Heterogeneity and Cross-Sectional Dependence in Panels: Heterogeneous *vs.* Homogeneous Estimators

Oguzhan Akgun[a], Alain Pirotte[a] and Giovanni Urga[b]

[a] *CRED, University Paris II Panthéon-Assas, France*

[b] *Cass Business School, London, United Kingdom, and Bergamo University, Italy*

This paper focuses on the comparison of homogeneous and heterogeneous panel data estimators, including partially heterogeneous ones, in presence of cross-sectional dependence generated by common factors and spatial error dependence. Our specifications allow us to consider and contrast weak cross-sectional dependence and strong cross-sectional dependence in a general linear heterogeneous panel data model. An overview of the estimation procedures, including heterogeneous, homogeneous and partially heterogeneous estimators, is presented. Then, an extensive Monte Carlo study is conducted using a general framework encompassing recent contributions in the literature especially in terms of considering common factors and spatial dependence simultaneously. Our simulation results show that, even for small individual and time dimensions, heterogeneous estimators perform better in terms of bias, root mean squared error, size and size adjusted power compared to homogeneous estimators. Last, the superiority of the heterogeneous estimators is confirmed by an empirical application relating fiscal decentralization and government size in 22 OECD countries over the period 1973-2017.

Keywords: Panel data models – heterogeneity – homogeneity – cross-sectional dependence – spatial panel – common factors – forecasting.

Email addresses: oguzhan.akgun@u-paris2.fr (O. Akgun), alain.pirotte@u-paris2.fr (A. Pirotte), G.Urga@city.ac.uk (G. Urga).

Hétérogénéité et Dépendance Inter-Individuelle sur Données de Panel : Estimateurs Hétérogènes *vs.* Homogènes

Cet article se focalise sur les estimateurs hétérogènes versus homogènes sur données de panel, y compris ceux partiellement hétérogènes, en présence de dépendances inter-individuelles via une structure de dépendance spatiale des perturbations et/ou en présence de facteurs communs observables/inobservables. Ces estimateurs sont liés à des spécifications qui permettent de combiner et de distinguer les dépendances inter-individuelles faibles (reliées à une matrice spatiale) des dépendances inter-individuelles fortes (i.e. les facteurs communs). Une présentation générale des trois catégories d'estimateurs ci-avant mentionnées est d'abord faite. Ensuite, une approche par simulation de Monte Carlo, plus générale que celles précédemment menées, est retenue pour étudier les propriétés de ces estimateurs en présence de facteurs communs et de dépendance spatiale. Les résultats montrent que, même pour des dimensions individuelle et temporelle faibles, les estimateurs hétérogènes fournissent de meilleurs résultats en termes de biais, de RMSE, de taille et de puissance des tests bilatéraux comparativement à ceux obtenus sur la base des estimateurs homogènes. Enfin, une application empirique, qui s'intéresse à l'impact de la décentralisation fiscale sur la taille des gouvernements de 22 pays de l'OCDE sur la période 1973-2017, confirme cette supériorité.

Mots-clés : Modèles sur données de panel – hétérogénéité – homogénéité – dépendance inter-individuelle – panel spatial – facteurs communs – prévision.

Classification JEL : C13, C23.

1 Introduction

The optimal estimation strategy for panels with heterogeneous slope coefficients is an open question in the econometric literature. The presence of cross-sectional dependence (CD) makes the choice between available estimators more difficult. In this paper our interest lies on identifying the optimal estimation methods in linear panel data models with random slope coefficients that display CD.

The main aim of the paper is to explore the impact of different types and strength of CD on the performance of homogeneous (pooled) and heterogeneous estimators, including partially heterogeneous ones, in presence of low and high degrees of slope heterogeneity using simulated and real data. To this end, we first review the most recent contributions to the literature on panel data models with CD, such as Pesaran (2006), Kapetanios and Pesaran (2007), Bai (2009) and Song (2013). To investigate the role of heterogeneity on the optimal strategy, in addition to fully heterogeneous and fully homogeneous estimators, we consider the partially heterogeneous frameworks proposed by Bonhomme and Manresa (2015a) and Su et al. (2016a). Our paper is the first in the literature to discuss this last class of estimators in a comparative manner.

To evaluate the small sample properties of the estimators under consideration, we conduct a Monte Carlo exercise using a general linear panel data model with CD where the correlations among panel units arise from both common factors and spatial dependence. Our framework allows us to study the impact of strong cross-sectional dependence (SCD) that results from common factors and weak cross-sectional dependence (WCD), generated by spatial dependence in the error terms, on the performance of heterogeneous and homogeneous estimators.

Our results indicate that heterogeneous estimators perform better than the homogeneous and partially heterogeneous estimators in terms of estimation bias and root mean squared error (RMSE) as well as correct inference which we evaluate by size and size adjusted power of hypothesis tests concerning the average slope coefficients. Contrary to the previous literature comparing the heterogeneous and homogeneous estimators in the absence of CD, our result on the dominance of the heterogeneous estimators is found to be valid even when time and individual dimensions are small. This is found to be the case in low and high degrees of heterogeneity observed commonly in practice. Notably, we study the properties of the estimation and inference procedures for heterogeneity degrees generally higher than the ones in previous Monte Carlo studies in the literature. We apply the methods in an empirical model relating the size of governments in OECD countries with their level of fiscal decentralization. This illustration also documents the importance of taking into account heterogeneity among panel units.

Motivation and related literature. It is now well known that pooling in presence of slope heterogeneity can produce misleading results on the magnitude of the average effects, inference based on them and on predictive performance. In a random coefficients model, average effects can be estimated consistently by pooled estimators with strictly exogenous regressors. However, in a seminal paper Pesaran and Smith (1995) show that the pooled estimators are not consistent for the average effect if the model contains weakly exogenous regressors. A large literature compare the heterogeneous and homogeneous estimators, though almost always in a setting which does not allow correlations among units. Some examples are Garcia-Ferrer et al. (1987), Baltagi and Griffin (1997), Baltagi et al. (2000), Hoogstrate et al. (2000), Baltagi et al. (2003), Baltagi et al. (2004), Mark and Sul (2012). These studies mostly reach the conclusion that the pooled estimators outperform the heterogeneous ones although some authors such as Hoogstrate et al. (2000) and Mark and Sul (2012) point out to the fact that the dominance of pooled estimators is a result of limited number of time series observations and the low degree of heterogeneity between units.

Depending on its strength and nature, CD can have similar consequences. Is it a result of local interactions generating spatial spillover effects or common factors which affect different units (see Chudik et al., 2011; Sarafidis and Wansbeek, 2012; Bailey et al., 2016b, and references therein)? With the increasing availability of data, we have more panels where both N and T are large. This offers new possibilities and challenges on the ways of characterizing CD. Several papers have distinguished between WCD and strong SCD. In terms of their effects on the statistical properties of conventional panel data estimators, the two types of CD can differ dramatically, therefore it is necessary to analyze them in a comparative way. The literature dealing with spatial interactions and common factors simultaneously is scarce, some exceptions being Bai (2009), Pesaran and Tosetti (2011), Bailey et al. (2016a), Shi and Lee (2018) and Kuersteiner and Prucha (2018).

However, as Sarafidis and Wansbeek (2012) underline, the literature does not provide a unique definition of these terms. In this paper, we adapt the definition by Chudik et al. (2011) where the former is a result of spatial interactions whereas the source of the latter is the common factors (see also Bailey et al., 2016b, and references therein). The factor and spatial econometric approaches tend to complement one another, with the factor approach being more suitable for modeling SCD, and the spatial approach generally requiring WCD, as defined in Chudik et al. (2011).

Organization. The paper is organized as follows. In Section 2, we present the heterogeneous panel data setup including simultaneously common factors and spatial error dependence. In this section, the associated estimation procedures are described. Section 3 deals with the Monte Carlo study including the design of experiments and the discussion of results. In Section 4, we present an empirical application using the methods described and compared in previous sections. Section 5 summarizes the main findings and provides some guidelines for future research.

2 The model and methods of estimation

2.1 Heterogeneous panel model with CD

We consider linear panel data models with random slope coefficients that display a CD structure in the disturbances. We use the following model with common factors and spatial error dependence:

$$y_{it} = \boldsymbol{\alpha}_i'\mathbf{d}_t + \boldsymbol{\beta}_i'\mathbf{x}_{it} + \boldsymbol{\gamma}_i'\mathbf{f}_t + u_{it}, \tag{1}$$

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{d}_t + \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{2}$$

with

$$u_{it} = \rho_i \sum_{j=1}^{N} w_{ij}\xi_{jt} + \varepsilon_{it}, \tag{3}$$

where y_{it} is the dependent variable for unit $i = 1, 2, \ldots, N$ at time $t = 1, 2, \ldots, T$, $\mathbf{x}_{it} = (x_{i1t}, x_{i2t}, \ldots, x_{ikt})'$ is a $(k \times 1)$ vector of observed regressors, $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{ik})'$ represents the corresponding $(k \times 1)$ slope parameters to be estimated. The regressors are assumed to be strictly exogenous but we also discuss the possibility of having weakly exogenous regressors. $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$ is a $(m \times 1)$ vector of unobserved common factors, $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{im})'$ is the associated $(m \times 1)$ vector of factor loadings, while $\mathbf{d}_t$ is a $(l \times 1)$ vector of observed common factors, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{il})'$ is its $(l \times 1)$ vector of factor loadings. $\mathbf{A}_i$ is the $(l \times k)$ matrix of factor loadings of the observed common factors and $\boldsymbol{\Gamma}_i$ is the $(m \times k)$ matrix of factor loadings of the unobserved common factors. The vector error process $\mathbf{v}_{it}$ is allowed to be autocorrelated and can exhibit spatial correlation. The number of unobserved common factors, m, is assumed to be fixed relative to N, in particular m is assumed to be strictly smaller than N. The error term u_{it} is assumed to follow a spatial pattern with ρ_i being its associated parameter. We assume that the slope coefficients are generated by a random coefficients model as

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\delta_i}, \quad \boldsymbol{\delta}_i \sim \text{IID}(\mathbf{0}, \boldsymbol{\Omega}_\delta), \quad i = 1, 2, \ldots, N, \tag{4}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$, $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2}, \ldots, \delta_{ik})'$, $\boldsymbol{\delta_i}$ are distributed independently of $\boldsymbol{\gamma}_j$, $\boldsymbol{\Gamma}_j$, ε_{jt}, $\mathbf{v}_{jt}$ and the common factors $\mathbf{d}_t$ and $\mathbf{f}_t$ for all $j = 1, \ldots, N$, $t = 1, 2, \ldots, T$. Throughout the paper our interest lies on the expected value of the random coefficients over panel units, namely $\boldsymbol{\beta}$. We also report estimates of the unit-specific marginal effects in our empirical application.

The model in (3) contains as special cases all commonly used spatial processes like spatial autoregression (SAR), spatial moving average (SMA), and spatial error components (SEC) as well as their higher order versions which can be obtained by alternating the definition of the variable ξ_{jt}. In all cases, we assume that the usual boundedness of the row and column sums of the spatial weights matrix defined by the double indexed sequence w_{ij} is satisfied (see, for instance, Pesaran and Tosetti, 2011).

The structure of CD described above combines WCD and SCD as defined by Chudik et al. (2011) where the former is a result of spatial error correlation whereas the source of the latter is the common factors. These two types of CD differ importantly in terms of their effect on parameter estimation. Unobserved common factors create an identification problem on the estimation of the slope parameters due to their potential correlation with the explanatory variables. Whereas spatial error dependence have an impact on the estimation efficiency and on inference on the model parameters.

To illustrate the two dimensions, we may refer to cross country growth regressions. On one hand, the SCD can be viewed as a result of a number of observed and or unobserved common factors that may have different effects on total factor productivity across countries. These include, for instance, aggregate technological shocks or oil price shocks that may affect total factor productivity through their effects on production costs. On the other hand, WCD can be viewed as a result of spatial spillover effects such as international technology diffusion which can be related to geographical distance due to transport costs or geographical barriers.

2.2 Methods of estimation

In this section, we review the recent studies dealing with the problem of unobserved common factors in heterogeneous and homogeneous linear panel data models as well as the partially heterogeneous estimators. Our main focus is on the common correlated effects (CCE) estimators proposed by Pesaran (2006), principal components (PC) approaches advanced by Kapetanios and Pesaran (2007), Bai (2009) and Song (2013), and partially heterogeneous estimators of Bonhomme and Manresa (2015a) and Su et al. (2016a).

2.2.1 Heterogeneous and homogeneous estimators

Pesaran (2006) shows that cross-sectional averages of dependent variable and explanatory variables can be used as observed proxies in order to estimate the slope parameters consistently when the number of cross-sectional units is large. The author works on the estimation of the model given in (1) and (2) using the cross-sectional averages of the explanatory variables and the dependent variable as observed proxies for the unobserved common factors as $N \to \infty$. For the unit-specific slope coefficient, the CCE estimator is given by

$$\widehat{\boldsymbol{\beta}}_{CCE,i} = \left(\mathbf{X}_{i.}'\bar{\mathbf{M}}_f\mathbf{X}_{i.}\right)^{-1}\mathbf{X}_{i.}'\bar{\mathbf{M}}_f\mathbf{y}_{i.}, \tag{5}$$

with

$$\bar{\mathbf{M}}_f = \mathbf{I}_T - \bar{\mathbf{H}}(\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-}\bar{\mathbf{H}}'. \tag{6}$$

where $\bar{\mathbf{H}} = (\mathbf{D}, \bar{\mathbf{Z}})$, $\mathbf{D} = (\mathbf{d}_1', \mathbf{d}_2', \ldots, \mathbf{d}_T')'$, $\bar{\mathbf{Z}} = (\bar{\mathbf{z}}_{.1}', \bar{\mathbf{z}}_{.2}', \ldots, \bar{\mathbf{z}}_{.T}')'$, $\bar{\mathbf{z}}_{.t} = N^{-1}\sum_{i=1}^{N}\mathbf{z}_{it}$ and $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$. Pesaran shows that the consistency of the estimator follows when $(N,T) \to \infty$ and if a rank condition on the factor loadings is satisfied given that the regressors are strictly exogeneous.[1] The estimator is asymptotically normal if $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$, with convergence rate being $\sqrt{T}$. Chudik and Pesaran (2015) show that with weakly exogenous regressors *CCE* estimation is still valid as long as a sufficient number of lags of the cross-sectional averages $\bar{\mathbf{z}}_{.t}$ are used to construct the matrix $\bar{\mathbf{M}}_f$.

Once the unit-specific parameters are estimated using (5), we can infer their mean from these estimates as the unit-specific slope parameters follow the random coefficient model given in (4). Pesaran considers a mean group estimator which is a simple average of the unit-specific slope parameter estimates. It is given by

$$\widehat{\boldsymbol{\beta}}_{CCEMG} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{CCE,i}. \tag{7}$$

The estimator does not require $T \to \infty$ for consistency; it is consistent for fixed T as $N \to \infty$. As $(N,T) \to \infty$, its distribution converges to a normal distribution at the rate of $\sqrt{N}$ and its asymptotic variance is determined by the variance of the random slope parameters.

Alternatively, one can use a pooled estimator which treats the slope coefficients as if they were the same across panel units. In particular, we can use

$$\widehat{\boldsymbol{\beta}}_{CCEP} = \left(\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}_f\mathbf{X}_{i.}\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}_f\mathbf{y}_{i.}, \tag{8}$$

where $\bar{\mathbf{M}}_f$ is defined as in (6). If the regressors $\mathbf{x}_{it}$ are strictly exogenous the estimator is consistent. Although it imposes homogeneity of the slope parameters, the estimator is asymptotically unbiased under (4). It is asymptotically normal with its convergence rate being $\sqrt{N}$ which is slower than the usual $\sqrt{NT}$ of the homogeneous panel data estimators.

The *CCE* estimators are based on the fact that cross-sectional averages of $\mathbf{z}_{it}$ can be used to remove the effect of common factors asymptotically. In general, cross-sectional averages of any subset of the elements of $\mathbf{z}_{it}$ provide similar results given that this subset satisfies the rank condition mentioned above. Moreover, additional exogenous variables can be used to improve estimation efficiency. In particular we can consider the estimator (5) by replacing $\bar{\mathbf{M}}_f$ with

$$\tilde{\mathbf{M}}_f = \mathbf{I}_T - \tilde{\mathbf{H}}(\tilde{\mathbf{H}}'\tilde{\mathbf{H}})^{-}\tilde{\mathbf{H}}', \tag{9}$$

[1]See Pesaran (2006) and Karabiyik et al. (2017) for a discussion on the rank condition.

with $\tilde{\mathbf{H}} = (\mathbf{D}, \bar{\mathbf{W}})$, $\bar{\mathbf{W}} = (\bar{\mathbf{w}}'_{.1}, \bar{\mathbf{w}}'_{.2}, \ldots, \bar{\mathbf{w}}'_{.T})'$, $\bar{\mathbf{w}}_{.t} = N^{-1} \sum_{i=1}^{N} \mathbf{w}_{it}$ where $\mathbf{w}_{it}$ is a vector of some exogenous variables which can include the explanatory variables themselves. This projection matrix uses only exogenous variables to construct observed proxies for the common factors. Its advantage is that the cross-sectional averages of $\mathbf{w}_{it}$ are exogenous by assumption, contrary to the cross-sectional averages of the dependent variable. In our paper, the pooled and mean group *CCE* estimators using only exogenous variables are called *CCEPX* and *CCEMGX*, respectively.

Another way to estimate the common factors is to apply principal component analysis (*PCA*) to the observed variables in the data set to estimate the unobserved common factors. Kapetanios and Pesaran (2007) suggest to use *PCA* to extract the common factors from $\mathbf{z}_{it}$. As shown by the authors, the small sample properties of the estimators based on this method are not satisfactory, potentially as a result of the *PCA* applied to the endogenous variable y_{it}. An alternative is to use only the exogenous variables to estimate the unobserved common factors. For the unit-specific slope parameters we use

$$\widehat{\boldsymbol{\beta}}_{PCX,i} = \left(\mathbf{X}'_{i.}\mathbf{M}_p\mathbf{X}_{i.}\right)^{-1}\mathbf{X}'_{i.}\mathbf{M}_p\mathbf{y}_{i.}, \tag{10}$$

with

$$\mathbf{M}_p = \mathbf{I}_T - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-}\mathbf{G}',$$

where $\mathbf{G} = (\mathbf{D}, \tilde{\mathbf{F}})$ and $\tilde{\mathbf{F}}$ being the matrix of observations on the principal components extracted from the matrix $\sum_{i=1}^{N} \mathbf{X}_{i.}\mathbf{X}'_{i.}$. A mean group estimator can be computed from these estimates as in the case of *CCE* estimators using

$$\widehat{\boldsymbol{\beta}}_{PCMGX} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{PCX,i}, \tag{11}$$

and the pooled estimator is defined as

$$\widehat{\boldsymbol{\beta}}_{PCPX} = \left(\sum_{i=1}^{N}\mathbf{X}'_{i.}\mathbf{M}_p\mathbf{X}_{i.}\right)^{-1}\sum_{i=1}^{N}\mathbf{X}'_{i.}\mathbf{M}_p\mathbf{y}_{i.}. \tag{12}$$

A related but different estimator is proposed by Bai (2009) in a homogeneous slope framework. The model that the author considers is similar to the one in (1) but the estimator proposed does not require the explanatory variables to be related to the unobserved common factors as in (2). Also the author does not consider the observed common factors which is taken into account in our implementation of the estimator. This estimator is the solution to the following non-linear equations:

$$\widehat{\boldsymbol{\beta}}_{IPCP} = \left(\sum_{i=1}^{N}\tilde{\mathbf{X}}'_{i.}\tilde{\mathbf{X}}_{i.}\right)^{-1}\sum_{i=1}^{N}\tilde{\mathbf{X}}'_{i.}(\tilde{\mathbf{y}}_{i.} - \widehat{\mathbf{F}}\widehat{\boldsymbol{\gamma}}_i), \tag{13}$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\tilde{\mathbf{y}}_{i.}-\tilde{\mathbf{X}}_{i.}\widehat{\boldsymbol{\beta}}_{IPCP}\right)\left(\tilde{\mathbf{y}}_{i.}-\tilde{\mathbf{X}}_{i.}\widehat{\boldsymbol{\beta}}_{IPCP}\right)'\right]\widehat{\mathbf{F}}=\widehat{\mathbf{F}}\widehat{\mathbf{V}}_{NT},$$

where $\widehat{\boldsymbol{\gamma}}_i = T^{-1}\widehat{\mathbf{F}}'(\tilde{\mathbf{y}}_{i.} - \tilde{\mathbf{X}}_{i.}\widehat{\boldsymbol{\beta}}_{IPCP})$, $\tilde{\mathbf{X}}_{i.} = \mathbf{M}_D\mathbf{X}_{i.}$, $\tilde{\mathbf{y}}_{i.} = \mathbf{M}_D\mathbf{y}_{i.}$, $\widehat{\mathbf{V}}_{NT}$ is a diagonal matrix containing the m largest eigenvalues of the matrix in the brackets on the left hand side of the equation and $\widehat{\mathbf{F}}$ are the corresponding eigenvectors. To obtain the final estimator of the slope parameters, one can iterate between these two equations until convergence is achieved.

Song (2013) generalizes this iterative estimation procedure to allow for heterogeneity in slope parameters among units. Our implementation of this unit-specific estimator is given by

$$\widehat{\boldsymbol{\beta}}_{IPC,i}=\left(\tilde{\mathbf{X}}_{i.}'\tilde{\mathbf{X}}_{i.}\right)^{-1}\tilde{\mathbf{X}}_{i.}'(\tilde{\mathbf{y}}_{i.}-\tilde{\mathbf{F}}\tilde{\boldsymbol{\gamma}}_i), \tag{14}$$

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(\tilde{\mathbf{y}}_{i.}-\tilde{\mathbf{X}}_{i.}\widehat{\boldsymbol{\beta}}_{IPC,i}\right)\left(\tilde{\mathbf{y}}_{i.}-\tilde{\mathbf{X}}_{i.}\widehat{\boldsymbol{\beta}}_{IPC,i}\right)'\right]\tilde{\mathbf{F}}=\tilde{\mathbf{F}}\tilde{\mathbf{V}}_{NT},$$

with $\tilde{\boldsymbol{\gamma}}_i = T^{-1}\tilde{\mathbf{F}}'(\tilde{\mathbf{y}}_{i.} - \tilde{\mathbf{X}}_{i.}\widehat{\boldsymbol{\beta}}_{IPC,i})$, $\tilde{\mathbf{V}}_{NT}$ is a diagonal matrix containing the m largest eigenvalues of the matrix in the brackets on the left hand side of the equation and $\tilde{\mathbf{F}}$ are the corresponding eigenvectors. The author does not consider the asymptotic distribution of a pooled estimator of the average value of the heterogeneous slopes. However, a mean group estimator based on the unit-specific estimates is used which is given by

$$\widehat{\boldsymbol{\beta}}_{IPCMG}=\frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{IPC,i}. \tag{15}$$

The advantage of the iterative *PC* estimators by Bai (2009) and Song (2013) is that they assume that the factor loadings and the factors are fixed. Therefore, they do not require any constraints on the correlation between the explanatory variables and the common factors. In particular, the data generating process for the explanatory variables is left unrestricted and does not have to be in the form of (2). For instance, the explanatory variables can be related to the common factors in a non-linear manner. In this case, the estimators which uses explanatory variables to estimate the common factors, e.g. *CCE* estimators or non-iterative *PC* estimators, can fail to estimate the slope parameters consistently. A disadvantage of these estimators is the fact that the number of common factors may not be known in practice. However, it is possible to use information criteria proposed by Bai and Ng (2002) to consistently estimate the number of common factors. Furthermore, Moon and Weidner (2015) show that the consistency of the estimator of Bai (2009) does not require a consistent estimation of the number of factors. In fact, as soon as the number of factors is not underestimated, the resulting estimators are asymptotically equivalent to the estimator based on the true number of common factors.

It is important to note that the estimators by Bai (2009) and Song (2013) are iterative estimators and in practice they need to be initialized by estimates of the slope parameters.

Originally, Bai (2009) used several initial estimators for the slope parameters, such as *OLS*, fixed effects (*FE*) and two-way fixed effects (*2WFE*) estimators. Jiang et al. (2017) show that unless the initial estimator of the slope parameters is consistent, the consistency of these iterative approaches are not guaranteed. In this paper, the iterative procedures are initialized using the estimators *PCPX* and *PCMGX*. These estimators are consistent as soon as the explanatory variables have the factor structure given in (2).

We consider some additional estimators based on the iterative *PC* methods. Using consistent initialization, as with the estimators *PCPX* and *PCMGX*, an option is to stop after the first iteration. This estimator is consistent and has the advantage of being computationally less demanding than the iterative estimators. In addition, if the data generating process (DGP) of the dependent variable contains common factors which do not appear in the process generating the explanatory variables, our results underline that this procedure produce less bias and more efficient estimates. These estimators are called *PCPX2S* and *PCMGX2S*.

2.2.2 Partially heterogeneous estimators

Recently, several papers have considered the possibility of having a grouped structure in the slope parameters of a panel data model in contrast to the fully heterogeneous or fully homogeneous setting taken into consideration in the previous subsection. Here we discuss two main approaches to the problem, namely the grouped fixed effects (*GFE*) approach of Bonhomme and Manresa (2015a) and the classifier Lasso (*C-Lasso*) approach of Su et al. (2016a). These two papers assume that the slope parameters in model (1) satisfy

$$\boldsymbol{\beta}_i = \sum_{g=1}^{K} \boldsymbol{\lambda}_g \mathbb{1}\{i \in G_g\}, \tag{16}$$

where K is the number of groups which is assumed here to be known and fixed, G_g is the set of indexes of N_g units which belong to group g, $\boldsymbol{\lambda}_g \neq \boldsymbol{\lambda}_{g'}$, $\boldsymbol{\phi}_g \neq \boldsymbol{\phi}_{g'}$, $\forall\ g \neq g'$ and $G_g \cap G_{g'} = \emptyset$, $\forall\ g \neq g'$.

In an extension of their original model, Bonhomme and Manresa (2015a) assume simultaneously grouped structures for slope parameters and factor loadings. Our implementation of their *GFE* estimator is defined as the value which minimizes the objective function

$$Q_{GFE} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \bar{y}_{.t} - \boldsymbol{\lambda}'_{g_i} (\mathbf{x}_{it} - \bar{\mathbf{x}}_{.t}) - \mu_{g_i t} \right)^2, \tag{17}$$

where $g_i \in \{1, 2, \ldots, K\}$ is a variable which states the group which ith panel unit belongs to and $\mu_{g_i t}$ are the group-specific time effects. The estimation approach uses an algorithm based on some initial values of the unknown parameters in (17) and iterates until convergence. More precisely, the iterative algorithm consists of the following 4 steps:

Step 1: Select some starting values $\boldsymbol{\lambda}_g^{(0)}$ and $\mu_{gt}^{(0)}$, $\forall g, t$. Set $s = 0$.
Step 2: Compute for all $i \in \{1, 2, \ldots, N\}$

$$g_i^{(s+1)} = \underset{g \in \{1,2,\ldots,K\}}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_{it} - \bar{y}_{.t} - \boldsymbol{\lambda}_g^{(s)\prime}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{.t}) - \mu_{gt}^{(s)} \right)^2 .$$

Step 3: Compute

$$\left(\boldsymbol{\lambda}_{(s+1)}^K, \boldsymbol{\mu}_{(s+1)}^K \right) = \underset{\boldsymbol{\lambda}^N, \boldsymbol{\phi}^N}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \bar{y}_{.t} - \boldsymbol{\lambda}_{g_i^{(s+1)}}^{\prime} (\mathbf{x}_{it} - \bar{\mathbf{x}}_{.t}) - \mu_{g_i^{(s+1)} t} \right)^2 ,$$

where $\boldsymbol{\lambda}^K$ and $\boldsymbol{\mu}^K$ are matrices which contain all $\boldsymbol{\lambda}_g$'s and $\boldsymbol{\mu}_g$'s, respectively.
Step 4: Set $s = s + 1$ and go to step 2. Repeat the process until numerical convergence.

In practice, Bonhomme and Manresa suggest to select many starting values and choose the final estimate as the one which gives the minimum of the objective function (17). They propose also alternative algorithms which are more efficient in certain situations such as a big number of groups, e.g. $K > 10$.[2] In our simulations, we use small values of K. Therefore, the iterative algorithm described above allows to save computing time.

As Bonhomme and Manresa highlight, the *GFE* estimator can be seen as an alternative to the estimator of Bai (2009), developed for the case of unit-specific factor loadings whereas *GFE* identifies homogeneity of these loadings within some groups. The *IPC* estimator is expected to work in a DGP suitable for *GFE* but *GFE* should be biased in the case of fully heterogeneous factor loadings. The *GFE* estimator uses time dummies within groups, therefore, it does not restrict the number of the common factors to be known or even finite. This is not the case for the *IPC* estimator.

The *GFE* estimator is a least squares estimator which selects the grouping index g_i and the parameters $\boldsymbol{\lambda}_{g_i}$, $\mu_{g_i t}$ such that the sum of squared residuals defined by (17) is minimized. To have the intuition behind the estimation procedure, let us suppose that the group membership index g_i is known. In this case, the least squares estimate of the slope parameters $\boldsymbol{\lambda}_{g_i}$ is the usual two-way fixed effects (*2WFE*) estimator applied to each group $g = 1, 2, \ldots, K$ separately. Bonhomme and Manresa show that the above algorithm chooses g_i consistently and inference can be drawn on the slope parameter estimates using this strategy in usual ways.

An alternative way to estimate the group membership and the slope parameters is proposed by Su et al. (2016a), relying on group Lasso literature. They considered both linear and non-linear panel data models. Importantly, they assumed that the cross-sectional units are independent of each other. In our implementation of their estimator, we use the *CCE* transformation on the dependent and explanatory variables to deal with the unobserved common

[2]See Bonhomme and Manresa (2015b) for alternative algorithms and their comparison.

factors. The objective function defining the *C-Lasso* estimator is given by

$$Q_{C\text{-}Lasso} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{y}_{it} - \boldsymbol{\beta}_i'\tilde{\mathbf{x}}_{it}\right)^2 + \frac{\phi}{N}\sum_{i=1}^{N}\prod_{g=1}^{K}||\boldsymbol{\beta}_i - \boldsymbol{\lambda}_g||, \tag{18}$$

where $\tilde{y}_{it}$ is the tth element of $\bar{\mathbf{M}}_f\mathbf{y}_{i.}$, $\tilde{\mathbf{x}}_{it}$ is the tth column of $\mathbf{X}_{i.}'\bar{\mathbf{M}}_f$, and ϕ is a tuning parameter. Su et al. (2016b) suggest an iterative algorithm similar to the one given above to compute the *C-Lasso* estimates. They prove the asymptotic normality of their estimator under general conditions but under cross-sectional independence. Here, we use CD-robust estimates of the variances as explained in Appendix.

Some remarks comparing these two estimators with each other and with the estimators of the previous subsection follow. First, the *GFE* estimator is obtained under a grouped structure of the factor loadings in the DGP of the dependent variable whereas *C-Lasso* estimator is robust under the usual factor structure with fully heterogeneous loadings. However, as long as the factor loadings $\boldsymbol{\gamma}_i$ are independent of the explanatory variables $\mathbf{x}_{it}$ both estimators are consistent for the group specific slope parameters $\boldsymbol{\lambda}_g$. Intuitively, the fact that the *2WFE* estimator is consistent under the assumption of uncorrelated loadings (Sarafidis and Wansbeek, 2012) implies the consistency of the *GFE* estimator, in the light of the connection between the two estimators discussed above.

Second, an important difference exist between the heterogeneity patterns defined by (4) and (16). For the former we assumed that the unit-specific coefficients are independent of the explanatory variables whereas in the latter case the group-specific parameters are assumed to be fixed. The fixed coefficients assumption is arguably more general than the random coefficients assumption as it allows for correlation between the group-specific parameters and the explanatory variables. As shown by Breitung and Salish (2020) in the case of correlated coefficients, homogeneous estimators are inconsistent for the expected values of the slope parameters whereas the mean group estimators are generally consistent. As they impose homogeneity within groups, it follows that the partially heterogeneous estimators are inconsistent as well if the slope parameters are fully heterogeneous and correlated with the explanatory variables.

3 Monte Carlo study

3.1 Design of the experiments

Our setup generalizes in several directions the framework in Pesaran (2006) and Pesaran and Tosetti (2011). The dependent and the explanatory variables are generated by

$$y_{it} = \alpha_{i1} d_{1t} + \beta_{i1} x_{i1t} + \beta_{i2} x_{i2t} + \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + u_{it}, \tag{19}$$

$$x_{ijt} = a_{ij1} d_{1t} + a_{ij2} d_{2t} + \gamma_{ij1} f_{1t} + \gamma_{ij3} f_{3t} + v_{ijt}, \quad j = 1, 2, \tag{20}$$

where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, x_{ijt}, $j = 1, 2$, are the observed explanatory variables, d_{jt}, $j = 1, 2$, and f_{jt}, $j = 1, 2, 3$, are the observed and unobserved common factors, respectively, and α_{ij}, β_{ij} and γ_{ijk} are their respective coefficients. The error term of the dependent variable carries spatial dependence and it is generated as a SAR using

$$u_{it} = \rho_i \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it}, \ \text{ with } \ \varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2), \ \ \sigma_i^2 \sim \text{IIDU}(0.5, 1.5),$$

where w_{ij} is the element of the spatial weight matrix $\mathbf{W}_N$ in row i and column j. An SMA is also considered as a generating process but the results are similar and they are not reported here. A rook-type spatial weight matrix is used. We consider two different cases for ρ_i. These two cases are based on Baltagi and Pirotte (2010), with the main difference being heterogeneity of the parameters in the (first order) SAR (or SMA) models, where $\rho_i = \rho = (0.2, 0.8)$ which corresponds to low and high spatial dependence, respectively. Similarly, we generate the heterogeneous coefficients using

$$\rho_i = \rho + e_i^{\rho}, \ \text{ with } \ \rho = \{0.2, 0.8\}, \ \ e_i^{\rho} \sim \text{U}(-0.1, 0.1).$$

The observed and unobserved common factors are generated as follows

$$d_{1t} = 1, \quad d_{2t} = \rho_d d_{2,t-1} + v_{dt}, \quad v_{dt} \sim \mathcal{N}(0, 1 - \rho_d^2), \quad \rho_d = 0.5, \quad d_{20} = 0,$$
$$f_{jt} = \rho_{fj} f_{j,t-1} + v_{fjt}, \quad v_{fjt} \sim \mathcal{N}(0, 1 - \rho_{fj}^2), \quad \rho_{fj} = 0.5, \quad f_{j0} = 0, \quad j = 1, 2, 3.$$

The disturbances associated to the explanatory variables are generated by a stationary AR(1) process which is given by

$$v_{ijt} = \rho_{v_{ij}} v_{ij,t-1} + \epsilon_{ijt}, \quad \epsilon_{ijt} \sim \mathcal{N}(0, 1 - \rho_d^2), \quad \rho_{v_{ij}} \sim \text{IIDU}(0.05, 0.95),$$

assuming that $v_{ij0} = 0$, $j = 1, 2$. The first 10 observations are discarded to minimize the impact of initial values. The slope coefficients β_{ij} are generated under two different assumptions corresponding to high and low heterogeneity. They are given by

$$\beta_{ij} = \beta_j + \eta_{ij}, \quad \beta_j = 1, \quad \eta_{ij} \sim \text{IID}\mathcal{N}(0, \sigma_{\eta_j}^2),$$

Table 1: Summary of Experiments

Cases	Description	Parametrization
- Case *1*	Low Spatial & Low Factor Dependence	$\rho = 0.2$, $\gamma_{i1}, \gamma_{i2} \sim \text{IID}\mathcal{N}(1, 0.1)$
- Case *2*	Low Spatial & High Factor Dependence	$\rho = 0.2$, $\gamma_{i1}, \gamma_{i2} \sim \text{IID}\mathcal{N}(2, 0.4)$
- Case *3*	High Spatial & Low Factor Dependence	$\rho = 0.8$, $\gamma_{i1}, \gamma_{i2} \sim \text{IID}\mathcal{N}(1, 0.1)$
- Case *4*	High Spatial & High Factor Dependence	$\rho = 0.8$, $\gamma_{i1}, \gamma_{i2} \sim \text{IID}\mathcal{N}(2, 0.4)$

where $\sigma^2_{\eta_j} = 0.15$ and $\sigma^2_{\eta_j} = 0.3$, $j = 1, 2$, correspond to low and high heterogeneity, respectively. These heterogeneity levels in both cases are higher compared to those of Pesaran (2006) and Pesaran and Tosetti (2011). The loadings of the observed factors are generated as follows:

$$\alpha_{i1} \sim \text{IID}\mathcal{N}(1, 1), \quad (a_{i11}, a_{i21}, a_{i12}, a_{i22})' \sim \text{IID}\mathcal{N}(0.5\boldsymbol{\tau}_4, 0.5\mathbf{I}_4),$$

where $\boldsymbol{\tau}_4 = (1, 1, 1, 1)'$ and $\boldsymbol{I}_4$, an identity matrix of dimension (4×4). The loadings of the unobserved common factors in the equations for the explanatory variables are generated as

$$\begin{pmatrix} \gamma_{i11} & \gamma_{i13} \\ \gamma_{i21} & \gamma_{i23} \end{pmatrix} \sim \begin{pmatrix} \text{IID}\mathcal{N}(0.5, 0.5) & \text{IID}\mathcal{N}(0, 0.5) \\ \text{IID}\mathcal{N}(0, 0.5) & \text{IID}\mathcal{N}(0.5, 0.5) \end{pmatrix}.$$

To calculate the *CCE* estimators which use only exogenous variables, an additional variable x_{i3t} is generated as

$$x_{i3t} = a_{i31} d_{1t} + a_{i32} d_{2t} + \gamma_{i31} f_{1t} + \gamma_{i33} f_{2t} + v_{i3t},$$

where the factor loadings are given by

$$a_{i31}, a_{i32} \sim \text{IID}\mathcal{N}(1.5, 1.02),\ \gamma_{i31}, \gamma_{i32} \sim \text{IID}\mathcal{N}(1, 0.1).$$

The other terms in (3.1) are defined in the same way as those contained in explanatory variable DGPs (20).

Contrary to the case of the factor loadings in the process generating the explanatory variable x_{it}, in this paper we follow Trapani and Urga (2009) and Phillips and Sul (2003) and draw loadings to generate low and high CD. This is controlled as follows

$$\gamma_{i1}, \gamma_{i2} \sim \begin{cases} \text{IID}\mathcal{N}(1, 0.1) \text{ for Low CD,} \\ \text{IID}\mathcal{N}(2, 0.4) \text{ for High CD.} \end{cases} \tag{21}$$

The chosen parameters in (21) induce average correlation coefficients among panel units of 0.5 and 0.8, respectively. Different cases considered are summarized in Table 1. We consider

$(N, T) = \{20, 50, 100\}$. For each experiment, 2,000 replications are performed, and twelve estimators are implemented: five heterogeneous estimators, five homogeneous ones and two partially heterogeneous estimators. These are summarized below. For *PC* estimators, we assume that the number of unobserved common factors are known. To summarize, the estimators involved in our simulations are:

(i) *CCEMG*, *CCEP*: They are in (7) and (8) and are suggested by Pesaran (2006) which use cross-sectional averages of the explanatory variables and the dependent variable to proxy the unobserved common factors;

(ii) *CCEMGX*, *CCEPX*: The estimators are defined by the equation (9) which use the cross-sectional averages of the explanatory variables and an additional exogenous variable to proxy the unobserved common factors;

(iii) *PCMGX*, *PCPX*: They are in (11) and (12) which use *PCA* to extract the unobserved common factors from the only explanatory variables in addition to the factors estimated in the first stage;

(iv) *IPCP*, *IPCMG*: They are in (13) and (15). For these estimators we use *PCMGX* and *PCPX* as initial slope parameter estimates, respectively;

(v) *PCMGX2S*, *PCPX2S* : They are the two-stage estimators which make use of the factor estimates obtained from the residuals of the *PCMGX* and *PCPX* in addition to the first-stage factor estimates of these respective estimators;

(vi) *C-Lasso*, *GFE*: They are defined as the argument which minimize the objective functions in (17) and (18), respectively.

3.2 Results

In this paper, we focus on two cases which are found to be the most interesting ones in terms of distinguishing between methodologies. These are Case *1* and Case *3* described in Table 1. The main results for the heterogeneous and homogeneous estimators are summarized in Tables 2 and 3, whereas the results for the partially heterogeneous estimators are reported in Tables 4 and 5. The results for higher degree of heterogeneity are discussed briefly but are not reported in order to save space. Each table reports bias and RMSE associated with the coefficient estimates of the slope average β_1. Full set of results which consider different combinations of WCD and SCD, different types of spatial dependence, and additional estimators are available from the authors upon request.

3.2.1 Bias and RMSE results

Bias and RMSE results for the heterogeneous and homogeneous estimators are summarized in Tables 2 and 3, whereas the results for the partially heterogeneous estimators are reported in Tables 4 and 5 for the case of low heterogeneity.

Under the assumption of low heterogeneity of the slope coefficients, Tables 2 and 3 consider low factor dependence and low factor dependence and high spatial dependence, respectively. Let us first focus on the low factor dependence case in Tables 2. All estimators which control for unobserved common factors, except *PCMGX* and *PCPX*, provide very small bias values and their RMSEs decline steadily with the increase of either N or T. For *PCMGX* and *IPCP*, the bias and RMSE values are only slightly higher and they also decline steadily when either N or T is getting large. *IPCP* has generally the highest bias values.[3] For instance, when $N, T = 20$ its bias is 1% in absolute terms. For the largest sample size with $N, T = 100$ this value is 0.05% which is negligible. Although more biased in small samples, for *IPCP*, RMSE values stay in the same bounds as those of the other consistent estimators.

For example, when $N, T = 20$ its RMSE is 12.73. For $N = 20$ and $T = 100$ this value equals 10.44 whereas for $N = 100$ and $T = 20$ it is 5.54. This shows that, for a fixed T, having a larger N improves the performance of the estimator more than the increase in T for a fixed N. Overall the best performer is *CCEMG* in terms of RMSE in this case. For this estimator let us focus on the small number of units first, namely $N = 20$. As T increases, we obtain 11.55, 9.59 and 8.99 for $T = 20, 50, 100$, respectively. When we fix $T = 20$, we obtain 7.92 and 5.30 for $T = 50$ and $T = 100$. As in the case of *IPCP*, this means that the RMSE decreases more rapidly when N gets large compared to T.

The RMSE values of the heterogeneous estimators are generally lower than that of the homogeneous estimators. For instance when $N, T = 20$, the RMSE of *CCEP* equals 12 which is slightly higher than that of *CCEMG* (11.55). This slight difference between the two estimators exists for other sample sizes as well. Only exception to this finding is the *PCMGX* and its homogeneous counterpart *PCPX*. Among these, although very slighyly, the RMSEs of *PCPX* are lower for small and moderate values of T. For instance, when $T, N = 50$, their respective RMSEs equal 6.78 and 6.77. For larger samples once again heterogeneous estimator dominates the homogeneous one.

In Table 3, the case of high spatial dependence is considered. Compared to the previous

[3]For the *IPCP* estimator, we also tried to extract three common factors instead of two from the residuals in the iterations. This framework allows to obtain higher size adjusted power values. This is possibly due to the heterogeneity of the slope parameters such that extracting the common factors in the DGP for explanatory variables reduces the variability of the error term. Nevertheless, we report the results obtained using two common factors to be in line with the original literature. Moreover, we applied the bias correction proposed by Bai (2009). The results did not improve significantly. Thus, we reported those without bias correction in the tables.

Table 2: Case *1*: Low Spatial Dependence & Low Factor Dependence

	Heterogeneous							*Homogeneous*					
	Bias (×100)			RMSE (×100)				Bias (×100)			RMSE (×100)		
N \ T	20	50	100	20	50	100		20	50	100	20	50	100
CCEMG							*CCEP*						
20	-0.31	0.15	-0.20	11.55	9.59	8.99	20	-0.50	0.23	-0.19	12.00	10.25	9.46
50	0.34	-0.32	0.05	7.92	5.97	5.71	50	0.05	-0.22	0.00	7.76	6.21	5.89
100	-0.06	-0.18	0.09	5.30	4.31	4.00	100	-0.09	-0.25	0.10	5.38	4.49	4.16
CCEMGX							*CCEPX*						
20	-0.25	0.17	-0.20	11.69	9.63	9.00	20	-0.52	0.21	-0.18	12.02	10.25	9.45
50	0.31	-0.32	0.05	8.01	5.98	5.72	50	0.03	-0.22	0.01	7.79	6.20	5.90
100	-0.08	-0.19	0.09	5.31	4.32	4.00	100	-0.10	-0.25	0.09	5.38	4.49	4.16
IPCMG							*IPCP*						
20	-0.61	0.21	-0.20	12.42	9.69	9.00	20	-0.92	-0.50	-0.41	12.73	11.31	10.44
50	0.41	-0.26	0.05	7.82	5.90	5.67	50	-0.30	-0.58	-0.29	8.11	6.71	6.61
100	0.06	-0.16	0.09	5.06	4.21	3.96	100	-0.40	-0.45	-0.05	5.54	5.03	4.78
PCMGX							*PCPX*						
20	0.74	0.61	-0.13	14.78	11.13	9.40	20	0.49	0.49	-0.19	13.91	11.19	9.57
50	0.65	-0.09	0.14	10.15	6.78	5.99	50	0.39	-0.04	0.08	9.22	6.77	6.10
100	0.42	0.05	0.11	6.72	4.87	4.24	100	0.38	-0.04	0.12	6.40	4.87	4.35
PCMGX2S							*PCPX2S*						
20	-0.08	0.35	-0.19	12.78	9.84	8.94	20	-0.44	0.21	-0.25	12.58	10.36	9.26
50	0.34	-0.25	0.09	8.64	6.03	5.71	50	0.00	-0.25	0.00	8.15	6.25	5.89
100	-0.01	-0.16	0.11	5.58	4.32	3.99	100	-0.07	-0.24	0.11	5.53	4.52	4.13
	Size (×100)			Size Adj. Power (×100)				Size (×100)			Size Adj. Power (×100)		
CCEMG							*CCEP*						
20	6.70	7.60	7.90	13.35	16.90	17.10	20	7.25	7.70	8.05	12.25	16.70	15.80
50	6.05	6.35	6.40	22.45	34.70	40.45	50	5.90	5.95	6.45	25.00	33.40	38.25
100	5.00	6.55	4.80	46.70	60.05	71.15	100	5.15	6.05	5.50	45.30	58.65	66.15
CCEMGX							*CCEPX*						
20	5.10	6.10	6.95	13.50	16.05	16.45	20	5.60	6.65	6.85	12.50	16.30	15.55
50	6.05	5.45	5.90	22.70	35.00	39.75	50	5.15	5.30	5.95	25.75	33.05	38.25
100	4.80	6.55	4.60	46.50	59.60	71.10	100	5.00	5.65	5.30	44.90	57.85	66.60
IPCMG							*IPCP*						
20	6.40	6.40	6.75	11.10	16.30	17.45	20	8.05	10.05	9.50	10.25	13.35	14.50
50	5.75	5.45	5.75	25.75	35.65	40.80	50	7.10	6.45	7.70	23.40	28.15	29.30
100	5.00	5.70	4.60	50.65	63.35	72.55	100	5.80	7.80	6.95	42.25	45.40	55.20
PCMGX							*PCPX*						
20	6.15	6.05	6.20	11.10	16.25	15.60	20	5.45	6.05	6.20	12.10	15.45	16.25
50	5.35	5.35	5.20	17.85	27.85	37.70	50	5.30	4.75	5.05	18.70	30.10	37.50
100	5.45	5.65	4.65	33.30	53.05	66.10	100	4.90	5.60	5.35	37.95	49.95	61.85
PCMGX2S							*PCPX2S*						
20	5.90	6.35	6.50	11.00	16.60	17.20	20	6.80	6.85	7.20	10.70	16.55	16.10
50	5.60	5.20	6.20	21.85	34.55	40.55	50	5.55	5.40	6.40	22.70	34.30	36.95
100	4.95	6.05	4.55	42.70	59.15	70.70	100	5.30	5.60	5.25	43.35	57.35	67.20

Notes: The individual slope coefficients are generated as $\beta_{ij} = \beta_j + \eta_{ij}$, $\beta_j = 1$, $\eta_{ij} \sim \text{IID}\mathcal{N}(0, \sigma^2_{\eta_j})$ with $\sigma^2_{\eta_j} = 0.15$, $j = 1, 2$, which corresponds to the case of Low Heterogeneity. The results concern the average coefficient β_1 of the DGP in (19).

Table 3: Case *3*: High Spatial Dependence & Low Factor Dependence

	Heterogeneous							*Homogeneous*					
T	Bias (×100)			RMSE (×100)				Bias (×100)			RMSE (×100)		
N	20	50	100	20	50	100		20	50	100	20	50	100
CCEMG							*CCEP*						
20	-0.28	0.30	-0.26	14.24	10.41	9.25	20	-0.48	0.40	-0.25	13.65	10.88	9.70
50	0.33	-0.36	0.02	10.35	6.61	5.93	50	0.00	-0.25	-0.03	9.30	6.73	6.09
100	0.00	-0.15	0.08	6.79	4.80	4.19	100	-0.07	-0.22	0.07	6.37	4.90	4.34
CCEMGX							*CCEPX*						
20	-0.07	0.41	-0.21	15.23	10.62	9.28	20	-0.62	0.39	-0.21	13.72	10.80	9.67
50	0.30	-0.37	0.01	10.82	6.67	5.95	50	-0.01	-0.26	-0.02	9.26	6.72	6.09
100	-0.10	-0.18	0.07	6.88	4.80	4.19	100	-0.10	-0.23	0.06	6.35	4.87	4.32
IPCMG							*IPCP*						
20	-0.22	0.09	-0.34	12.48	10.20	9.44	20	-1.92	-1.51	-1.26	13.44	11.95	11.04
50	0.32	-0.19	0.07	8.90	6.47	6.02	50	-1.27	-1.45	-1.41	9.19	7.49	7.20
100	0.20	-0.02	0.26	6.01	4.70	4.26	100	-0.98	-0.92	-0.53	6.22	5.34	4.93
PCMGX							*PCPX*						
20	0.85	0.75	-0.08	19.59	12.62	9.91	20	0.44	0.59	-0.18	17.27	12.38	10.02
50	0.52	-0.11	0.11	12.84	7.54	6.26	50	0.23	-0.08	0.06	10.98	7.38	6.33
100	0.38	0.06	0.12	8.07	5.37	4.44	100	0.34	-0.03	0.11	7.26	5.30	4.52
PCMGX2S							*PCPX2S*						
20	0.06	0.40	-0.23	15.30	10.56	9.20	20	-0.30	0.30	-0.20	13.88	10.82	9.48
50	0.48	-0.21	0.06	11.18	6.62	5.89	50	-0.11	-0.21	0.00	9.50	6.71	6.05
100	0.10	-0.16	0.09	7.18	4.82	4.17	100	-0.04	-0.25	0.09	6.53	4.93	4.31
	Size (×100)			Size Adj. Power (×100)				Size (×100)			Size Adj. Power (×100)		
CCEMG							*CCEP*						
20	6.30	7.85	7.80	11.65	15.05	16.80	20	6.90	7.00	8.10	10.80	15.85	16.30
50	5.75	5.95	6.25	16.30	28.25	37.75	50	5.75	5.15	6.10	18.15	30.10	36.65
100	5.05	6.60	4.55	31.25	49.95	67.70	100	5.25	6.35	5.25	34.20	48.40	62.10
CCEMGX							*CCEPX*						
20	3.05	5.05	5.70	10.45	14.65	17.20	20	3.35	5.00	6.10	10.45	14.95	15.85
50	4.85	4.70	5.50	15.10	29.30	37.30	50	3.70	4.40	4.85	18.30	28.90	38.00
100	4.10	6.15	4.30	31.70	50.50	66.90	100	4.30	5.85	5.20	35.00	49.05	61.85
IPCMG							*IPCP*						
20	6.55	6.45	6.45	10.20	15.25	16.40	20	8.65	7.75	8.40	9.10	12.60	12.50
50	6.30	6.00	5.00	21.10	30.15	38.60	50	7.35	7.10	6.85	16.25	20.70	21.45
100	4.95	6.35	5.05	41.25	53.55	65.65	100	6.60	7.50	6.05	30.85	39.55	49.05
PCMGX							*PCPX*						
20	6.10	6.45	6.10	8.40	12.55	14.95	20	5.45	6.45	5.95	8.85	12.85	14.60
50	5.90	5.45	5.15	12.25	23.80	35.55	50	5.90	5.05	5.15	12.40	25.85	34.20
100	4.95	5.65	4.85	26.00	44.80	61.10	100	4.70	5.75	5.10	29.60	43.30	59.65
PCMGX2S							*PCPX2S*						
20	5.50	7.15	6.90	10.20	13.70	16.30	20	5.85	6.45	6.85	9.50	14.80	15.65
50	5.00	5.70	5.50	16.80	29.80	38.85	50	5.45	5.30	6.00	17.20	30.30	35.40
100	5.25	6.35	4.75	28.90	51.80	66.60	100	4.85	6.35	5.25	33.80	48.30	62.85

Notes: See notes of Table 2.

case, the first difference we notice concerns RMSE values which are higher for all estimators. This is expected as the spatial dependence affects only the disturbances in (1). For instance when $N, T = 20$ the RMSE of *CCEMG* is 14.24 which is almost 25% higher than the respective value in the previous case. Similar conclusions apply to other estimators. Second important finding is that in this case heterogeneous estimators which deal with common factors using *PC* perform better than those using the *CCE* methodology. In terms of precision, the best performing estimator is *IPCMG* in small to moderate samples. For instance, when $N, T = 50$ its RMSE equals 6.47 while that of the second best performer, *CCEMG*, is 6.61. Third finding is on the comparison of the heterogeneous and the homogeneous estimators. In this case of higher spatial dependence, the homogeneous estimators perform better than the heterogeneous one in the smallest samples in general. For instance if we focus on the *CCEMG* and *CCEP*, we see that their RMSEs are 14.24 and 13.65, respectively. This finding applies to other estimators as well, with the exception of *IPCMG* and *IPCP*. Overall the best performer is still *IPCMG* and for moderate to larger samples the dominance of heterogeneous estimators continue to be valid.

Overall, a general feature that emerges is that the consistent heterogeneous estimators perform better than their homogeneous counterparts, even if the results are more contrasted when the degree of spatial dependence is high. Also we can conclude that the estimators using the *PC* methodology are more robust to high degrees of spatial dependence.

We also simulated the model with a higher degree of heterogeneity in the slope coefficients, namely with $\sigma^2_{\eta_j} = 0.3$, $j = 1, 2$. In this case, all consistent heterogeneous estimators are superior to homogeneous ones in all cases except when $T = 20$. Here again, *IPCMG* turns out to be a better choice in terms of bias and RMSE than *CCEMG* in the case of high spatial dependence. The results for this case are available upon request.

Tables 4 and 5 report the results on the partially heterogeneous estimators *C-Lasso* and *GFE* estimators in the case of low heterogeneity. The two estimators are time consuming compared to the others. For this reason, we perform for each experiment 1,000 replications instead of 2,000. Two values are considered for the number of groups: $K = 2$ and $K = 3$. The statistics of these estimators are computed by taking the averages of the values over the groups.

As can be seen in Table 4 which concerns the case of low spatial dependence, *C-Lasso* is more biased compared to the other estimators controlling for the unobserved common factors discussed above. Whereas *GFE* shows very low biases. For instance, when $N, T = 20$ and $K = 2$, the bias of *C-Lasso* is 1.62. Furthermore, the estimator does not seem to improve with increases in the sample size or the number of groups: the corresponding value for $N, T = 100$ is 1.74 and for $N, T = 20$ and $K = 3$ is 2.16. In terms of RMSE, however, *C-Lasso* dominates *GFE* for small values of N. For $N, T = 20$ and $K = 2$ the RMSE of *C-Lasso* is as small as

Table 4: Partially Heterogeneous Estimators

Case *1*: Low Spatial Dependence & Low Factor Dependence

	$K = 2$						$K = 3$					
T / N	Bias (×100)			RMSE (×100)			Bias (×100)			RMSE (×100)		
	20	50	100	20	50	100	20	50	100	20	50	100
GFE												
20	0.15	0.31	-0.16	12.33	10.68	10.18	0.11	0.55	-0.11	12.18	10.34	10.00
50	-0.28	-0.10	-0.38	7.27	6.72	6.64	-0.19	-0.19	-0.39	7.19	6.59	6.29
100	0.20	-0.17	-0.27	5.44	4.76	4.75	0.10	-0.11	-0.24	5.34	4.63	4.64
C-Lasso												
20	1.62	2.27	1.71	11.17	10.45	10.05	2.16	3.01	2.53	11.24	10.69	10.26
50	1.84	2.03	1.75	7.27	6.89	6.41	2.90	3.05	2.68	7.52	7.22	6.80
100	2.15	1.45	1.74	5.88	4.91	4.78	3.10	2.70	2.70	6.04	5.28	5.08
	Size (×100)			Size Adj. Power (×100)			Size (×100)			Size Adj. Power (×100)		
GFE												
20	30.50	33.60	38.60	11.40	10.70	9.60	38.40	40.80	44.10	9.00	10.10	9.70
50	22.50	29.40	33.50	22.40	20.10	19.20	27.40	32.40	38.10	18.00	18.10	15.00
100	21.20	29.80	30.80	37.00	36.60	36.90	25.40	33.00	34.10	33.40	29.50	36.80
C-Lasso												
20	18.80	20.80	21.50	11.50	12.30	13.70	31.40	34.10	38.70	11.00	11.20	9.40
50	13.20	13.20	12.90	22.80	28.40	25.10	22.60	27.20	27.30	16.60	18.20	17.40
100	12.60	11.50	13.30	36.20	38.80	35.00	23.50	24.50	25.20	31.00	28.70	26.50

Notes: See notes of Table 2.

11.17 whereas the corresponding value for *GFE* is 12.33. It is important to note that in this case the RMSE of *C-Lasso* is lower than that of the best performing heterogeneous estimator, namely *CCEMG* for which the corresponding number is 11.55. These results are confirmed with the case of high spatial dependence for which the results are given in Table 5. In this case biases are reinforced. For *GFE*, the absolute biases are similar to the heterogeneous and homogeneous estimators discussed above. However the bias and RMSE of the estimator remain higher than those of the well performing heterogeneous estimators.

3.2.2 Size and size adjusted power results

The size and size adjusted power constitute the second part of Tables 2-5. The nominal size is set to 5% and the size is computed using a two-sided test under the null hypothesis $H_0 : \beta_1 = 1$. The size adjusted power is investigated by testing $H_1 : \beta_1 = 0.9$.

The empirical sizes are very close to the nominal size for all values of N and T for most of the estimators taking into account the unobserved common factors. Among these estimators, it seems that the tests based on *IPCP* over-reject the null hypothesis. For instance, as seen

Table 5: Partially Heterogeneous Estimators

Case *3*: High Spatial Dependence & Low Factor Dependence

	K = 2						K = 3					
$N \backslash T$	Bias (×100)			RMSE (×100)			Bias (×100)			RMSE (×100)		
	20	50	100	20	50	100	20	50	100	20	50	100
GFE												
20	0.72	0.23	-0.15	13.40	11.34	10.79	0.39	0.36	-0.24	13.36	11.44	10.51
50	-0.39	-0.21	-0.48	7.90	6.95	6.81	-0.24	-0.31	-0.33	7.79	6.77	6.52
100	0.18	-0.20	-0.29	5.86	4.90	4.86	0.12	-0.18	-0.32	5.73	4.76	4.74
C-Lasso												
20	2.33	2.14	0.60	13.29	10.75	9.82	2.04	2.96	1.59	13.36	10.86	9.97
50	2.20	1.58	1.38	8.23	6.98	6.54	2.92	2.68	2.43	8.17	7.26	6.72
100	2.58	1.54	1.40	6.77	5.36	4.93	3.20	2.80	2.53	6.76	5.57	5.32
	Size (×100)			Size Adj. Power (×100)			Size (×100)			Size Adj. Power (×100)		
GFE												
20	28.20	29.10	32.50	9.40	12.90	10.90	33.60	36.20	36.40	10.70	11.40	8.80
50	18.80	25.50	29.00	19.00	20.20	23.90	20.30	26.40	28.80	20.60	24.10	18.80
100	16.00	25.30	27.60	37.00	37.30	39.40	18.70	25.20	27.40	29.30	32.40	33.40
C-Lasso												
20	19.50	19.70	17.70	10.60	11.50	9.70	30.90	35.80	33.60	9.50	10.70	10.20
50	15.40	11.70	12.50	20.90	21.90	19.20	20.40	23.50	27.90	15.80	17.30	16.50
100	13.30	13.70	14.10	28.20	38.20	38.30	18.70	24.80	24.30	27.20	27.40	26.50

Notes: See notes of Table 2.

in Table 2, its size is 8.05% when $N, T = 20$. Although it performs better than *IPCP*, tests based on *CCEP* are over-sized as well in small samples. The corresponding number for this estimator is 7.25%. The conclusions are similar in the case of high spatial dependence for which the results can be seen in Table 3. Once more, most of the estimators are performing well in small samples except *IPCP* which over-rejects the null hypothesis.

The tests based on *C-Lasso* and *GFE* are hugely over-sized. This is an expected result as these estimators wrongly assume a grouped factor structure instead of taking into account the heterogeneity in unit-specific coefficients.

Figure 1 reports the RMSE and size adjusted power of these estimators considering the alternative hypothesis $H_1 : \beta_1 = 0.7$. When $N = 50$ as T increases, it is seen that the RMSEs tend to 0 whatever the estimator considered, whereas the size adjusted power values tend to 100. Compared to the tables we discussed above, this figure adds additional information because we also consider $T < 20$. It appears that for $T < 15$, the homogeneous *CCEP* and *IPCP* estimators have lower RMSE and higher size adjusted power compared to their homogeneous counterparts *CCEMG* and *IPCMG*. As soon as $T \geq 15$, *CCEMG* and *IPCMG* tend to dominate all the other estimators, especially the homogeneous ones.

4 Empirical application

In this section, we present an empirical application which demonstrates the estimation procedures introduced in the previous sections. The main objective is to investigate the role of heterogeneity and CD on the estimates of the effect of fiscal decentralization on government size in OECD countries.

4.1 Empirical model and the data

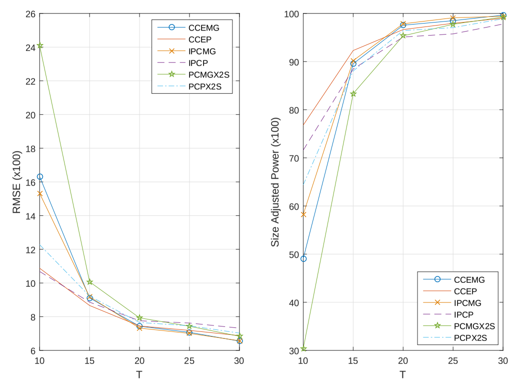
Our empirical model is based on that of Jin and Zou (2002) where the dependent variable, which we denote $size$, is one of the following measures of size of government: aggregate government size, national government size, and subnational government size. They are defined as the total tax revenue, tax revenue in central level, and tax revenue in state and local levels combined, respectively. All these variables are measured as shares in GDP. The main focus is on the effect on aggregate government, however, following Jin and Zou we also report some results after disaggregating the government size. The main argument behind the relation between the level of decentralization and government size is that decentralization would lead to fiscal discipline such that, tax autonomy or local power on public spending decisions encourage local governments' fiscal responsibility (see Martinez-Vazquez et al., 2017, for a detailed survey of the literature).

We model government size as

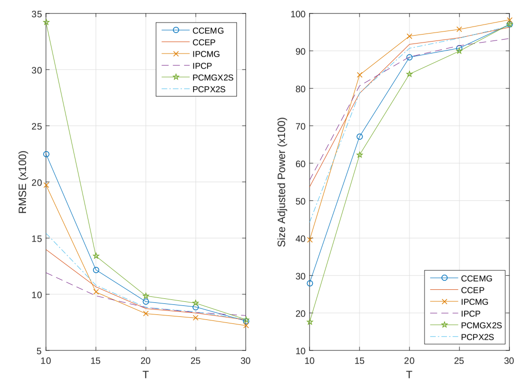
$$size_{it}^{(l)} = \alpha_{i1} + \alpha_{i2}t + \beta_{i1}dec_{it} + \beta_{i2}inf_{it} + \beta_{i3}gr_{it} + \beta_{i4}urb_{it} + \beta_{i5}open_{it} + e_{it}, \quad (22)$$

where $l = 1, 2, 3$ denotes aggregate government size, national government size, and subnational government size, respectively, dec is the variable of interest denoting the degree of fiscal decentralization measured as the ratio of the sum of tax revenue in state and local levels to total tax revenue. Control variables are inf, the annual consumer price inflation, gr, the annual GDP growth, urb, the urban population as a share of total population, and $open$, the openness index defined as the ratio of the sum of imports and exports in GDP. In addition to these control variables Jin and Zou use two time invariant variables (borrowing constraints and federal vs. unitary indicator), an indicator for elected versus non-elected subnational governments and an indicator for political central banks. As these variables rarely have statistically significant effects on government size and because of their (nearly or completely) time invariant nature, we do not consider them in our model. On the contrary, as will be discussed below, the variables of interest follow some heterogeneous trends in our case. Hence, we added a trend in the model as an observed common factor.

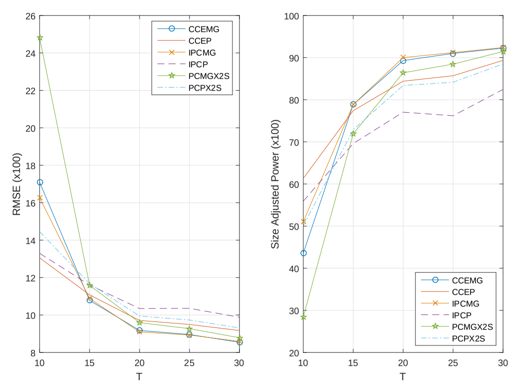
The data on fiscal variables $size$ and dec come from the OECD Fiscal Decentralisation Database. We collected the data on other variables from the World Development Indicators database of the World Bank. Our final data set contains 10 variables on 22 OECD countries between the years 1973 and 2017. The countries considered are AUS, AUT, BEL, CAN, CHE, DEU, DNK, ESP, FIN, FRA, GBR, GRC, IRL, ITA, JPN, LUX, NLD, NOR, NZL, PRT, SWE and USA. In addition to the variables in the model above, we have data on inv, investment ratio, and pop, annual population growth. In the original data set there were occasional missing values. In GRC, fiscal variables for the years 1973 and 1974, and in NZL pop were missing for 1991. Instead of dropping the countries from the data set, we extrapolated these values using a linear trend.



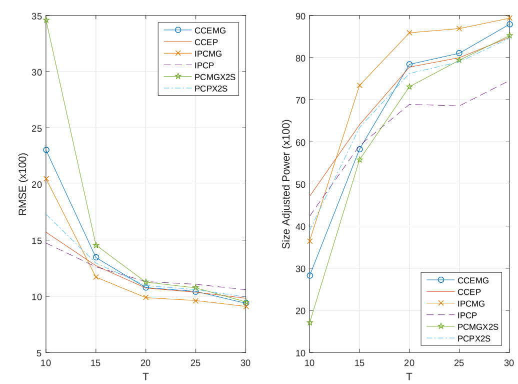
(a) Low heterogeneity, Case *1*



(b) Low heterogeneity, Case *3*



(c) High heterogeneity, Case *1*



(d) High heterogeneity, Case *3*

Figure 1: RMSE and size adjusted power of various estimators for $N = 50$

4.2 Preliminary analysis

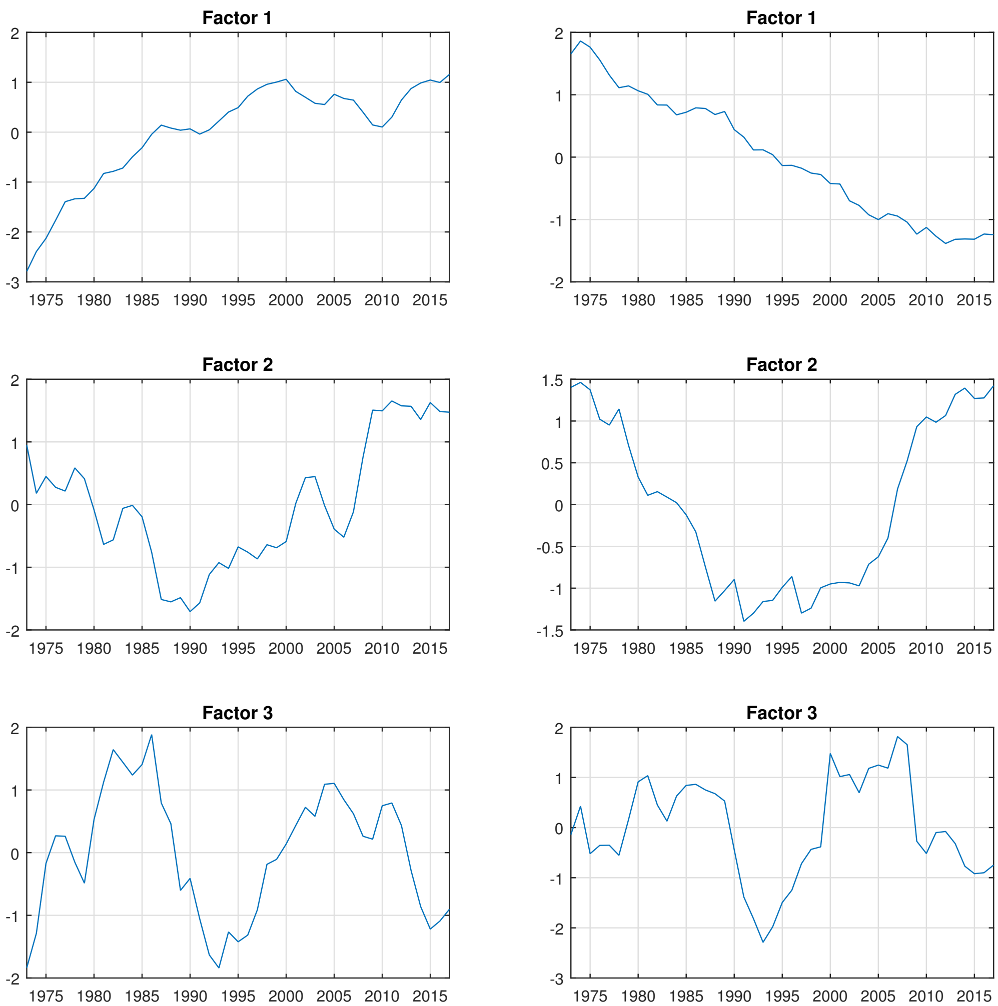
In this subsection, we report a preliminary analysis of the variables in our data set. The descriptive statistics of the variables are reported in Table 6.

Table 6: Descriptive Statistics

Variable	*Notation*	*Mean*	*Std. Dev.*	*Min.*	*Max.*
Fiscal Variables					
Aggregate Government Size	$size^{(1)}$	0.343	0.070	0.166	0.495
National Government Size	$size^{(2)}$	0.283	0.073	0.108	0.417
Subnational Government Size	$size^{(3)}$	0.060	0.046	0.000	0.167
Decentralization	dec	0.175	0.135	0.000	0.502
Control Variables					
Consumer Inflation	inf	0.048	0.051	-0.045	0.310
GDP Growth	gr	0.024	0.026	-0.091	0.252
Urbanisation	urb	0.761	0.104	0.400	0.980
Openness	$open$	0.741	0.514	0.131	4.084
Additional Variables					
Investment Ratio	inv	0.233	0.038	0.115	0.387
Population Growth	pop	0.006	0.005	-0.019	0.038

We applied two tests of CD to each variable in the data set. These are the LM test of Breusch and Pagan (1980) and a modified version of this test, the Modified BP Test. The first test follows a χ^2 distribution with $N(N-1)/2$ degrees of freedom as T goes to infinity for fixed N under the null hypothesis of no CD, and the second one is distributed as a standard normal for large T and N (see Pesaran, 2015, for details). The results are reported in Table 7 where the statistics can be found for each original variable and their defactored versions. As can be seen, for all variables in the data set the hypothesis of no CD can be rejected in any significance level. Moreover, even when we remove the common factors which are discussed in what follows, the no CD hypothesis can be rejected. This shows that allowing for common factors is not sufficient to model the CD in these variables.

Following these CD test results, we estimated the common factors in fiscal variables aggregate government size and decentralization. We extracted the first three PCs from these series using the estimation methodology in Bai and Ng (2002) and Bai (2003). These time series are shown in Figure 2 and their loadings are reported in Table 8. As is seen in the first graphs



(a) Aggregate Government Size

(b) Decentralization

Figure 2: Common Factors in Fiscal Variables

Table 7: CD Test Results

Variable	$size^{(1)}$	$size^{(2)}$	$size^{(3)}$	dec	inf	gr	urb	$open$	inv	pop
				Panel a: Original Data						
Breusch-Pagan LM Test	2926.20	2543.30	3448.50	2845.50	6772.10	2253.20	7990.50	6005.80	2432.70	1546.70
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Modified BP Test	125.40	107.60	149.70	121.60	304.30	94.10	361.00	268.70	102.40	61.20
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
				Panel b: Defactored Data						
Breusch-Pagan LM Test	87.82	89.52	83.25	85.11	84.34	72.43	167.22	130.96	111.08	86.29
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Modified BP Test	30.10	30.90	28.00	28.90	28.50	23.00	67.00	50.20	40.90	29.40
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: For each variable x_{it}, the Breusch-Pagan LM Test statistics are computed as $CD_{BP} = T\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\widehat{\kappa}_{ij}^2$ where $\widehat{\kappa}_{ij}$ is the correlation coefficient between x_{it} and x_{jt}. Under the null of no CD, the asymptotic distribution of the test statistic is χ_q^2 with $q = N(N-1)/2$. The Modified BP Test statistics are computed as $CD_M = [N(N-1)]^{-1/2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(T\widehat{\kappa}_{ij}^2 - 1)$ which is distributed as $N(0,1)$ under the null of no CD. p-values are in parentheses. The test statistics given in *Panel b* are computed after removing country fixed effects and the unobserved common factors estimated using PC methods. For each variable the number of common factors are chosen using the information criterion IC_{p_1} of Bai and Ng (2002).

of each panel, both variables have a strong trend component which is roughly linear in almost the entire period under consideration. For government size, this trend breaks in 2000 and gets flatter, whereas for decentralization such a change occurs after 2010. The loadings of the first common factor of the government size series given in the first column of Table 8 show that, for the majority of countries government size tended to increase over this period as the loadings are positive. Three exceptions are GBR, IRL and NLD which have negative loadings. For the countries where the loadings of the second factor is negative, this trend is compensated until 1990 as the second common factor has a downward trend until this year. Last common factor of this series shows a more irregular behavior compared to the others. Overall, we can see that the first common factor captures the long run movements whereas the last one shows the higher frequency ones.

The results are very similar for decentralization. The loadings of the first common factor are positive for roughly half of the countries for this variable. As the first common factor is downward sloping, the countries with a positive factor loading experienced a decreasing level of decentralization over the period. However, as in the previous case, the second factor compensates this trend for some of the countries. Once again, the third common factor captures the higher frequency component.

As a conclusion to this subsection, we can say that the fiscal variables show trends and they are strongly cross-sectionally correlated in the sample in hand. As a result, we can expect to be able to better model them using methods which take into account common factors.

Table 8: Factor Loadings for Fiscal Variables

	Aggregate Government Size			*Decentralization*		
Country	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 1*	*Factor 2*	*Factor 3*
AUS	0.76	-0.33	0.25	0.03	-0.36	-0.86
AUT	0.91	0.05	-0.04	0.87	-0.03	-0.10
BEL	0.87	0.16	0.21	-0.86	0.27	-0.06
CAN	0.57	-0.61	-0.30	-0.81	0.28	-0.19
CHE	0.87	0.35	0.12	0.91	0.17	-0.03
DEU	0.24	0.29	-0.29	0.46	0.65	0.18
DNK	0.90	-0.11	-0.07	0.25	-0.75	-0.09
ESP	0.94	-0.07	-0.03	-0.96	0.05	0.07
FIN	0.85	-0.13	-0.29	-0.09	0.72	-0.26
FRA	0.94	0.09	0.03	-0.91	0.01	-0.19
GBR	-0.17	-0.02	0.74	0.80	0.31	0.28
GRC	0.85	0.44	-0.04	0.08	0.70	0.03
IRL	-0.03	-0.81	0.21	0.57	0.53	-0.14
ITA	0.92	0.16	-0.07	-0.90	0.12	0.21
JPN	0.73	0.07	-0.05	0.12	-0.43	0.34
LUX	0.73	0.27	0.45	0.86	-0.24	-0.03
NLD	-0.22	-0.56	-0.13	-0.91	-0.15	0.13
NOR	0.17	-0.22	0.82	0.86	-0.09	-0.40
NZL	0.48	-0.68	-0.10	0.29	0.82	-0.05
PRT	0.92	0.25	-0.04	-0.88	-0.25	-0.01
SWE	0.55	-0.66	0.18	-0.72	0.40	-0.21
USA	0.46	-0.42	-0.30	-0.43	0.28	-0.17

4.3 Results

The results of estimation of the model (22) by heterogeneous and homogeneous estimators using aggregate government size, $size^{(1)}$, as the dependent variable are reported in Table 9. Whereas the results for the partially heterogeneous estimators are reported in Table 10.

Following Bresson et al. (2011), we compare the models and estimation methods by their predictive performance. Hence, each table reports in-sample RMSE values averaged over countries. To calculate these RMSE values, we follow the Auxiliary Variables Approach by Akgun et al. (2020). This method, originally developed for out-of-sample forecasting, is based on the estimation of the unobserved common factors using the explanatory variables of the model as well as some exogenous variables which do not enter into the estimation equation. The forecast methodology consists of four steps. In the first step, the slope parameters are estimated by any estimator robust to unobserved common factors. In the second step, the unobserved common factors are estimated by PCA from a number of auxiliary variables which can include the explanatory variables of the model. In the third step, the residuals from the first step are regressed on these estimated common factors to estimate the factor loadings. In the final step, the forecasts are computed using the estimated slope coefficients, common factors and factor loadings.

In our application, we use investment ratio and annual population growth as additional variables to estimate the unobserved common factors. These variables are used also for the estimators *Ind. CCEX*, *CCEMGX* and *CCEPX*. For the calculation of RMSE, we use the unit-specific estimates in the case that the related estimator assumes heterogeneity. In each table concerning the estimation of the model, we report the estimates of the parameter $\beta_1 = \mathrm{E}(\beta_{i1})$, whereas the prediction RMSEs are computed using the unit-specific estimates for the heterogeneous estimators. Standard errors reported are computed using the variance formulas in Appendix.

In Table 9, the estimates of the parameter of decentralization are always negative, which is in line with the existing literature. It varies between -0.08 for *CCEMG* and -0.37 for *PCMGX*. The estimator which gives the smallest average RMSE is *Ind. CCEX*. The mean group estimator based on this estimator, *CCEMGX*, provides an estimate equal to -0.24. This implies that, on average, one percentage point increase in tax decentralization diminishes the government size by 0.24 percentage points. When we compare this estimate and its respective RMSE with the ones provided by homogeneous estimators we see that the latter ones always have larger RMSEs and point out to a smaller effect in absolute terms. Based on in-sample prediction precision measured by average RMSEs, in this analysis our preferred estimate is *Ind. CCEX* and the mean group estimator based on it, *CCEMGX*.

In Figure 3 the country-specific estimates of the marginal effect of decentralization are reported using the preferred estimator *Ind. CCEX* together with their 95% confidence bands

Table 9: Estimation Results for Homogeneous and Heterogeneous Estimators

	(1) *CCEMG*	(2) *CCEP*	(3) *CCEMGX*	(4) *CCEPX*	(5) *PCMGX*	(6) *PCPX*	(7) *IPCMG*	(8) *IPCP*	(9) *PCP2SX*	(10) *PCMGX2S*
Decentralisation	-0.08	-0.14	-0.24	-0.12	-0.15	-0.14	-0.37	-0.22	-0.33	-0.15
	(0.234)	(0.078)	(0.209)	(0.104)	(0.235)	(0.083)	(0.151)	(0.127)	(0.156)	(0.112)
Consumer Inflation	0.10	0.01	0.06	-0.01	-0.02	-0.03	0.06	0.03	0.06	0.05
	(0.045)	(0.048)	(0.046)	(0.047)	(0.036)	(0.034)	(0.048)	(0.04)	(0.038)	(0.023)
GDP Growth	-0.09	-0.11	-0.12	-0.11	-0.08	-0.12	-0.15	-0.16	-0.16	-0.11
	(0.044)	(0.032)	(0.051)	(0.033)	(0.031)	(0.044)	(0.039)	(0.035)	(0.033)	(0.033)
Urbanisation	-0.16	-0.20	-0.06	-0.16	-0.44	0.37	0.58	0.33	0.44	0.22
	(0.234)	(0.233)	(0.422)	(0.382)	(0.453)	(0.444)	(0.429)	(0.414)	(0.313)	(0.36)
Openness	-0.01	-0.04	0.03	-0.04	-0.04	0.01	0.04	-0.04	0.03	-0.04
	(0.025)	(0.02)	(0.031)	(0.048)	(0.027)	(0.027)	(0.044)	(0.031)	(0.038)	(0.026)
Average RMSE (x100)	0.831	0.946	0.803	0.951	0.968	0.958	1.166	0.957	1.111	0.955

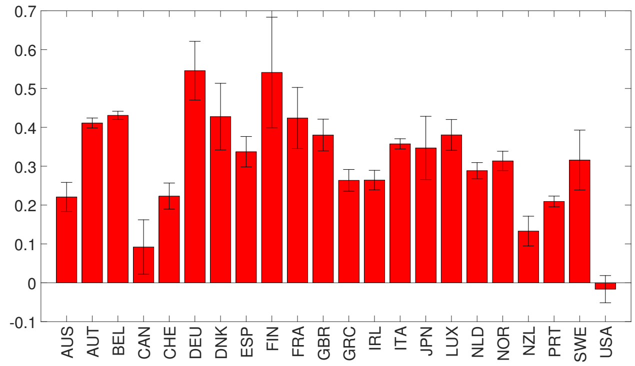
Notes: Standard errors in parentheses are computed using the formulas in Appendix. “Average RMSE” denotes in-sample RMSE values averaged over countries.

computed using the Newey and West (1987) estimator in (A.1). This figure also decomposes the effect of decentralization into the size of different levels of government. First of all, in all panels we see that a considerable amount of heterogeneity exists among countries in terms of the effect of decentralization. Shown in part (a), all country-specific coefficient estimates are positive, except USA, and they range between roughly zero for USA and 0.6 for DEU and FIN. The average of these estimates is 0.31 which is much larger than the estimate of Jin and Zou, 0.12. The effect on national government size, reported in part (b) is the source of the negative effect of decentralization on total government size. Here, there are two countries which have estimated coefficients much larger than unity in absolute terms. For some countries we also see positive effects although they are statistically insignificant. This is similar for the effect on total government size shown in part (c). Once again, we see countries with very large negative effects, some with small negative effects and some with positive effects. In what follows we report the results on partially heterogeneous estimators to see if these groups can be identified by *GFE* and *C-Lasso*.

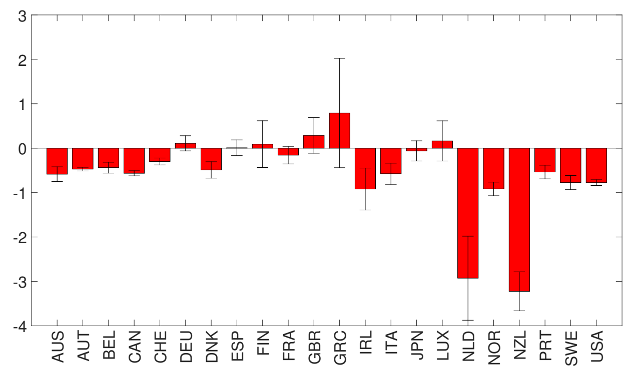
The estimation results for each group selected by *GFE* and *C-Lasso* are reported in Table 10. As we have a relatively small number of panel units, we chose small numbers of groups $K = 2$ and $K = 3$. These are also the numbers of groups we used in the Monte Carlo simulations.

We see from the estimation of the model that the RMSE values are smaller for $K = 3$ using any estimator. The smallest average RMSE is reached by *C-Lasso* in this sample. The estimator identified the countries with negative effects, some with smaller negative effects and some with positive effects. The first group of countries has a coefficient of -0.66 which is more than twice as big as the average effect estimated by the homogeneous estimators. This group has seven countries which are reported in Table 11. All countries in this group have the same sign as estimated by the unit-specific estimators except GRC which had a positive but insignificant coefficient when estimated with *Ind. CCEX*. The second group also contains seven countries and their estimated coefficient is 0.53. All of these countries had positive coefficients estimated by *Ind. CCEX* except IRL. The last group is the countries with moderate negative effect which is estimated as -0.20. This group has eight countries for half of which the coefficients are statistically significant when estimated by *Ind. CCEX*. Overall, we can see that *C-Lasso* groups the countries reasonably with respect to the estimates provided by the unit-specific estimator. The average estimated coefficient over groups equals -0.11 which is smaller in absolute terms compared to the estimates provided by the heterogeneous estimator *CCEMGX*.

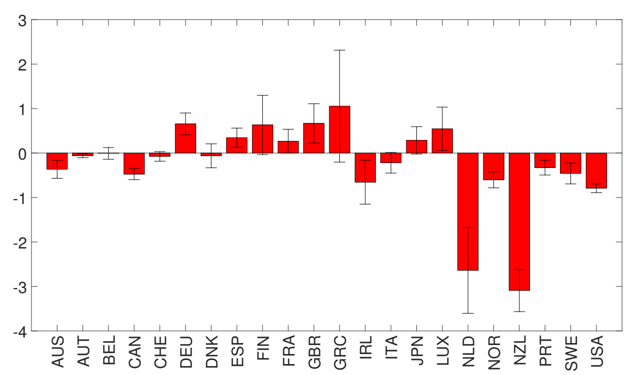
When we compare the partially heterogeneous estimators with the fully heterogeneous ones, we see that the in-sample predictions of the latter are better compared to those of the former. The in-sample RMSE of *C-Lasso* is 0.854 with $K = 3$. The corresponding value for the *Ind.*



(a) Effect on Subnational Government Size



(b) Effect on National Government Size



(c) Effect on Aggregate Government Size

Figure 3: Estimated Marginal Effects of Decentralization on Government Size and 95% Confidence Bands

CCEX is 0.803 as mentioned above. To conclude, we see that the optimal heterogeneous estimators outperform the homogeneous and partially heterogeneous ones.

5 Conclusion

In this paper, we evaluated the performance of alternative homogeneous, heterogeneous and partially heterogeneous panel data estimators. The comparison was performed by Monte Carlo simulations as well as real data using several models with cross-sectional dependence modeled by spatial error dependence and common factors. These specifications allowed us to compare and contrast the case of weak cross-sectional dependence with the case of strong cross-sectional dependence. We revisited the recent literature on alternative estimation procedures accounting for the nature and the degree of cross-sectional dependence. We compare the performance of twelve estimators using an extensive Monte Carlo exercise under low and high levels of heterogeneity as well as weak and strong cross-sectional dependence generated by spatial error dependence and unobserved common factors structures, respectively.

Our main results can be summarized as follows: (i) Even for small T and N, heterogeneous estimators, especially *CCEMG* of Pesaran (2006) and *IPCMG* Song (2013), outperform their homogeneous counterparts. However, most of the estimators considered show desirable small sample properties. (ii) The dominance of the heterogeneous estimators are more pronounced for the case of high heterogeneity, as expected. This main result holds for different degrees of spatial dependence and factor dependence. (iii) The main difference on the performance of the two methods of dealing with unobserved common factors, namely common correlated effects (*CCE*) and principal components (*PC*), occurs when we change from low to high spatial dependence whereas changing from low to high factor dependence does not make a big difference in their comparative performance. The estimators based on *PC* methods are found to be more robust to spatial dependence compared to those based on *CCE*. This result shows that both methodology work equally good against unobserved factors. (iv) Among the two partially heterogeneous estimators assuming a grouped structure of heterogeneity, the *GFE* of Bonhomme and Manresa (2015a) performs well in terms of bias and RMSE whereas *C-Lasso* of Su et al. (2016a) based on *CCE* transformation gives less satisfactory results. The performance of *GFE* improves as we increase the number of groups assumed in the estimation.

Finally, we applied the methods to the estimation of the effect of fiscal decentralization on the size of government using data from 22 OECD countries over the period 1973-2017. These findings are confirmed using real data. We documented that a considerable amount of heterogeneity exists among countries with respect to the estimated effect of fiscal decentralization on the size of government which is an important finding from a policy perspective. Though, our findings are in line with the previous literature in terms of the estimated average effect over

Table 10: Estimation Results for Partially Heterogeneous Estimators

	GFE					*C-Lasso*				
	$K = 2$		$K = 3$			$K = 2$		$K = 3$		
	Group 1	*Group 2*	*Group 1*	*Group 2*	*Group 3*	*Group 1*	*Group 2*	*Group 1*	*Group 2*	*Group 3*
Decentralization	0.16	-0.07	0.05	0.44	-0.36	-0.43	0.38	-0.66	0.53	-0.20
	(0.105)	(0.054)	(0.068)	(0.104)	(0.099)	(0.087)	(0.251)	(0.077)	(0.076)	(0.049)
Consumer Inflation	-0.25	0.00	-0.26	-0.08	-0.05	0.02	0.11	-0.03	0.12	0.15
	(0.072)	(0.106)	(0.213)	(0.091)	(0.057)	(0.045)	(0.078)	(0.057)	(0.086)	(0.058)
GDP Growth	-0.18	-0.16	-0.14	-0.12	-0.19	-0.06	-0.17	-0.08	-0.20	-0.14
	(0.054)	(0.083)	(0.134)	(0.061)	(0.052)	(0.034)	(0.033)	(0.029)	(0.04)	(0.079)
Urbanisation	-0.03	-0.11	0.04	0.43	-0.26	-0.33	-0.01	-0.20	0.00	-0.57
	(0.105)	(0.116)	(0.3)	(0.092)	(0.052)	(0.342)	(0.271)	(0.333)	(0.326)	(0.511)
Openness	-0.13	0.02	-0.17	0.01	0.07	0.03	-0.04	0.05	-0.03	0.00
	(0.011)	(0.01)	(0.04)	(0.012)	(0.021)	(0.02)	(0.019)	(0.031)	(0.021)	(0.02)
In-Sample RMSE (x100)	0.962		0.934			0.863		0.854		

Notes: Standard errors in parentheses are computed using the formulas in Appendix. "Average RMSE" denotes in-sample RMSE values averaged over countries.

Table 11: Country Groups Selected by *C-Lasso*

Groups	Coefficients	Countries							
Group 1	-0.66	CHE	GRC	NLD	NOR	NZL	SWE	USA	
Group 2	0.53	DEU	ESP	FIN	FRA	GBR	IRL	LUX	
Group 3	-0.20	AUS	AUT	BEL	CAN	DNK	ITA	JPN	PRT

countries. The result show that, heterogeneous estimators outperform the homogeneous and heterogeneous ones in terms of in-sample prediction precision.

The main findings in this paper suggest some interesting further developments. First, in our Monte Carlo simulations we assumed that the number of factors is known. If this is not the case, the number should be estimated and investigated from the sample. Second, *CCE* estimators require a rank condition to hold. As Westerlund and Urbain (2013) show *CCE* methods can even turn inconsistent in this case, it would be interesting to study the consequences of rank deficiency. Third, we evaluated the performance of the estimators assuming grouped structure of heterogeneity. It will be interesting to extend our analysis to the case of shrinkage estimators that can be considered as a hybrid solution between homogeneous and heterogeneous estimators (see Maddala et al., 1994, 1997; Hsiao et al., 1999). This is an ongoing research agenda.

Acknowledgements

We wish to thank participants in the New York Camp Econometrics XIII (Syracuse University, United States, 6-9 April 2018), the 17th Spatial Econometrics and Statistics Workshop (University of Burgundy, France, 24-25 May 2018), the 24th International Panel Data Conference (Sogang University, Korea, 19-20 June 2018), the 20th OxMetrics User Conference (Cass Business School, United Kingdom, 10-11 September 2018), and Badi Baltagi, Neil R. Ericsson, Wang Fa, Lynda Khalaf for useful comments and suggestions. The usual disclaimer applies. This paper is a part of Oguzhan Akgun's Ph.D. research project, and it was developed while Giovanni Urga was visiting professor at University Paris II Panthéon-Assas - CRED (France) during the academic years 2016-2017, 2017-2018 and 2018-2019, and Alain Pirotte was visiting professor at Bergamo University during the academic years 2016-2017, 2017-2018 and 2018-2019 under the scheme "Progetto ITALY® - Azione 2: Grants for Visiting Professor and Scholar": we wish to thank both institutions for financial support. Financial support is also acknowledged from the Centre for Econometric Analysis (CEA) of Cass Business School for several visits of Alain Pirotte and Oguzhan Akgun at CEA.

References

Akgun, O., Pirotte, A., and Urga, G. (2020). Forecasting using heterogeneous panels with cross-sectional dependence. *International Journal of Forecasting*. In Press, available online 12 February 2020.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

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

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

Bailey, N., Holly, S., and Pesaran, M. H. (2016a). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, 31(1):249–280.

Bailey, N., Kapetanios, G., and Pesaran, M. H. (2016b). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, 31(6):929–960.

Baltagi, B. H., Bresson, G., Griffin, J. M., and Pirotte, A. (2003). Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from french regional gasoline consumption. *Empirical Economics*, 28(4):795–811.

Baltagi, B. H., Bresson, G., and Pirotte, A. (2004). Tobin q: Forecast performance for hierarchical bayes, shrinkage, heterogeneous and homogeneous panel data estimators. *Empirical Economics*, 29(1):107–113.

Baltagi, B. H. and Griffin, J. M. (1997). Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics*, 77(2):303–327.

Baltagi, B. H., Griffin, J. M., and Xiong, W. (2000). To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics*, 82(1):117–126.

Baltagi, B. H. and Pirotte, A. (2010). Panel data inference under spatial dependence. *Economic Modelling*, 27(6):1368–1381.

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

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

Breitung, S. and Salish, N. (2020). Estimation of heterogeneous panels with systematic slope variations. *Forthcoming in Journal of Econometrics.*

Bresson, G., Hsiao, C., and Pirotte, A. (2011). Assessing the contribution of R&D to total factor productivity—a Bayesian approach to account for heterogeneity and heteroskedasticity. *AStA Advances in Statistical Analysis*, 95(4):435–452.

Breusch, T. S. and Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1):239–253.

Chudik, A. and Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2):393–420.

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal*, 14(1):C45–C90.

Garcia-Ferrer, A., Highfield, R. A., Palm, F., and Zellner, A. (1987). Macroeconomic forecasting using pooled international data. *Journal of Business & Economic Statistics*, 5(1):53–67.

Hoogstrate, A. J., Palm, F. C., and Pfann, G. A. (2000). Pooling in dynamic panel-data models: An application to forecasting GDP growth rates. *Journal of Business & Economic Statistics*, 18(3):274–283.

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (1999). Bayes estimation of short-run coefficients in dynamic panel data models. In C. Hsiao, K. Lahiri, L.-F. Lee, M. H. Pesaran, Eds, Analysis of Panels and Limited Dependent Variable Models: In Honour of G. S. Maddala, Cambridge University Press, Cambridge, Chapter 11, 268–297.

Jiang, B., Yang, Y., Gao, J., and Hsiao, C. (2017). Recursive estimation in large panel data models: Theory and practice. Working Paper no. 05/17, Monash Business School.

Jin, J. and Zou, H.-f. (2002). How does fiscal decentralization affect aggregate, national, and subnational government size? *Journal of Urban Economics*, 52(2):270–293.

Kapetanios, G. and Pesaran, M. H. (2007). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. In G. Phillips and E. Tzavalis, Eds, The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis, Cambridge University Press, Cambridge, Chapter 10, 239-281.

Karabiyik, H., Reese, S., and Westerlund, J. (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics*, 197(1):60–64.

Kelejian, H. H. and Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1):131–154.

Kuersteiner, G. M. and Prucha, I. R. (2018). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *arXiv preprint arXiv:1802.01755*.

Maddala, G. S., Li, H., and Srivastava, V. K. (1994). A comparative study of different shrinkage estimators for panel data models. *Annals of Economics and Finance*, 2(1):1–30.

Maddala, G. S., Trost, R. P., Li, H., and Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimator. *Journal of Business and Economic Statistics*, 15(1):90–100.

Mark, N. C. and Sul, D. (2012). When are pooled panel-data regression forecasts of exchange rates more accurate than the time-series regression forecasts? In James, J., Marsh, I. W., and Sarno, L., editors, *Handbook of Exchange Rates*, chapter 9, pages 265–281. Wiley.

Martinez-Vazquez, J., Lago-Peñas, S., and Sacchi, A. (2017). The impact of fiscal decentralization: A survey. *Journal of Economic Surveys*, 31(4):1095–1129.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

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

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34(6-10):1089–1117.

Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79–113.

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

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

Sarafidis, V. and Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews*, 31(5):483–531.

Shi, W. and Lee, L.-F. (2018). A spatial panel data model with time varying endogenous weights matrices and common factors. *Regional Science and Urban Economics*, 72:6–34.

Song, M. (2013). Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Unpublished Manuscript, Department of Economics, Columbia University.

Su, L., Shi, Z., and Phillips, P. C. B. (2016a). Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264.

Su, L., Shi, Z., and Phillips, P. C. B. (2016b). Supplement to "Identifying latent structures in panel data". *Econometrica*, 84(6):2215–2264.

Trapani, L. and Urga, G. (2009). Optimal forecasting with heterogeneous panels: A Monte Carlo study. *International Journal of Forecasting*, 25(3):567–586.

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

Appendix Variance formulas

For unit-specific estimates that we report in our empirical application, we use Newey and West (1987)'s heteroskedasticity and autocorrelation consistent (HAC) covariance estimator defined as

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}_{H,i}) = \frac{1}{T^2}\sum_{t,s=1}^{T} k_T\left(\frac{|t-s|}{p+1}\right)\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}_{is}'\widehat{u}_{it}\widehat{u}_{is}, \tag{A.1}$$

where $k_T(\cdot)$ is the kernel function with p being its associated bandwidth, $\tilde{\mathbf{x}}_{it}$ is the tth column of $\mathbf{X}_{i.}'\mathbf{M}_H$, $\widehat{u}_{it} = y_{it} - \widehat{\boldsymbol{\alpha}}_i'\mathbf{d}_t + \widehat{\boldsymbol{\beta}}_{H,i}\mathbf{x}_{it}$, $\widehat{\boldsymbol{\alpha}}_i = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'(\mathbf{y}_{i.} - \mathbf{X}_{i.}\widehat{\boldsymbol{\beta}}_i)$ and $\widehat{\boldsymbol{\beta}}_{H,i} = (\mathbf{X}_{i.}'\mathbf{M}_H\mathbf{X}_{i.})^{-1}\mathbf{X}_{i.}'\mathbf{M}_H\mathbf{y}_{i.}$ being a generic estimator defined by the matrix $\mathbf{M}_H$. For instance, for *CCE* estimator in (5), we have $\mathbf{M}_H = \bar{\mathbf{M}}_f$, for *PCX* in (10) we have $\mathbf{M}_H = \mathbf{M}_p$ etc. For the iterative estimator (14), the matrix $\mathbf{M}_H$ is defined by the common factor estimates achieved after numerical convergence. To implement this estimator we use the Bartlett kernel with a bandwidth equal to $\lfloor 2T^{1/2} \rfloor$ where $\lfloor\cdot\rfloor$ denotes the largest integer smaller than its argument.

In presence of heterogeneous slopes and an error term which contains common factors and spatial effects, Pesaran and Tosetti (2011) show that the non-parametric variance estimators proposed by Pesaran (2006) are valid and can be used to obtain robust standard errors for the slope coefficient estimates. In our Monte Carlo simulations, we use similar variance formulas for heterogeneous and homogeneous estimators. For the mean group and pooled estimators we respectively use

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}_{H,MG}) = \frac{1}{N\,(N-1)}\sum_{i=1}^{N}\underline{\widehat{\boldsymbol{\beta}}}_{H,i}\underline{\widehat{\boldsymbol{\beta}}}'_{H,i}, \tag{A.2}$$

and

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}_{H}) = \frac{1}{N}\mathbf{Q}_H^{-1}\boldsymbol{\Lambda}_H\mathbf{Q}_H^{-1}, \tag{A.3}$$

with

$$\mathbf{Q}_H = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{X}'_{i.}\mathbf{M}_H\mathbf{X}_{i.}}{T}\right),$$

$$\boldsymbol{\Lambda}_H = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\mathbf{X}'_{i.}\mathbf{M}_H\mathbf{X}_{i.}}{T}\right)\underline{\widehat{\boldsymbol{\beta}}}_{H,i}\underline{\widehat{\boldsymbol{\beta}}}'_{H,i}\left(\frac{\mathbf{X}'_{i.}\mathbf{M}_H\mathbf{X}_{i.}}{T}\right),$$

where $\underline{\widehat{\boldsymbol{\beta}}}_{H,i} = \widehat{\boldsymbol{\beta}}_{H,i} - \widehat{\boldsymbol{\beta}}_{H,MG}$, $\widehat{\boldsymbol{\beta}}_{H,i} = (\mathbf{X}'_{i.}\mathbf{M}_H\mathbf{X}_{i.})^{-1}\,\mathbf{X}'_{i.}\mathbf{M}_H\mathbf{y}_{i.}$ and $\widehat{\boldsymbol{\beta}}_{H,MG} = N^{-1}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{H,i}$. The matrix $\mathbf{M}_H$ is defined differently for different estimators as explained above.

If the slope parameters are in fact homogeneous, these variance estimators are not valid. Pesaran and Tosetti (2011) propose non-parametric variance estimators that combine Newey and West (1987)'s HAC and spatial HAC procedure of Kelejian and Prucha (2007). They are computed as

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}_{H,MG}) = \frac{1}{(NT)^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}k_S\left(\frac{\phi_{ij}}{\phi_N}\right)k_T\left(\frac{|t-s|}{p+1}\right)\underset{\sim}{\mathbf{x}}_{it}\underset{\sim}{\mathbf{x}}'_{js}\widehat{u}_{it}\widehat{u}_{js}, \tag{A.4}$$

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}_{H}) = \mathbf{Q}_H^{-1}\left[\frac{1}{(NT)^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}k_S\left(\frac{\phi_{ij}}{\phi_N}\right)k_T\left(\frac{|t-s|}{p+1}\right)\tilde{\mathbf{x}}_{it}\tilde{\mathbf{x}}'_{js}\widehat{u}_{it}\widehat{u}_{js}\right]\mathbf{Q}_H^{-1}, \tag{A.5}$$

where $k_S(\cdot)$ is the cross-section kernel function, ϕ_{ij} is the distance between the units i and j, ϕ_N is the threshold distance which is an increasing function of N, $\underset{\sim}{\mathbf{x}}_{it}$ is the tth column of $\underset{\sim}{\mathbf{X}}'_{i.} = \left(T^{-1}\mathbf{X}'_{i.}\mathbf{M}_H\mathbf{X}_{i.}\right)^{-1}\mathbf{X}'_{i.}\mathbf{M}_H$ and $\widehat{u}_{it} = y_{it} - \widehat{\boldsymbol{\alpha}}'_i\mathbf{d}_t + \widehat{\boldsymbol{\beta}}'\mathbf{x}_{it}$ with $\widehat{\boldsymbol{\beta}}$ being $\widehat{\boldsymbol{\beta}}_{H,MG}$ or $\widehat{\boldsymbol{\beta}}_H$ for the respective estimator.

As suggested by Bonhomme and Manresa (2015a), when heterogeneity follows a grouped pattern, one can use the estimator defined in (A.5) to estimate the covariance matrix of the *GFE* estimator. If the interest lies on inference concerning the group specific parameters $\boldsymbol{\lambda}_g$ we can apply this formula to each group in the panel to compute the covariance matrices of the *GFE* and *C-Lasso* estimators. To calculate the variance of these estimators, within each group g, we set $k_T\left(\frac{|t-s|}{p+1}\right) = 1$ if $t = s$, zero otherwise and $k_S\left(\frac{\phi_{ij}}{\phi_N}\right) = 1$ for all i, j.