ij})} \end{bmatrix}, \underset{T_{ij}\times k_x}{\mathbf{X}_{ij}} = \begin{bmatrix} \mathbf{x}'_{ij(1)} \\ \vdots \\ \mathbf{x}'_{ij(n_{ij})} \end{bmatrix}, \underset{T_{ij}\times k_s}{\mathbf{S}_{ij}} = \begin{bmatrix} \mathbf{s}'_{i(1)} \\ \vdots \\ \mathbf{s}'_{i(n_{ij})} \end{bmatrix},$$

$$\underset{n_{ij}\times k_d}{\mathbf{D}_{ij}} = \begin{bmatrix} \mathbf{d}'_{j(1)} \\ \vdots \\ \mathbf{d}'_{j(n_{ij})} \end{bmatrix}, \underset{n_{ij}\times k_f}{\mathbf{F}_{ij}} = \begin{bmatrix} \mathbf{f}^{s\prime}_{(1)} \\ \vdots \\ \mathbf{f}^{s\prime}_{(n_{ij})} \end{bmatrix}, \underset{T_{ij}\times 1}{\mathbf{E}_{ij}} = \begin{bmatrix} \varepsilon^*_{ij(1)} \\ \vdots \\ \varepsilon^*_{ij(n_{ij})} \end{bmatrix}.$$

Here we express the time index inside (.) to highlight different initial and last time periods respectively for each cross-section pair (ij). Then, the 3D-PCCE estimator of $\boldsymbol{\theta}$ is obtained by

$$\hat{\boldsymbol{\theta}}_{PCCE} = \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{W}'_{ij} \mathbf{M}_{H_{ij}} \mathbf{W}_{ij} \right)^{-1} \left(\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbf{W}'_{ij} \mathbf{M}_{H_{ij}} \mathbf{Y}_{ij} \right) \tag{42}$$

where $\mathbf{M}_{H_{ij}} = \mathbf{I}_{T_{ij}} - \mathbf{H}_{ij} \left(\mathbf{H}'_{ij}\mathbf{H}_{ij} \right)^{-1} \mathbf{H}'_{ij}$.

Next, we consider the 3D model (1) with the general error components specification (13). To develop the two-step consistent estimation procedure for unbalanced panels, we multiply the model by $s_{ij,t} = 1$ to get:

$$y^s_{ijt} = \boldsymbol{\beta}' \mathbf{x}^s_{ijt} + \boldsymbol{\gamma}' \mathbf{s}^s_{it} + \boldsymbol{\delta}' \mathbf{d}^s_{jt} + \boldsymbol{\kappa}' \mathbf{q}^s_t + \boldsymbol{\varphi}' \mathbf{z}^s_{ij} + \mu^s_{ij} + \upsilon^s_{it} + \zeta^s_{jt} + \pi_{ij}\lambda^s_t + \varepsilon^s_{ijt} \tag{43}$$

where $y^s_{ijt} = s_{ij,t}y_{ijt}$ and similarly for others. Taking the cross-section averages of (43) over i and j, we have:

$$\bar{y}^s_{..t} = \boldsymbol{\beta}' \bar{\mathbf{x}}^s_{..t} + \boldsymbol{\gamma}' \bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}' \bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}' \mathbf{q}_t + \boldsymbol{\varphi}' \bar{\mathbf{z}}^s_{..t} + \bar{\mu}^s_{..t} + \bar{\upsilon}^s_{.t} + \bar{\zeta}^s_{.t} + \bar{\pi}^s_{..t}\lambda_t + \bar{\varepsilon}^s_{..t} \tag{44}$$

where $\bar{v}^s_{.t} = \sum_{i=1}^{N_1} w_{it} v_{it}$ with $w_{it} = n_{it}/n_t$, $\bar{\zeta}^s_{.t} = \sum_{j=1}^{N_2} w_{jt}\zeta_{jt}$ with $w_{jt} = n_{jt}/n_t$, and see (37) for other definitions. As $n_t \to \infty$,

$$\bar{v}^s_{.t} = \bar{v} + o_p(1) \text{ and } \bar{\zeta}^s_{.t} = \bar{\zeta} + o_p(1) \tag{45}$$

where $\bar{v} = N_1^{-1}\sum_{i=1}^{N_1} v_{it} \to_p 0$ and $\bar{\zeta} = N_2^{-1}\sum_{j=1}^{N_2}\zeta_{jt} \to_p 0$. Using (38) and (45), we can approximate $\bar{y}^s_{..t}$ and λ_t by

$$\bar{y}^s_{..t} = \boldsymbol{\beta}'\bar{\mathbf{x}}^s_{..t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}} + \bar{\mu} + \bar{v} + \bar{\zeta} + \bar{\pi}\lambda_t + \bar{\varepsilon}_{..t} + o_p(1)$$

$$\lambda_t = \frac{1}{\bar{\pi}}\left\{\bar{y}^s_{..t} - \left(\boldsymbol{\beta}'\bar{\mathbf{x}}^s_{.t} + \boldsymbol{\gamma}'\bar{\mathbf{s}}^s_{.t} + \boldsymbol{\delta}'\bar{\mathbf{d}}^s_{.t} + \boldsymbol{\kappa}'\mathbf{q}_t + \boldsymbol{\varphi}'\bar{\mathbf{z}} + \bar{\mu} + \bar{v} + \bar{\zeta} + \bar{\varepsilon}_{..t}\right)\right\} + o_p(1).$$

Hence, we augment the model (43) with the cross-section averages as:

$$y^s_{ijt} = \boldsymbol{\beta}'\mathbf{x}^s_{ijt} + \boldsymbol{\gamma}'\mathbf{s}^s_{it} + \boldsymbol{\delta}'\mathbf{d}^s_{jt} + \boldsymbol{\psi}'_{ij}\mathbf{f}^s_t + \tau^s_{ij} + \mu^s_{ij} + v^s_{it} + \zeta^s_{jt} + \varepsilon^{*s}_{ijt}, \tag{46}$$

where $\varepsilon^{*s}_{ijt} = s_{ij,t}\varepsilon^*_{ijt}$ with $\varepsilon^*_{ijt} = \varepsilon_{ijt} - \frac{\pi_{ij}}{\bar{\pi}}\left(\bar{\varepsilon}_{..t} + \bar{\mu} + \bar{v} + \bar{\zeta}\right) \to_p \varepsilon_{ijt}$ and see (39) for other definitions.

To derive the appropriate 3D-within transformation directly for unbalanced panels (46), we consider the simple specification:

$$y^s_{ijt} = \mu^s_{ij} + v^s_{it} + \xi^s_{jt} + \varepsilon^s_{ijt}, \tag{47}$$

and examine the property of the transformed data given by

$$\tilde{y}^s_{ijt} = y^s_{ijt} + s_{ij,t}\left(-\bar{y}^s_{ij.} - \bar{y}^s_{.jt} - \bar{y}^s_{i\cdot t} + \bar{y}^s_{..t} + \bar{y}^s_{.j.} + \bar{y}^s_{i..} - \bar{y}^s_{...}\right) \tag{48}$$

Then, it is straightforward to show:

$$\begin{aligned} & \left(-\bar{y}^s_{ij.} - \bar{y}^s_{.jt} - \bar{y}^s_{i\cdot t} + \bar{y}^s_{..t} + \bar{y}^s_{.j.} + \bar{y}^s_{i..} - \bar{y}^s_{...}\right) \\ = \; & -\left(\mu_{ij} + v_{it} + \xi_{jt}\right) + D_1 + D_2 + D_3 + D_4 + D_5 \end{aligned}$$

where

$$D_1 = -\left(\bar{v}^s_{ij.} - \sum_{j=1}^{N_2}\frac{n_{ij}}{n_i}\bar{v}^s_{ij.}\right) + \left(\sum_{i=1}^{N_1}\frac{n_{ij}}{n_j}\bar{v}^s_{ij.} - \sum_{i=1}^{N_1}\frac{n_i}{n}\sum_{j=1}^{N_2}\frac{n_{ij}}{n_i}\bar{v}^s_{ij.}\right)$$

$$D_2 = -\left(\bar{\xi}^s_{ij.} - \sum_{i=1}^{N_1}\frac{n_{ij}}{n_j}\bar{\xi}^s_{ij.}\right) + \left(\sum_{j=1}^{N_2}\frac{n_{ij}}{n_i}\bar{\xi}^s_{ij.} - \sum_{j=1}^{N_2}\frac{n_j}{n}\sum_{i=1}^{N_1}\frac{n_{ij}}{n_j}\bar{\xi}^s_{ij.}\right)$$

$$D_3 = -\left(\bar{\mu}_{.jt} - \sum_{t=1}^{T}\frac{n_{jt}}{n_j}\bar{\mu}_{.jt}\right) - \left(\bar{\mu}_{i.t} - \sum_{t=1}^{T}\frac{n_{it}}{n_i}\bar{\mu}_{i.t}\right) + \left(\bar{\mu}_{..t} - \sum_{t=1}^{T}\frac{n_t}{n}\bar{\mu}_{..t}\right)$$

$$D_4 = -\left(\bar{v}^s_{.jt} - \sum_{j=1}^{N_2}\frac{n_{jt}}{n_t}\bar{v}^s_{.jt}\right),\; D_5 = -\left(\bar{\xi}_{i.t} - \sum_{i=1}^{N_1}\frac{n_{it}}{n_t}\bar{\xi}_{i.t}\right)$$

with $\bar{v}^s_{ij.} = \frac{1}{n_{ij}}\sum_{t=1}^T s_{ij,t}v_{it}$, $\bar{\xi}^s_{ij.} = \frac{1}{n_{ij}}\sum_{t=1}^T s_{ij,t}\xi_{jt}$, $\bar{\mu}^s_{.jt} = \frac{1}{n_{jt}}\sum_{i=1}^{N_1} s_{ij,t}\mu_{ij}$, $\bar{\mu}^s_{i.t} = \frac{1}{n_{it}}\sum_{j=1}^{N_2} s_{ij,t}\mu_{ij}$, $\bar{\mu}^s_{..t} = \frac{1}{n_t}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} s_{ij,t}\mu_{ij}$, $\bar{v}^s_{.jt} = \frac{1}{n_{jt}}\sum_{i=1}^{N_1} s_{ij,t}v_{it}$, and $\bar{\xi}^s_{i.t} = \frac{1}{n_{it}}\sum_{j=1}^{N_2}\xi_{jt}$. Notice in the balanced panels that $D_1 = D_2 = D_3 = D_4 = D_5 = 0$. Also, as $\left(\min_i n_i, \min_j n_j, \min_t n_t, \min_{(ij)} n_{ij}\right) \to \infty$, $D_i \to_p 0$ for $i = 1, ..., 5$. Therefore, we obtain:

$$\left(-\bar{y}^s_{ij.} - \bar{y}^s_{.jt} - \bar{y}^s_{i\cdot t} + \bar{y}^s_{\cdot\cdot t} + \bar{y}^s_{.j.} + \bar{y}^s_{i\cdot\cdot} - \bar{y}^s_{\cdots}\right) = -\left(\mu_{ij} + v_{it} + \xi_{jt}\right) + o_p(1) \quad (49)$$

Using (49) and applying (48) to (47), we obtain the desired result:

$$\tilde{y}^s_{ijt} = \tilde{\varepsilon}^s_{ijt}, \quad (50)$$

where $\tilde{\varepsilon}^s_{ijt} = \varepsilon^s_{ijt} - s_{ij,t}\left(\bar{\varepsilon}^s_{ij.} - \bar{\varepsilon}^s_{.jt} - \bar{\varepsilon}^s_{i\cdot t} + \bar{\varepsilon}^s_{\cdot\cdot t} + \bar{\varepsilon}^s_{.j.} + \bar{\varepsilon}^s_{i\cdot\cdot} - \bar{\varepsilon}^s_{\cdots}\right)$.

We now apply the 3D-within transformation (48) to (46) and obtain:

$$\tilde{y}^s_{ijt} = \boldsymbol{\beta}'\tilde{\mathbf{x}}^s_{ijt} + \tilde{\boldsymbol{\psi}}'_{ij}\mathring{\mathbf{f}}^s_{ijt} + \tilde{\varepsilon}^{*s}_{ijt}, \quad (51)$$

where $\tilde{\boldsymbol{\psi}}_{ij} = \boldsymbol{\psi}'_{ij} - \left(\frac{1}{n_{jt}}\sum_{i=1}^{N_1}\boldsymbol{\psi}'_{ij}\right) - \left(\frac{1}{n_{it}}\sum_{j=1}^{N_2}\boldsymbol{\psi}'_{ij}\right) + \left(\frac{1}{n_t}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\boldsymbol{\psi}'_{ij}\right)$, $\mathring{\mathbf{f}}^s_{ijt} = s_{ij,t}\tilde{\mathbf{f}}^s_{ij}$ with $\tilde{\mathbf{f}}^s_{ij} = \mathbf{f}^s_t - \bar{\mathbf{f}}^s_{ij}$ and $\bar{\mathbf{f}}^s_{ij} = n_{ij}^{-1}\sum_{t=1}^T s_{ij,t}\mathbf{f}^s_t$, and $\mathbf{f}^s_t$ is defined in (40). Collecting only the n_{ij} observations with $s_{ij,t} = 1$ from (51), we have:

$$\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{X}}_{ij}\boldsymbol{\beta} + \tilde{\mathbf{F}}_{ij}\tilde{\boldsymbol{\psi}}_{ij} + \tilde{\mathbf{E}}^*_{ij}, \quad (52)$$

where

$$\underset{n_{ij}\times 1}{\tilde{\mathbf{Y}}_{ij}} = \begin{bmatrix} \tilde{y}_{ij(1)} \\ \vdots \\ \tilde{y}_{ij(n_{ij})} \end{bmatrix}, \underset{n_{ij}\times k_x}{\tilde{\mathbf{X}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ij(1)} \\ \vdots \\ \tilde{\mathbf{x}}'_{ij(n_{ij})} \end{bmatrix}, \underset{n_{ij}\times k_f}{\tilde{\mathbf{F}}_{ij}} = \begin{bmatrix} \tilde{\mathbf{f}}^{s\prime}_{ij(1)} \\ \vdots \\ \tilde{\mathbf{f}}^{s\prime}_{ij(n_{ij})} \end{bmatrix}, \underset{n_{ij}\times 1}{\tilde{\mathbf{E}}^*_{ij}} = \begin{bmatrix} \tilde{\varepsilon}^*_{ij(1)} \\ \vdots \\ \tilde{\varepsilon}^*_{ij(n_{ij})} \end{bmatrix}.$$

As before we employ the time index inside (.) to highlight different initial and last time periods respectively for each cross-section pair (ij). Then, the 3D-PCCE estimators of $\boldsymbol{\beta}$ are obtained by

$$\tilde{\boldsymbol{\beta}}_{PCCE} = \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\tilde{\mathbf{X}}'_{ij}\mathbf{M}_{ij}\tilde{\mathbf{X}}_{ij}\right)^{-1}\left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\tilde{\mathbf{X}}'_{ij}\mathbf{M}_{ij}\tilde{\mathbf{Y}}_{ij}\right) \quad (53)$$

where $\mathbf{M}_{ij} = \mathbf{I}_T - \tilde{\mathbf{F}}_{ij}\left(\tilde{\mathbf{F}}'_{ij}\tilde{\mathbf{F}}_{ij}\right)^{-1}\tilde{\mathbf{F}}'_{ij}$. As $\left(\min_t n_t, \min_{(ij)} n_{(ij)}\right) \to \infty$, both PCCE estimators, (42) and (53), will follow the asymptotic normal distribution.

4.2 4D Model Extensions

BMW propose the following baseline 4D model:

$$y_{ijst} = \mathbf{x}'_{ijst}\boldsymbol{\beta} + u_{ijst}, \quad (54)$$

$$u_{ijst}=\mu_{ijs}+\theta_{ijt}+\zeta_{jst}+v_{ist}+\varepsilon_{ijst} \tag{55}$$

for $i=1,...,N_1$, $j=1,...,N_2$, $s=1,...,N_3$, and $t=1,...,T$. BMW derive the following 4D within transformation to eliminate pair-wise interaction effects, μ_{ijs}, v_{ist}, ζ_{jst}, and λ_{ijt} from (54):

$$\begin{aligned}\tilde{y}_{ijst} &= y_{ijst}-\bar{y}_{.jst}-\bar{y}_{i.st}-\bar{y}_{ij.t}-\bar{y}_{ijs.}+\bar{y}_{..st}+\bar{y}_{.j.t}+\bar{y}_{.js.}+\bar{y}_{i..t}+\bar{y}_{i.s.}+\bar{y}_{ij..}\\ &\quad -\bar{y}_{...t}-\bar{y}_{..s.}-\bar{y}_{.j..}-\bar{y}_{i...}-\bar{y}_{....}\end{aligned} \tag{56}$$

and estimate $\boldsymbol{\beta}$ consistently from the transformed specification:

$$\tilde{y}_{ijst}=\tilde{\mathbf{x}}'_{ijst}\boldsymbol{\beta}+u_{ijst}. \tag{57}$$

Further, BBMP propose the feasible GLS random effects estimator of $\boldsymbol{\beta}$ under the assumption that u_{ijst} and its components individually have zero mean, and error components are pairwise uncorrelated.

To introduce CSD explicitly into (54), we consider the following extension:

$$y_{ijst}=\mathbf{x}'_{ijst}\boldsymbol{\beta}+\mu_{ijs}+\theta_{ijt}+\zeta_{jst}+v_{ist}+\pi_{ijs}\lambda_t+\varepsilon_{ijst}, \tag{58}$$

where λ_t is the global factor with heterogeneous coefficients π_{ijs}. In the presence of such CSD, both the 4D-FE estimator and the 4D-RE estimator would be likely to be biased due to the correlation between $\mathbf{x}_{ijst}$ and λ_t. Thus, we develop the two-step consistent estimation procedure. Taking the cross-section averages of (58) over i, j and s, we obtain:

$$\begin{aligned}\bar{y}_{...t} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\left(\boldsymbol{\beta}'\mathbf{x}_{ijst}+\mu_{ijs}+\theta_{ijt}+\zeta_{jst}+v_{ist}+\pi_{ijs}\lambda_t+\varepsilon_{ijst}\right)\\ &= \boldsymbol{\beta}'\bar{\mathbf{x}}_{...t}+\bar{\mu}_{...}+\bar{\theta}_{..t}+\bar{\zeta}_{..t}+\bar{v}_{..t}+\bar{\pi}_{...}\lambda_t+\bar{\varepsilon}_{...t}\end{aligned} \tag{59}$$

where $\bar{\mathbf{x}}_{...t}=\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\mathbf{x}_{ijst}$, $\bar{\varepsilon}_{...t}=\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\varepsilon_{ijst}$, $\bar{\mu}_{...}=\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\mu_{ijs}$, $\bar{\pi}_{...}=\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\pi_{ijs}$, $\bar{\theta}_{..t}=\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N_2}\sum_{j=1}^{N_2}\theta_{ijt}$, $\bar{\zeta}_{..t}=\frac{1}{N_2}\sum_{j=1}^{N_2}\frac{1}{N}\sum_{s=1}^{N_3}\zeta_{jst}$, and $\bar{v}_{..t}=\frac{1}{N_1}\sum_{i=1}^{N_1}\frac{1}{N}\sum_{s=1}^{N_3}v_{ist}$. From (59) we have:

$$\lambda_t=\frac{1}{\pi_{...}}\left\{\bar{y}_{...t}-\left(\boldsymbol{\beta}'\bar{\mathbf{x}}_{...t}+\bar{\mu}_{...}+\bar{\theta}_{..t}+\bar{\zeta}_{..t}+\bar{v}_{..t}+\bar{\pi}_{...}\lambda_t+\bar{\varepsilon}_{...t}\right)\right\}.$$

Thus, we derive the cross-section augmented version of (58) by

$$y_{ijst}=\boldsymbol{\beta}'\mathbf{x}_{ijst}+\boldsymbol{\psi}'_{ijs}\mathbf{f}_t+\mu_{ijs}+\theta_{ijt}+\zeta_{jst}+v_{ist}+\varepsilon^*_{ijst}, \tag{60}$$

where $\mathbf{f}_t=(\bar{y}_{...t},\bar{\mathbf{x}}'_{...t})'$, $\boldsymbol{\psi}'_{ijs}=\left(\psi_{0,ijs},\boldsymbol{\psi}'_{ijs}\right)=\left(\frac{\pi_{ijs}}{\bar{\pi}_{...}},-\frac{\pi_{ijs}}{\bar{\pi}_{...}}\boldsymbol{\beta}'\right)$ and $\varepsilon^*_{ijst}=\varepsilon_{ijst}-\frac{\pi_{ijs}}{\bar{\pi}_{...}}\left(\bar{\varepsilon}_{...t}+\bar{\theta}_{..t}+\bar{\zeta}_{..t}+\bar{v}_{..t}\right)$. As $N_1,N_2,N_3\to\infty$, $\varepsilon^*_{ijst}\to_p\varepsilon_{ijst}$ because of the following approximations: $\bar{\mu}_{...}\to_p 0$, $\theta_{..t}\to_p 0$, $\zeta_{..t}\to_p 0$, $v_{..t}\to_p 0$, and $\bar{\varepsilon}_{...t}\to_p 0$.

Next, we apply the 4D-within transformation (56) to (60), and obtain:

$$\tilde{y}_{ijst} = \boldsymbol{\beta}'\tilde{\mathbf{x}}_{ijst} + \tilde{\boldsymbol{\psi}}'_{ijs}\tilde{\mathbf{f}}_t + \tilde{\varepsilon}^*_{ijt}. \tag{61}$$

where

$$\tilde{\boldsymbol{\psi}}'_{ijs} = \left(\boldsymbol{\psi}_{ijs} - \boldsymbol{\psi}_{.js} - \boldsymbol{\psi}_{i.s} - \boldsymbol{\psi}_{ij.} + \boldsymbol{\psi}_{..s} + \boldsymbol{\psi}_{.j.} + \boldsymbol{\psi}_{i..} - \boldsymbol{\psi}_{...}\right)' \text{ and } \tilde{\mathbf{f}}_t = \left(\mathbf{f}_t - \bar{\mathbf{f}}\right).$$

We rewrite (61) as

$$\tilde{\mathbf{Y}}_{ijs} = \tilde{\mathbf{X}}_{ijs}\boldsymbol{\beta} + \tilde{\mathbf{F}}\tilde{\boldsymbol{\psi}}_{ijs} + \tilde{\boldsymbol{\varepsilon}}^*_{ijs}, \tag{62}$$

where

$$\underset{T\times 1}{\tilde{\mathbf{Y}}_{ijs}} = \begin{bmatrix} \tilde{y}_{ijs1} \\ \vdots \\ \tilde{y}_{ijsT} \end{bmatrix}, \ \underset{T\times k_x}{\tilde{\mathbf{X}}_{ijs}} = \begin{bmatrix} \tilde{\mathbf{x}}'_{ijs1} \\ \vdots \\ \tilde{\mathbf{x}}'_{ijsT} \end{bmatrix}, \ \underset{T\times k_f}{\tilde{\mathbf{F}}} = \begin{bmatrix} \tilde{\mathbf{f}}'_1 \\ \vdots \\ \tilde{\mathbf{f}}'_T \end{bmatrix}.$$

Then, it is straightforward to derive the PCCE estimator of $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}}_{PCCE} = \left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{s=1}^{N_3}\tilde{\mathbf{X}}'_{ijs}\mathbf{M}_{\tilde{\mathbf{F}}}\tilde{\mathbf{X}}_{ijs}\right)^{-1}\left(\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{s=1}^{N_3}\tilde{\mathbf{X}}'_{ijs}\mathbf{M}_{\tilde{\mathbf{F}}}\tilde{\mathbf{Y}}_{ijs}\right), \tag{63}$$

where $\mathbf{M}_{\tilde{\mathbf{F}}} = \mathbf{I}_T - \tilde{\mathbf{F}}\left(\tilde{\mathbf{F}}'\tilde{\mathbf{F}}\right)^{-1}\tilde{\mathbf{F}}'$.

Further, we can follow Kapetanios and Shin (2017) and develop 4D models with the hierarchical multi-factor error structure. To this end define the global factor λ_t which affects all (ijs) pairs, the regional factors τ_{it}, τ^*_{jt}, τ^{**}_{st}, and finally the local factors τ_{ijt}, τ^*_{ist} and τ^{**}_{jst}. This logic suggests the following model:

$$y_{ijst} = \mathbf{x}'_{ijst}\boldsymbol{\beta} + \mu_{ijs} + v_{js}\tau_{it} + v^*_{is}\tau_{jt} + v^{**}_{ij}\tau_{st} + \zeta_s\tau_{ijt} + \zeta^*_j\tau^*_{ist} + \zeta^{**}_i\tau^{**}_{jst} + \pi_{ij}\lambda_t + \varepsilon_{ijst}. \tag{64}$$

Such higher-dimensional setups involve several layers of factor specifications (a number that grows with the dimensions), rendering the estimation of and inference on models non-trivial and challenging.

5 Monte Carlo Study

In this section we conduct the Monte Carlo studies and investigate the small sample properties of the CTFE and two versions of the PCCE estimator for the models, (4), (11) and (17), respectively. We consider the two data generating processes (DGP). We construct DGP1 by

$$y_{ijt} = \beta' x_{ijt} + \mu_{ij} + \pi_{ij}\lambda_t + \varepsilon_{ijt}, \tag{65}$$

$$x_{ijt} = \mu^x_{ij} + \mu_{ij} + \pi^x_{ij}\lambda_t + v_{ijt}, \tag{66}$$

for $i = 1, ..., N_1$, $j = 1, ..., N_2$, and $t = 1, ..., T$. The global common factor, λ_t and idiosyncratic errors, ε_{ijt} and v_{ijt} are generated independently as *iid* processes with zero mean and unit variance: namely, $\lambda_t \sim iidN(0,1)$,

$\varepsilon_{ijt} \sim iidN(0,1)$ and $v_{ijt} \sim iidN(0,1)$. We generate pairwise individual effects independently as $\mu_{ij} \sim iidN(0,1)$ and $\mu_{ij}^x \sim iidN(0,1)$. The factor loadings, π_{ij} and π_{ij}^x, are then independently generated from $U[1,2]$ random variables.

Next, we construct DGP2 by

$$y_{ijt} = \beta' x_{ijt} + \mu_{ij} + \upsilon_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt}. \tag{67}$$

$$x_{ijt} = \mu_{ij}^x + \mu_{ij} + \pi_{ij}^x \lambda_t + v_{ijt},, \tag{68}$$

for $i = 1, ..., N_1$, $j = 1, ..., N_2$, and $t = 1, ..., T$. In addition, we follow BBMP and generate υ_{it} and ζ_{jt} independently as:

$$\upsilon_{it} \sim U(-1,1) \text{ and } \zeta_{jt} \sim U(-1,1) \text{ for } i = 1, ..., N_1, \ j = 1, ..., N_2, \ t = 1, ...T.$$

In both DGP1 and DGP2 we set $\beta = 1$.

For each experiment we evaluate the following summary statistics:
Bias: $\hat{\beta}_R - \beta_0$, where $\beta_0 (= 1)$ is the true parameter value, and $\hat{\beta}_R = R^{-1}\sum_{r=1}^R \hat{\beta}_r$ is the mean of $\hat{\beta}_r$.
Root mean square error (RMSE): $\sqrt{R^{-1}\sum_{r=1}^R \left(\hat{\beta}_r - \beta_0\right)^2}$.
Size: the empirical rejection probability of the t-statistic for the null hypothesis $\beta = \beta_0$ against $\beta \neq \beta_0$ at the 5% significance.

We conduct each experiment 1,000 times for the (N_1, N_2, T) triples with $N_1, N_2 = \{25, 49, 100\}$[12] and $T = \{50, 100, 200, 400\}$. The simulation results are provided respectively in Table 1 for DGP1 and Table 2 for DGP2.

In Table 1 we find that biases of the 2D and 3D PCCE estimators of β are mostly negligible even for the relatively small sample size at $(N_1, N_2, T) = (25, 25, 50)$. On the other hand, the CTFE estimator displays substantial biases for most cases. As N increases, the biases become smaller but still non-negligible. RMSE results are qualitatively similar to the bias patterns. RMSEs of both PCCE estimators decrease sharply with $N_1 (N_2)$ or T whilst RMSEs of the CTFE estimator fall only with $N_1 (N_2)$. Turning to the empirical sizes of the t-test for the null hypothesis, $\hat{\beta} = \beta_0 (= 1)$, we find that the CTFE over-rejects the null in all cases and tends to 1 as $N_1 (N_2)$ or T rises. By contrast the size of the 2D PCCE estimator is reasonably close to the nominal 5% level in all cases while the 3D PCCE estimator tends to slightly over-reject the null when N_1 or N_2 is relatively small. As expected, overall performance of the 2D PCCE estimator is the best under DGP1.

Simulation results in Table 2 are qualitatively similar to those in Table 1. Biases of both PCCE estimators are almost negligible in all cases and their RMSEs decrease rapidly with $N_1 (N_2)$ or T. Empirical sizes of the t-test for $\hat{\beta} = \beta_0 (= 1)$ are still close to the nominal 5% level in almost all cases for the 2D PCCE estimator. The 3D PCCE estimator tends to slightly over-reject the null, but its size performance improves quickly as $N_1 (N_2)$ or T increases. On the

[12] Namely, 25 are pairs of 5 units, 49 pairs of 7 units and 100 pairs of 10 units.

Table 1: Simulation results for β under the DGP1

	CTFE							
	Bias							
$(N_1 N_2, T)$	50	100	200	400				
25	0.0829	0.0832	0.0833	0.0822				
49	0.0347	0.0341	0.0338	0.0344				
100	-0.0307	-0.0315	-0.0313	-0.0316				
	RMSE							
$(N_1 N_2, T)$	50	100	200	400				
25	0.0914	0.0871	0.0854	0.0832				
49	0.0420	0.0383	0.0357	0.0353				
100	0.0347	0.0336	0.0324	0.0322				
	Size							
$(N_1 N_2, T)$	50	100	200	400				
25	0.7610	0.9590	0.9980	1.0000				
49	0.4020	0.6290	0.8810	0.9950				
100	0.5530	0.8410	0.9910	1.0000				
	2D PCCE				3D PCCE			
	Bias							
$(N_1 N_2, T)$	50	100	200	400	50	100	200	400
25	0.0017	0.0011	0.0014	0.0003	0.0017	0.0008	0.0012	0.0002
49	0.0006	-0.0005	0.0002	0.0004	0.0000	-0.0001	0.0000	0.0006
100	0.0004	-0.0005	-0.0001	-0.0003	0.0009	-0.0002	0.0000	-0.0003
	RMSE							
$(N_1 N_2, T)$	50	100	200	400	50	100	200	400
25	0.0290	0.0202	0.0146	0.0101	0.0290	0.0202	0.0146	0.0101
49	0.0207	0.0156	0.0100	0.0071	0.0207	0.0156	0.0100	0.0071
100	0.0142	0.0103	0.0070	0.0051	0.0142	0.0103	0.0070	0.0051
	Size							
$(N_1 N_2, T)$	50	100	200	400	50	100	200	400
25	0.049	0.045	0.052	0.052	0.133	0.124	0.132	0.114
49	0.041	0.062	0.047	0.047	0.095	0.113	0.097	0.086
100	0.042	0.048	0.044	0.055	0.081	0.093	0.074	0.093

Notes: We report the simulation results for three estimators for DGP1, (65) and (66). CTFE refers to the 3D within estimator given by (5), 2D PCCE is the PCCE estimator given by (11) and 3D PCCE is the PCCE estimator given by (18).

contrary the CTFE estimator suffers from substantial biases and size distortions, and its performance does not improve in large samples. We note in passing that such good performance of the 2D PCCE estimator is rather surprising as we expect that the 3D PCCE estimator will dominate under DGP2. Overall simulation results support the simulation findings reported under the 2D panels by Kapetanios and Pesaran (2005) and Pesaran (2006).

Table 2: Simulation results for β under the DGP2

	CTFE							
	Bias							
(N_1N_2, T)	50	100	200	400				
25	0.0835	0.0829	0.0830	0.0827				
49	0.0143	0.0144	0.0155	0.0156				
100	-0.0365	-0.0371	-0.0362	-0.0370				
	RMSE							
(N_1N_2, T)	50	100	200	400				
25	0.0921	0.0872	0.0850	0.0839				
49	0.0272	0.0220	0.0194	0.0177				
100	0.0400	0.0388	0.0371	0.0374				
	Size							
(N_1N_2, T)	50	100	200	400				
25	0.7780	0.9450	1.0000	1.0000				
49	0.1420	0.2060	0.3630	0.5650				
100	0.7120	0.9400	0.9940	1.0000				
	2D PCCE				3D PCCE			
	Bias							
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	-0.0001	0.0008	0.0009	0.0001	0.0012	0.0006	0.0009	0.0005
49	-0.0002	0.0000	0.0006	0.0001	-0.0001	0.0000	0.0005	0.0005
100	0.0000	-0.0001	0.0001	-0.0002	0.0001	-0.0003	0.0001	-0.0002
	RMSE							
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	0.0295	0.0201	0.0145	0.0104	0.0368	0.0250	0.0181	0.0130
49	0.0208	0.0147	0.0102	0.0072	0.0238	0.0169	0.0120	0.0083
100	0.0148	0.0103	0.0069	0.0049	0.0161	0.0114	0.0076	0.0054
	Size							
(N_1N_2, T)	50	100	200	400	50	100	200	400
25	0.0510	0.0470	0.0630	0.0570	0.1290	0.1240	0.1330	0.1260
49	0.0480	0.0540	0.0460	0.0500	0.0910	0.1010	0.1070	0.0850
100	0.0620	0.0510	0.0450	0.0490	0.0930	0.0800	0.0660	0.0700

Notes: We report the simulation results for three estimators for DGP2, (67) and (68). CTFE refers to the 3D within estimator given by (5), 2D PCCE is the PCCE estimator given by (11) and 3D PCCE is the PCCE estimator given by (18).

6 Empirical Application to the Gravity Model of the Intra-EU Trade

Anderson and van Wincoop (2003) show that "The gravity equation tells us that bilateral trade, after controlling for size, depends on the bilateral trade barriers but relative to the product of their Multilateral Resistance Indices (MTR)." Bilateral barrier relative to average trade barriers that both regions face with all their trading partners. Omitting MTR induces potentially severe bias (e.g. Baldwin and Taglioni, 2006). Subsequent research has focused on estimating the model with directional country-specific fixed effects to control for unobservable MTRs (e.g. Feenstra, 2004).

A large number of studies have established an importance of taking into account multilateral resistance and bilateral heterogeneity, simultaneously, in the 2D panels. Serlenga and Shin (2007) is the first paper to develop the panel gravity model by incorporating observed and unobserved factors. Alternatively, Behrens et al. (2012) develop the spatial econometric specification, to control for multilateral cross-sectional correlations across trade flows. Mastromarco et al. (2016b) compare the factor- and the spatial-based gravity models to investigate the Euro impact on intra-EU trade flows over 1960-2008 for 190 country-pairs of 14 EU and 6 non-EU OECD countries. They document evidence that the CD test confirms that the factor-based model is more appropriate. Furthermore, Gunnella et al. (2015) propose the panel gravity models which accommodate both strong and weak CSD, simultaneously, through the use of unobserved factors and endogenously selected spatial clusters estimated by the nonlinear threshold techniques advanced by Kapetanios et al. (2014).

When we analyse the 3D panel gravity models, we should control for the potential origin of biases presented by unobserved time-varying MTRs, if they are correlated with covariates. Baltagi et al. (2003) propose the 3D panel data model with CTFE specification. This approach has been popularly applied to measure the impacts of (unobserved) MTRs of the exporters and the importers in the structural gravity studies (e.g. Baltagi et al., 2015 and Chapter 10). As discussed in Section 3, however, the 3D panel data model, typically estimated by CTFE or CTRE estimators, fails to accommodate (strong and heterogeneous) CSD. The presence of cross-sectional correlations across (ij) pairs suggests that the appropriate econometric techniques be required, in order to avoid biased and misleading estimation results.

In this regard we apply our proposed approach to the dataset covering the period 1960-2008 (49 years) for 182 country-pairs amongst 14 EU member countries (Austria, Belgium-Luxemburg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden, United Kingdom).[13] Our sample period consists of several important economic integrations, such as the Custom Union in 1958, the European Monetary System in

[13] It is the extended dataset analysed by Serlenga and Shin (2007), who provide the Data Appendix for detailed data description and sources. Belgium and Luxemburg are treated as a single country. Denmark, Sweden and The UK although nonmember countries, as part of the EU, experienced similar history and faced similar legislation and regulation to euro area

1979 and the Single Market in 1993, all of which can be regarded as promoting intra-EU trades.[14]

We consider the following generalised panel gravity specification:

$$\ln EXP_{ijt} = \beta_0 + \beta_1 CEE_{ijt} + \beta_2 EMU_{ijt} + \beta_3 SIM_{ijt} + \beta_4 RLF_{ijt} + \beta_5 \ln GDP_{it} \\ + \beta_6 \ln GDP_{jt} + \beta_7 RER_t + \gamma_1 DIS_{ij} + \gamma_2 BOR_{ij} + \gamma_3 LAN_{ij} + u_{ijt} \tag{69}$$

where the dependent variable, EXP_{ijt} is the export flow from country i to country j at time t, CEE and EMU are dummies for European Community membership and for European Monetary Union, SIM and RLF measure similarity in size and difference in relative factor endowments, RER represents the logarithm of common real exchange rates, GDP_{it} and GDP_{jt} are logged GDPs of exporter and importer, and finally the logarithm of geographical distance (DIS) and the dummies for common language (LAN) and for common border (BOR) represent time-invariant bilateral barriers.

We report the estimation results of (69), employing the four estimators, namely the two-way within estimator with $u_{ijt} = \mu_{ij} + \lambda_t + \varepsilon_{ijt}$, the CTFE estimator with $u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \varepsilon_{ijt}$, the 2D PCCE estimator with $u_{ijt} = \mu_{ij} + \pi_{ij}\lambda_t + \varepsilon_{ijt}$, and the 3D PCCE estimator with $u_{ijt} = \mu_{ij} + v_{it} + \zeta_{jt} + \pi_{ij}\lambda_t + \varepsilon_{ijt}$. We also report the CD test results applied to the residuals from each of the four estimation methods and the estimates of the CSD exponent (α). Following BKP, we compute the CD test and estimate α sequentially.

Following the structural gravity literature, our main focus is on investigating the impacts of t_{ij} that contain both barriers and incentives to trade between i and j.[15] Here we focus on the two dummy variables, CEE (equal to one when both countries belong to the European Community) and EMU (equal to one when both trading partners adopt the same currency). Both are expected to exert a positive impact on bilateral export flows. The main motivation behind the EMU project is that a single currency will reduce the transaction costs within member countries. But, the empirical evidence on the common currency effect on trade flows is rather mixed. Rose (2001), Frankel and Rose (2002), Glick and Rose (2002) and Frankel (2008), document a huge positive effect whilst a number of studies report negative or insignificant effects (e.g., Persson, 2001, Pakko and Wall, 2002, De Nardis and Vicarelli, 2003). More recent studies by Serlenga and Shin (2007), and Mastromarco et al. (2016b), and Gunnella et al. (2015) highlight an importance of controlling for strong CSD, and find a small but significant effect (7 to 10%) of the euro on intra-EU trade flows.[16]

countries.

[14] To mitigate the potentially negative impact of the global financial crisis on our analysis, we exclude the data after 2008

[15] In the current study we cannot consistently estimate the coefficients associated with time-invariant regressors, DIS_{ij}, BOR_{ij}, and LAN_{ij} and/or because the within transformation wipes out them. Similar identification issues are also applied to the coefficients on GDP_{it} and GDP_{jt}. Following Serlenga and Shin (2007) and Chapter 3 we will investigate this issue in future study.

[16] After the Brexit, the issue on potential benefits of joining the currency union will be

In addition to the standard mass covariates, GDP_{it} and GDP_{jt}, we also consider the impact of (logarithm of) bilateral real exchange rates (RER), which is defined as the price of the foreign currency per the home currency unit and is meant to capture the relative price effects. Further, following the New Trade Theory advanced by Krugman (1979) and Helpman (1987), we add RLF and SIM. RLF is the logarithm of the absolute value of the difference between per capita GDPs of trading countries, and measures the difference in terms of relative factor endowments. The higher RLF results in a higher volume of inter-industry trade and a lower share of intra-industry trade. SIM is the logarithm of an index that captures the relative size of two countries in terms of GDP, and is bounded between zero (absolute divergence in size) and 0.5 (equal country size).

Table 3 reports the estimation results for the 3D panel gravity specification in (69). The two-way FE estimation results are all statistically significant except RER. The impacts of home and foreign country GDPs on exports are positive, but surprisingly, the former is twice larger than the latter. The impact of similarity in size (SIM) is negative and significant, inconsistent with *a priori* expectations. Importantly, we find that trade and currency union memberships (CEE and EMU) significantly boost export flows, though their magnitudes seem to be too high. However, the CD test applied to the residuals rejects the null of no or weak CSD convincingly. The estimate of α is 0.99 with the confidence band containing unity, suggesting that the residuals are strongly correlated. Thus, the FE estimation results are likely to be biased and unreliable.

Next, we turn to the CTFE estimation results, which has been popularly applied in the structural gravity literature. This approach aims to control for MTRs through bilateral pair-fixed effects and origin and destination country-time fixed effects. The CD test results indicate that the CTFE residuals do not suffer from any strong CSD, suggesting that the 3D within transformation may be able to remove strong CSD. This rather surprising result is not supported by the estimate of α, which is 0.91. Though the confidence band does not include unity, this estimate is still pretty high and close to 1. Further, we find that all the coefficients become insignificant except for CEE. Focusing on the impacts on CEE and EMU, the former is still substantial (0.29) while the latter turns out to be negligible (-0.011). Combining these results together, we may conclude that the CTFE results are rather unreliable.

Moving to the 2D PCCE estimation results, we find that all the coefficients are significant with the expected signs except for EMU. The impact of foreign country GDP on exports is substantially larger than home GDP. The RER coefficient is positive, confirming that a depreciation of the home currency increases exports. The impact of CEE is smaller (0.186), but the EMU effect is insignificant and negligible (0.017). But, the PCCE estimator still suffers from strong CSD residuals together with the estimate of α being 0.87. This may explain the conflict finding against the existing studies reporting a significant effect of the

more hotly debated. In retrospect, the UK Treasury made a bold prediction in 2003 that the pro-trade effect of the Euro on UK would be over 40%.

euro on trades in the 2D panels.

Finally, the 3D PCCE estimation results show that all the coefficients are significant with expected signs. The CD test fails to strongly reject the null, suggesting that this approach is able to successfully deal with strong and/or weak CSD in the 3D panels. This is also supported by the smaller estimate of α (0.77), which is close to a moderate range of weak CSD.[17] Focusing on the CEE and EMU impacts on exports, we find that the former turns out to be still substantial (0.335) while the latter becomes modest at 0.081, close to the consensus magnitudes reported in the recent 2D panel studies (e.g. Baldwin, 2006, Gunnella et al., 2015). Combining these results, we may conclude that the 3D PCCE estimation results are mostly reliable, providing a general support for the thesis that the potential trade-creating effects of the Euro should be viewed in the long-run historical and multilateral perspectives rather than simply focusing on the formation of a monetary union as an isolated event.

The CTFE estimator is proposed to capture the (unobserved) multilateral resistance terms and trade costs, which are likely to exhibit history and time dependence (e.g. Herwartz and Weber, 2010). However, it fails to accommodate strong cross-section correlations among MTRs, which are present in our sample of the EU countries (confirmed by CD tests and CSD exponent estimates). To capture such complex interlinkages among trading partners, we should model the time-varying interdependency of bilateral export flows in a more flexible manner than simply introducing deterministic country-time specific dummies. Baldwin (2006) stresses an importance of taking into account the fact that many omitted pair-specific variables reflect time-varying factors such as multilateral trade costs or union membership. MTRs arise from the bilateral country-pair specific reactions to global shocks or the local spillover effects across a small number of countries or both. In order to avoid biased and misleading results, we propose a novel econometric technique, called the 3D-PCCE estimator, which is the first step to developing the multidimensional models with the hierarchical multi-factor error structure whereby an external shock can alter the trade costs for individual country relative to all other countries in a heterogeneous and time-dependent way (e.g. Kapetanios and Shin, 2017).

7 Conclusion

Given the growing availability of the big dataset which contain information on multiple dimensions, the recent literature on the panel data have focused more on extending the two-way error components models to the multidimensional setting. We propose novel estimation techniques to accommodate cross-sectional error dependence within the 3D panel data models. Despite the massive development of modelling residual CSD through unobserved factors in the 2D panel data models (e.g. Pesaran, 2006; Bai, 2009), our approach is the first attempt to introduce strong CSD into the multi-dimensional error components, and well

[17]BHP show that the values of $\alpha \in [1/2, 3/4)$ represent a moderate degree of CSD.

Table 3: 3D panel gravity model estimation results for bilateral export flows

	FE			CTFE		
	Coeff	se	t-ratio	Coeff	se	t-ratio
gdph	2.185	0.041	52.97			
gdpf	1.196	0.041	28.98			
sim	-0.263	0.052	-5.069	-0.055	0.074	-0.754
rlf	0.031	0.006	5.011	0.006	0.005	1.294
rer	0.005	0.007	0.791	0.031	0.072	0.436
cee	0.302	0.014	22.05	0.290	0.017	16.99
emu	0.204	0.019	10.71	-0.011	0.036	-0.315
CD stat	620.1			-2.676		
	$\alpha_{0.05}$	α	$\alpha_{0.95}$	$\alpha_{0.05}$	α	$\alpha_{0.95}$
CSD exponent	0.925	0.992	1.059	0.865	0.914	0.963
	2D PCCE			3D PCCE		
	Coeff	se	t-ratio	Coeff	se	t-ratio
gdph	0.289	0.095	3.033			
gdpf	1.491	0.095	15.69			
sim	0.042	0.105	0.401	1.032	0.111	9.290
rlf	0.007	0.005	1.420	-0.004	0.005	-0.748
rer	0.144	0.019	7.427	0.168	0.114	1.471
cee	0.187	0.014	13.20	0.335	0.022	15.10
emu	0.018	0.015	1.160	0.081	0.045	1.793
CD stat	76.11			-4.19		
	$\alpha_{0.05}$	α	$\alpha_{0.95}$	$\alpha_{0.05}$	α	$\alpha_{0.95}$
CSD exponent	0.837	0.867	0.897	0.724	0.775	0.826

Notes: Using the annual dataset over 1960-2008 for 182 country-pairs amongst 14 EU member countries, we estimate the 3D panel gravity specification, (69). FE stands for the two-way fixed effects estimator with country-pair and time effects. CTFE refers to the 3D within estimator given by (5). 2D PCCE estimator is given by (11) with factors $\mathbf{f}_t = \left\{\overline{gdp}_{.t}, \overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{cee}_{..t}, \overline{rer}_t, t\right\}$. 3D PCCE estimator is given by (18) with factors $\mathbf{f}_t = \left\{\overline{sim}_{..t}, \overline{rlf}_{..t}, \overline{rer}_t, t\right\}$. CD test refers to testing the null hypothesis of residual cross-sectional error independence or weak dependence and is defined in (34). α denotes the estimate of CSD exponent jointly with the 90% confidence bands.

suited to the analysis of sophisticated error CSD across the triple or higher dimensions.

We develop the two-step consistent estimation procedure, called the 3D-PCCE estimator. We discuss the extent of cross-section dependence and develop a diagnostic test for the null hypothesis of (pairwise) residual cross-section independence or weak dependence in the 3D panels. The empirical usefulness and superiority of the proposed the 3D-PCCE estimator are demonstrated via the Monte Carlo studies and the empirical application to the 3D panel gravity model of the intra-EU trade.

At this stage, it seems appropriate to mention the number of obvious and challenging extensions and generalisations. First, as discussed in Section 4, we will address the number of extensions to the analysis of incomplete panel dataset and 4D or higher dimensional models. Second, as an ongoing research, we develop the general multi-dimensional heterogenous panel data models with hierarchical multi-factor error structure (e.g. Kapetanios and Shin, 2017). Third, our proposed approach can be easily extended to dynamic panels. Finally and more importantly, we aim to develop the most challenging models by combining both the spatial-based and the factor-based techniques within the 3D or higher dimensional models. Bailey et al. (2016) develop the multi-step estimation procedure that can distinguish the relationship between spatial units that is purely spatial from that which is due to the effect of common factors. Furthermore, Mastromarco et al. (2016a) propose the technique for allowing weak and strong CSD in modelling technical efficiency of stochastic frontier panels by combining the exogenously driven factor-based approach and an endogenous threshold regime selection by Kapetanios et al. (2014). Bai and Li (2015) and Shi and Lee (2014,5) have developed the framework for jointly modelling spatial effects and interactive effects. See also Gunnella et al. (2015) and Kuersteiner and Prucha (2015). This is the most recent research trend, and thus the successful development of the general combined approach within the multi-dimensional panels may broaden its appeal further.

Reference

Abowd, J.M., F. Kramarz and D.N. Margolis (1999): "High Wage Workers and High Wage Firms", *Econometrica* 67: 251-333.

Anderson, J. and E. van Wincoop (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review* 93: 170-92.

Bai, J. (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica* 77: 1229-1279.

Bai, J. and K. Li (2014): "Spatial Panel Data Models with Common Shocks," *mimeo.*, Columbia University.

Bai, J. and K. Li (2015): "Dynamic Spatial Panel Data Models with Common Shocks," *mimeo.*, Columbia University.

Bai, J., Y. Liao and J. Yang (2015): "Unbalanced Panel Data Models with Interactive Effects," in *The Oxford Handbook of Panel Data* ed. by B.H. Baltagi. Oxford University Press: New York, 149-170.

Baier, S.L. and J.H. Bergstrand (2007): "Do Free Trade Agreements Actually Increase Members International Trade?" *Journal of International Economics* 71:72-95.

Bailey, N., S. Holly and M.H. Pesaran (2016): "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-sectional Dependence," forthcoming in *Journal of Applied Econometrics*.

Bailey, N., G. Kapetanios and M.H. Pesaran (2016): "Exponent of Cross-sectional Dependence: Estimation and Inference," forthcoming in *Journal of Applied Econometrics*.

Balazsi, L., M. Bun, F. Chan and M.N. Harris (2016): "Models with Endogenous Regressors," Chapter 3 in *The Econometrics of Multi-dimensional Panels* ed. by L. Matyas. Springer: Berlin.

Balazsi, L., B.H. Baltagi, L. Matyas and D. Pus (2016): "Modelling Multi-dimensional Panel Data: A Random Effects Approach", *mimeo.*, Central European University.

Balazsi, L., L. Matyas and T. Wansbeek (2015): "The Estimation of Multi-dimensional Fixed Effects Panel Data Models", forthcoming in *Econometric Reviews*.

Baldwin, R.E. (2006): *In or Out: Does it Matter? An Evidence-Based Analysis of the Euro's Trade Effects*. Centre for Economic Policy Research.

Baldwin, R.E. and D. Taglioni (2006): "Gravity for Dummies and Dummies for Gravity Equations," NBER Working Paper 12516.

Baltagi B.H and G. Bresson (2016): "Modelling Housing Using Multi-dimensional Panel Data," Chapter 12 in *The Econometrics of Multi-dimensional Panels* ed. by L. Matyas. Springer: Berlin.

Baltagi, B.H., P. Egger, P. and M. Pfaffermayr (2003): "A Generalized Design for Bilateral Trade Flow Models," *Economics Letters* 80: 391-397.

Baltagi, B.H., P. Egger, P. and M. Pfaffermayr (2015): "Panel Data Gravity Models of International Trade," in *The Oxford Handbook of Panel Data* ed. by B.H. Baltagi. Oxford University Press: New York, 608-641.

Baltagi, B.H., P. Egger and K. Erhardt (2016): "The Estimation of Gravity Models in International Trade," Chapter 3 in *The Econometrics of Multi-*

dimensional Panels ed. by L. Matyas. Springer: Berlin.

Behrens, K., C. Ertur andW. Kock (2012): "Dual Gravity: Using Spatial Econometrics to Control For Multilateral Resistance," *Journal of Applied Econometrics* 27: 773-794.

Bertoli, S. and J. Fernandez-Huertas Moraga (2013): "Multilateral Resistance to Migration," *Journal of Development Economics* 102: 79-100.

Chudik, A., M.H. Pesaran and E. Tosetti (2011): "Weak and Strong Cross-section Dependence and Estimation of Large Panels," *Econometrics Journal* 14: 45-90.

Davis, P. (2002): "Estimating Multi-way Error Components Models with Unbalanced Data Structures," *Journal of Econometrics* 106, 67-95.

De Nardis, S. and C. Vicarelli (2003): "Currency Unions and Trade: The Special Case of EMU," *World Review of Economics* 139: 625-49.

Feenstra, R.C. (2004): *Advanced International Trade*. Princeton University Press: Princeton, NJ.

Frankel, J.A. (2008): "The Estimated Effects of the Euro on Trade: Why are They below Historical Effects of Monetary Unions among Smaller Countries?" NBER Working Paper 14542.

Frankel, J.A. and A.K. Rose (2002): "An Estimate of the Effect of Common Currencies on Trade and Income," *Quarterly Journal of Economics* 117: 437-466.

Glick R. and A.K. Rose (2002): "Does a Currency Union Affect Trade? The Time Series Evidence," *European Economic Review* 46: 1125-1151.

Gunnella V., C. Mastromarco, L. Serlenga and Y. Shin (2015): "The Euro Effects on Intra-EU Trade Flows and Balances: Evidence from the Cross Sectionally Correlated Panel Gravity Models," *mimeo.*, University of York.

Hausman, J.A. and W.E. Taylor (1981): "Panel Data and Unobservable Individual Effect," *Econometrica* 49: 1377-1398.

Helpman, E. (1987): "Imperfect competition and international trade: evidence from fourteen industrialized countries, " *Journal of the Japanese and International Economies*, 1: 62?81.

Herwartz, H. and H. Weber (2010): "The Euro's Trade Effect under Cross-sectional Heterogeneity and Stochastic Resistance", Kiel Working Paper No. 1631, Kiel Institute for the World Economy, Germany.

Kapetanios, G. and M.H. Pesaran (2005): "Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns,", CESifo Working Paper Series 1416, CESifo Group Munich.

Kapetanios, G. and Y. Shin (2017): "Multi-dimensional Heterogeneous Panel Data with Hierarchical Multi-factor Error Structure," *mimeo*, University of York.

Kramarz, F., S.J. Machin and A. Ouazad (2008): "What Makes a Test Score? The Respective Contributions of Pupils, Schools, and Peers in Achievement in English Primary Education," INSEAD Working Paper.

Kuersteiner, G.M. and I.R. Prucha (2015): "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," *mimeo.*, University of

Maryland.

Krugman, P.R. (1997): "Increasing Returns, Monopolistic Competition and International Trade," *Journal of International Economics*, 9: 469-479.

Le Gallo, J. and A. Pirotte (2016): "Models for Spatial Panels," Chapter 9 in *The Econometrics of Multi-dimensional Panels* ed. by L. Matyas. Springer: Berlin.

Mastromarco, C., L. Serlenga and Y. Shin (2016a): "Modelling Technical Inefficiency in Cross Sectionally Dependent Stochastic Frontier Panels," *Journal of Applied Econometrics* 31, 281-297.

Mastromarco, C., L. Serlenga and Y. Shin, (2016b): "Multilateral Resistance and Euro Effects on Trade Flows," *Spatial Econometric Interaction Modelling* eds. by G. Arbia and R. Patuelli. Springer: Berlin, 253-278.

Matyas, L. (1997): "Proper Econometric Specification of the Gravity Model," *The World Economy* 20: 363-369.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica* 46, 69-85.

Pakko, M.R. and H.J. Wall (2002): "Reconsidering the Trade-creating Effects of a Currency Union," *Federal Reserve Bank of St Louis Review* 83: 37-46.

Pesaran, M.H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica* 74: 967-1012.

Pesaran, M.H. (2015): "Testing Weak Cross-sectional Dependence in Large Panels," *Econometric Reviews* 34: 1089-1117.

Persson T. (2001): "Currency Union and Trade: How Large is the Treatment Effect?" *Economy Policy* 33: 435-448.

Rose, A.K. (2001): "Currency Unions and Trade: The Effect is Large," *Economic Policy* 33: 449-61.

Serlenga, L. and Y. Shin (2007): "Gravity Models of Intra-EU Trade: Application of the CCEP-HT Estimation in Heterogeneous Panels with Unobserved Common Time-specific Factors," *Journal of Applied Econometrics* 22: 361-381.

Shi, W. and L.F. Lee (2014): "Spatial Dynamic Panel Data Models with Interactive Fixed Effects," *mimeo.*, Ohio State University.

Wooldridge, J.M. (2010): "Correlated Random Effects Models with Unbalanced Panels," *mimeo.*, Michigan State University.