

# Large Panels with Common Factors and Spatial Correlation\*

M. Hashem Pesaran  
Cambridge University and USC

Elisa Tosetti  
Cambridge University

May 31, 2010

## Abstract

This paper considers methods for estimating the slope coefficients in large panel data models that are robust to the presence of various forms of error cross section dependence. It introduces a general framework where error cross section dependence may arise because of unobserved common effects and/or error spill-over effects due to spatial or other forms of local dependencies. Initially, this paper focuses on a panel regression model where the idiosyncratic errors are spatially dependent and possibly serially correlated, and derives the asymptotic distributions of the mean group and pooled estimators under heterogeneous and homogeneous slope coefficients, and for these estimators proposes non-parametric variance matrix estimators. The paper then considers the more general case of a panel data model with a multifactor error structure and spatial error correlations. Under this framework, the Common Correlated Effects (CCE) estimator, recently advanced by Pesaran (2006), continues to yield estimates of the slope coefficients that are consistent and asymptotically normal. Small sample properties of the estimators under various patterns of cross section dependence, including spatial forms, are investigated by Monte Carlo experiments. Results show that the CCE approach works well in the presence of weak and/or strong cross sectionally correlated errors.

**Keywords:** Panels, Common Factors, Spatial Dependence, Common Correlated Effects Estimator.

**JEL Classification:** C10, C31, C33

---

\*We are grateful to the Editor (Cheng Hsiao), an Associate Editor and three anonymous referees, Badi Baltagi, Alexander Chudik and George Kapetanios for helpful comments and suggestions.

# 1 Introduction

Over the past few years there has been a growing literature, both empirical and theoretical, on econometric analysis of panel data models with cross sectionally dependent error processes. Such cross correlations can arise for a variety of reasons, such as omitted common factors, spatial spill-overs, and interactions within socioeconomic networks. Conditioning on variables specific to the cross section units alone does not deliver cross section error independence; an assumption required by the standard literature on panel data models. In the presence of such dependence, conventional panel estimators such as fixed or random effects can result in misleading inference and even inconsistent estimators (Phillips and Sul (2003)). Further, conventional panel estimators may be inconsistent if regressors are correlated with unobserved common factors that might be causing the error cross section dependence (Andrews (2005)).

Currently, there are two main strands in the literature for dealing with error cross section dependence in panels where  $N$  is large relative to  $T$ , namely the residual multifactor and the spatial econometric approaches. The multifactor approach assumes that the cross dependence can be characterized by a *finite* number of unobserved common factors, possibly due to economy-wide shocks that affect all units, albeit with different intensities. Under this framework, the error term is a linear combination of few common time-specific effects with heterogeneous factor loadings plus an idiosyncratic (individual-specific) error term. Estimation of a panel with such multifactor residual structure can be addressed by using statistical techniques commonly adopted in factor analysis, such as the maximum likelihood (Robertson and Symons (2000); Robertson and Symons (2007)), and the principal components procedures (Coakley, Fuertes, and Smith (2002); Bai (2009)). Recently, Pesaran (2006) has suggested an estimation method, referred to as Common Correlated Effects (CCE), that consists of approximating the linear combinations of the unobserved factors by cross section averages of the dependent and explanatory variables and then running standard panel regressions augmented with these cross section averages. An advantage of this approach is that it yields consistent estimates under a variety of situations, such as serial correlation in errors, unit roots in the factors and possible contemporaneous dependence of the observed regressors with the unobserved factors (Coakley, Fuertes, and Smith (2006); Kapetanios and Pesaran (2007); Kapetanios, Pesaran, and Yagamata (2009)).

The spatial approach assumes that the structure of cross section correlation is related to location and distance among units, defined according to a pre-specified metric. Proximity need not be measured in terms of physical space, but can be defined using other types of metrics, such as economic (Conley (1999); Pesaran, Schuermann, and Weiner (2004), policy, or social distance (Conley and Topa (2002)). Hence, cross section correlation is represented by means of a spatial process, which explicitly relates each unit to its neighbours (Whittle (1954)). Estimation of panels with spatially correlated errors can be based on maximum likelihood (ML) techniques (Lee (2004)), or on the generalized method of moments (GMM) (Kelejian and Prucha (1999); Lee (2007); Kelejian and Prucha (2009)). Recently, non-parametric methods based on heteroskedasticity and autocorrelation consistent estimators applied to spatial models have also been proposed (Conley (1999); Kelejian and Prucha (2007); Bester, Conley, and Hansen (2009)).

In this paper we build on the existing literature and consider a general panel data model where error cross section dependence is due to unobserved common factors and/or spatial dependence, whilst at the same time allow

for the errors to be serially correlated. We focus on estimation and inference procedures that are robust to the presence of various forms of cross sectional and temporal dependencies in the error processes. Robust methods are needed because the source and extent of error cross section dependence is often unknown. The error cross section dependence can take many different forms and its nature could differ at micro and macro levels. For instance, at a micro-level, individual consumption behaviour can be influenced by economy-wide factors, such as changes in taxation and interest rates, and by local neighbourhood effects such as keeping up with the Jones's (Cowan, Cowan, and Swann (2004)). In macroeconomics, several studies have argued business cycle fluctuations could be the result of both strategic interactions as well as aggregate technological shocks (Cooper and Haltiwanger (1996)). Our econometric specification, by allowing for the presence of both sources of contemporaneous error correlations, is sufficiently general and includes the models proposed in the literature as special cases.

We focus on estimation of slope coefficients in the case of a number of different specifications. Initially, we concentrate on a panel data model without unobserved factors where the errors are spatially dependent and possibly serially correlated, and derive the asymptotic distribution of the mean group and pooled estimators, under alternative assumptions regarding the slope coefficients. In the presence of heterogeneous slopes, we show that the non-parametric approach advanced by Pesaran (2006) continues to be applicable and can be used to obtain standard errors that are robust to both spatial and serial error correlations. However, in the case of homogeneous slopes the CCE procedure will not be applicable. In this case we propose a non-parametric variance matrix estimator that adapts the Newey and West (1987)'s heteroskedasticity autocorrelation consistent (HAC) procedure to allow for the spatial effects along the lines recently advanced by Kelejian and Prucha (2007). We refer to this variance estimator as spatial, heteroskedasticity, autocorrelation (SHAC) estimator. We then consider the more general case where the error term in the panel data model is composed of a multifactor structure and a spatial process, and show that Pesaran's CCE approach continues to be valid and yields consistent estimates of the slope coefficients and their standard errors. We also show how to obtain consistent estimates of the errors in the panel to be used in tests of cross section independence, and for further analysis of the underlying spatial processes.

Using Monte Carlo techniques, we investigate the small sample performance of the estimators under various patterns of error cross section dependence, with and without error serial correlation, under both cases of heterogeneous and homogeneous slopes. We examine the performance of the alternative estimators when the errors only display spatial dependence, when they are subject to unobserved common factors as well as spatial dependence, and in the case where the source of cross section dependence changes over time. Our results indicate that the mean group and pooled estimators with robust standard errors do work well under certain regularity conditions outlined in our theorems. However, under slope homogeneity or in the presence of unobserved common factors these estimators fail to provide correct inference. The results also document the tendency of the tests based on HAC type standard errors to over reject the null hypothesis in small samples even in the case of error cross section dependence which is purely spatial. In contrast, our Monte Carlo experiments clearly show that the augmentations of panel regressions with cross section averages, as formulated by the CCE procedure, eliminates the effects of all forms of spatial and temporal correlations, irrespective of whether these are due to spatial and/or unobserved common factors. The small sample properties of CCE estimators do not seem to be affected by the heterogeneity

assumptions on slope coefficients, or by the presence of error serial correlations. It is this level of robustness of the CCE estimator which particularly commends it for use in empirical analysis.

The plan of the remainder of the paper is as follows: Section 2 sets out a panel regression model with unobserved common factors and general spatial and temporal error processes. Section 3 develops the asymptotic distribution of the mean group and pooled estimators in the presence of spatial error dependence and error serial correlation. Section 4 considers the more general case where the errors also contain unobserved common factors, and establishes the validity of the CCE estimators for this class of models. Consistent estimation of the residuals from such models is considered in Section 5, where the necessary identification conditions are stated. Section 6 describes the Monte Carlo experiments and report the results. Section 7 ends with some concluding remarks.

**Notation:**  $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \dots \geq \lambda_n(\mathbf{A})$  are the eigenvalues of a matrix  $\mathbf{A} \in \mathbb{M}^{n \times n}$ , where  $\mathbb{M}^{n \times n}$  is the space of real  $n \times n$  matrices.  $\mathbf{A}^-$  denotes a generalized inverse of  $\mathbf{A}$ . The column norm of  $\mathbf{A} \in \mathbb{M}^{n \times n}$  is  $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$ . The row norm of  $\mathbf{A}$  is  $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$ . The Euclidean norm of  $\mathbf{A}$  is  $\|\mathbf{A}\|_2 = [\text{Tr}(\mathbf{A}\mathbf{A}')]^{1/2}$ .  $K$  is used for a fixed positive constant.  $(N, T) \xrightarrow{j} \infty$  denotes  $N$  and  $T$  tending to infinity jointly but in no particular order.

## 2 Heterogenous panels with unobserved common factors and spatial error correlation

We begin with a general specification where the dependent variable is a function of a set of individual-specific regressors, a linear combination of common observed and unobserved factors, and includes errors that are serially and spatially correlated. Let  $y_{it}$  be the observation on the  $i^{\text{th}}$  cross section unit at time  $t$  for  $i = 1, 2, \dots, N$ ;  $t = 1, 2, \dots, T$ , and suppose that it is generated as

$$y_{it} = \boldsymbol{\alpha}'_i \mathbf{d}_t + \boldsymbol{\beta}'_i \mathbf{x}_{it} + \boldsymbol{\gamma}'_i \mathbf{f}_t + e_{it}, \quad (1)$$

where  $\mathbf{d}_t = (d_{1t}, d_{2t}, \dots, d_{nt})'$  is a  $n \times 1$  vector of observed common effects, and  $\mathbf{x}_{it}$  is a  $k \times 1$  vector of observed individual-specific regressors on the  $i^{\text{th}}$  cross section unit at time  $t$ ,  $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{mt})'$  is an  $m$ -dimensional vector of unobservable common factors,  $\boldsymbol{\gamma}_i = (\gamma_{1i}, \gamma_{2i}, \dots, \gamma_{mi})'$  is the associated  $m \times 1$  vector of factor loadings. The number of factors,  $m$ , is assumed to be fixed relative to  $N$ , and in particular  $m < N$ . The common factors,  $\mathbf{f}_t$  simultaneously affect all cross section units, albeit with different degrees as measured by  $\boldsymbol{\gamma}_i$ . For instance, a rise in the interest rate may affect household consumption and firm investment decisions; oil price shocks may influence firm production costs; real shocks, such as a decline in the aggregate demand and employment could simultaneously slow growth in a number of countries (see Andrews (2005)).

Finally, the unit-specific or idiosyncratic errors,  $e_{it}$ , are assumed to be spatially and temporally correlated. The most widely used spatial models are the Spatial Moving Average (SMA), the Spatial Autoregressive (SAR) model, and the Spatial Error Component (SEC) specifications. These models differ in the range of dependence implied by their covariance matrices, but under certain invertibility conditions they can all be written as special

cases of

$$\mathbf{e}_t = \mathbf{R}_t \boldsymbol{\varepsilon}_t, \text{ for } t = 1, 2, \dots, T, \quad (2)$$

where  $\mathbf{e}_t = (e_{1t}, \dots, e_{Nt})'$ ,  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$  and  $\mathbf{R}_t$  is a given  $N \times N$  matrix.

We shall make use of the following assumptions.

**ASSUMPTION 1** For each  $i$ ,  $\varepsilon_{it}$  follows the linear stationary process with absolute summable autocovariances:

$$\varepsilon_{it} = \sum_{s=0}^{\infty} a_{is} \epsilon_{is},$$

where  $\epsilon_{is} \sim IID(0, 1)$  with finite fourth-order cumulants.

**ASSUMPTION 2**  $\mathbf{R}_t$  has bounded row and column norms for all  $t$ .

**ASSUMPTION 3** The slope coefficients  $\boldsymbol{\beta}_i$  follow the random coefficient model

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{v}_i, \quad \mathbf{v}_i \sim IID(\mathbf{0}, \boldsymbol{\Omega}_v) \quad \text{for } i = 1, 2, \dots, N,$$

where  $\|\boldsymbol{\beta}\|_2 < K$ ,  $\|\boldsymbol{\Omega}_v\|_2 < K$ ,  $\boldsymbol{\Omega}_v$  is a symmetric non-negative definite matrix, and the random deviations,  $\mathbf{v}_i$ , are distributed independently of  $\varepsilon_{jt}$ ,  $\mathbf{x}_{jt}$ , and  $\mathbf{d}_t$ , for all  $i, j$  and  $t$ .

**ASSUMPTION 4**  $(\mathbf{d}'_t, \mathbf{x}'_{it})'$  and  $\boldsymbol{\varepsilon}_{js}$  are independently distributed for all  $t, s, i$  and  $j$ .

Note that under Assumption 1 we have

$$Var(\varepsilon_{it}) = \sum_{s=0}^{\infty} a_{is}^2 = \sigma_i^2 \leq K < \infty,$$

and the covariance matrix of  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})'$  has bounded row and column norms, for all  $i$ . Assumption 2 implies that the spatial error process, (2), carries weak cross section dependence at all points in time, namely that its weighted averages converges to zero for all set of weights satisfying certain regularity conditions (see also Lemma A.1 in the Appendix). Notions of weak and strong cross section dependence are developed and discussed in Chudik, Pesaran, and Tosetti (2010). We note that Assumption 2 holds for most widely used spatial models that are subject to a set of regularity conditions that are standard in the spatial econometrics literature. These regularity conditions ensure consistency and asymptotic normality of quasi-ML and GMM estimators of spatial parameters (see Kelejian and Prucha (1999), Lee (2004), and Mardia and Marshall (1984) for details).

The model outlined in equations (1)-(2) is quite general and renders a variety of linear panel data models as special cases. The coefficients  $\boldsymbol{\alpha}_i$  may be treated as fixed or random, possibly correlated with the other variables in the panel. The vector  $\mathbf{d}_t$  could contain deterministic terms such as an intercept or linear trends, or common observed variables such as oil prices. For example, in the case of panel data models with fixed effects we would set  $n = 1$ , and  $d_{1t} = 1$ , for  $t = 1, 2, \dots, T$ .

The focus of this paper is on estimating the slope coefficients  $\beta_i$ , their cross section means,  $\beta = E(\beta_i)$ , and unit-specific errors,  $u_{it}$ . We shall consider four cases of interest. Initially, we abstract from unobserved common factors, and concentrate exclusively on the effects of weak spatial error dependence. Accordingly, we impose  $\gamma_i = \mathbf{0}$  in equation (1), and consider estimation of the slope coefficients and their cross section means in a panel regression model where  $e_{it}$  follows an invertible spatial process of type (2). We then investigate the properties of the proposed estimators under the special case of homogeneous slopes, namely when  $\beta_i = \beta$  for  $i = 1, 2, \dots, N$  (i.e.,  $\Omega_v = \mathbf{0}$  in Assumption 3). Next, we turn to the more general specification where  $\gamma_i \neq \mathbf{0}$ , and allow the unobserved common factors to be correlated with the individual-specific regressors,  $\mathbf{x}_{it}$ . Initially, we deal with the case of heterogeneous slopes and then consider the special case of  $\Omega_v = \mathbf{0}$ . There are two further specifications that may be derived from (1)-(2). These are the cases of common factors and no spatial error correlation with either heterogeneous or homogeneous slopes. However, these specifications have already been discussed in Pesaran (2006) and will not be considered here. Consistent estimators of the residuals in the general case is addressed in Section 5.

### 3 Estimating panels with spatial error correlation

The literature on spatial econometrics typically considers the problem of spatial dependence under strong assumptions of homogeneity and temporal independence. Only recently, a strand of literature in spatial econometrics has considered the incorporation of unobserved heterogeneity in spatial panel data models, where  $N$  is usually assumed to be large relative to  $T$ . Baltagi, Song, and Koh (2003) and Kapoor, Kelejian, and Prucha (2007) have focused on ML and GMM estimation of panels where the error term is the sum of an individual-specific component and a spatially autocorrelated idiosyncratic error. Baltagi, Egger, and Pfaffermayr (2009) generalized their earlier work by allowing for spatial correlations in both the individual means and the remainder error components, with possibly different spatial autoregressive parameters. Fingleton (2008) extended Kapoor, Kelejian, and Prucha (2007) contribution on GMM estimation of spatial random effects panels to the case where the idiosyncratic error term follows a spatial moving average process, while Egger, Pfaffermayr, and Winner (2005) have focused on extensions to the case of unbalanced panels. Lee and Yu (2008) considered estimation of a spatial panel data model with individual-specific fixed effects, and proposed a “transformation approach” to eliminate the fixed effects and then apply quasi-ML to the transformed model. Yu, de Jong, and Lee (2007), Yu, de Jong, and Lee (2008) and Yu and Lee (2007) focused on the properties of the quasi-ML estimator in the case of dynamic, possibly non-stationary, panels with fixed effects and spatial error correlation, assuming both  $N$  and  $T$  large.

It is worth noting that application of ML techniques requires the serial correlation processes of the error terms, if any, to be fully specified. In panels where  $N$  is relatively large this could be quite demanding, since different dynamic specifications might be appropriate across different cross sectional units. The GMM method is less demanding but still requires moment conditions that correctly take account of specific spatial and serial correlation patterns of the errors. The use of quasi-ML and GMM becomes even more involved if the errors also depend on unobserved common factors. It is, therefore, of interest to develop estimation and inference procedures for panels that are reasonably robust to the presence of cross section and temporal dependencies in the error processes.

In this section we focus on two estimators that can be used for estimating the mean,  $\beta$ , of the slope coefficients in equation (1), when errors are spatially correlated. The first, known as mean group (MG) estimator of  $\beta$ , is given by (see Pesaran and Smith (1995))

$$\hat{\beta}_{MG} = N^{-1} \sum_{i=1}^N \hat{\beta}_i, \quad (3)$$

where

$$\hat{\beta}_i = (\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{y}_i, \quad (4)$$

with  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$ ,  $\mathbf{X}'_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})$ ,  $\mathbf{M}_D = \mathbf{I}_T - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ , and  $\mathbf{D}' = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_T)$ . Alternatively, we can use the fixed effects, or pooled, estimator of  $\beta$

$$\hat{\beta}_P = \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_D \mathbf{y}_i. \quad (5)$$

To derive the asymptotic distribution of the above estimators, we make the following additional assumption on the individual-specific regressors and observed common factors.

**ASSUMPTION 5** *We assume:*

- (a) *For each  $i = 1, 2, \dots, N$ , the  $k \times k$  observation matrix  $T^{-1}\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i$  is non-singular for the sample size  $T$  under consideration and tends to a finite non-singular matrix,  $\mathbf{Q}_i$  as  $T \rightarrow \infty$ . Also, the elements of the  $k \times T$  matrix  $\mathbf{W}'_i = (T^{-1}\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{M}_D$  are uniformly bounded, and  $T^{-1}\mathbf{W}'_i \mathbf{W}_i = \left( \frac{\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i}{T} \right)^{-1}$  has bounded elements as  $T \rightarrow \infty$ .*
- (b) *The  $k \times k$  pooled observation matrix  $(NT)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i$  is finite and non-singular for sample sizes  $N$  and  $T$  under consideration and tends to a finite non-singular matrix,  $\mathbf{Q}$ , as  $(N, T) \xrightarrow{j} \infty$ .*

We have

$$\sqrt{N} \left( \hat{\beta}_{MG} - \beta \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[ (T^{-1}\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i)^{-1} T^{-1}\mathbf{X}'_i \mathbf{M}_D \mathbf{e}_i \right], \quad (6)$$

$$\sqrt{N} \left( \hat{\beta}_P - \beta \right) = \left( N^{-1} \sum_{i=1}^N T^{-1}\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N [T^{-1}\mathbf{X}'_i \mathbf{M}_D (\mathbf{X}_i \mathbf{v}_i + \mathbf{e}_i)]. \quad (7)$$

where  $\mathbf{e}_i = (e_{i1}, e_{i2}, \dots, e_{iT})'$ . The asymptotic distributions of (6) and (7) are summarized in the following theorems.

**Theorem 1 (MG estimator - Heterogeneous slopes, spatial corr. and no common factors)** *Consider the panel data model (1) with errors  $e_{it}$  following the spatial process given by (2). Suppose that Assumptions 1-4 and 5(a) hold and that  $\gamma_i = \mathbf{0}$ , for  $i = 1, 2, \dots, N$ . Then for the mean group estimator,  $\hat{\beta}_{MG}$ , given by (3), as  $(N, T) \xrightarrow{j} \infty$  we have*

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

where  $\Sigma_{MG} = \Omega_v$ .

**Theorem 2 (Pooled estimator - Heterogeneous slopes, spatial corr. and no common factors)** Consider the panel data model (1)-(2). Suppose that Assumptions 1-4 and 5(b) hold and that  $\gamma_i = \mathbf{0}$ , for  $i = 1, 2, \dots, N$ . Then for the pooled estimator,  $\hat{\beta}_P$ , given by (5), as  $(N, T) \xrightarrow{j} \infty$ , we have

$$\sqrt{N} \left( \hat{\beta}_P - \beta \right) \xrightarrow{d} N(\mathbf{0}, \Sigma_P),$$

where

$$\Sigma_P = \mathbf{Q}^{-1} \mathbf{\Lambda} \mathbf{Q}^{-1}, \quad (8)$$

with

$$\begin{aligned} \mathbf{Q} &= \lim_{N, T \rightarrow \infty} N^{-1} \sum_{i=1}^N \left( \frac{\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i}{T} \right), \\ \mathbf{\Lambda} &= \lim_{N, T \rightarrow \infty} \left[ N^{-1} \sum_{i=1}^N \left( \frac{\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i}{T} \right) \Omega_v \left( \frac{\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i}{T} \right) \right]. \end{aligned} \quad (9)$$

■

The proofs are provided in the Appendix. We observe that, to obtain the asymptotic distribution of both estimators, we have premultiplied them by  $\sqrt{N}$  rather than the usual  $\sqrt{NT}$ . This follows from the random coefficients hypothesis stated in Assumption 3, since the time-invariant variability of  $\beta_i$  dominates the other sources of randomness in the model. Robust estimators for  $\Sigma_P$  and  $\Sigma_{MG}$  can be obtained following the non-parametric approach employed in Pesaran (2006), which makes use of estimates of  $\beta$  computed for different cross sectional units. A consistent estimator of the asymptotic variance of the mean group estimator is given by

$$\widehat{Asy.Var} \left( \hat{\beta}_{MG} \right) = \frac{1}{N(N-1)} \sum_{i=1}^N \left( \hat{\beta}_i - \hat{\beta}_{MG} \right) \left( \hat{\beta}_i - \hat{\beta}_{MG} \right)'. \quad (10)$$

Similarly, a consistent non-parametric estimator of the asymptotic variance of the pooled estimator is

$$\widehat{Asy.Var} \left( \hat{\beta}_P \right) = \frac{1}{N} \mathbf{Q}_{NT}^{-1} \mathbf{\Lambda}_{NT} \mathbf{Q}_{NT}^{-1}, \quad (11)$$

where

$$\begin{aligned} \mathbf{Q}_{NT} &= \frac{1}{N} \sum_{i=1}^N (T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i), \\ \mathbf{\Lambda}_{NT} &= \frac{1}{N-1} \sum_{i=1}^N \left[ \left( \frac{\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i}{T} \right) \left( \hat{\beta}_i - \hat{\beta}_{MG} \right) \left( \hat{\beta}_i - \hat{\beta}_{MG} \right)' \left( \frac{\mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i}{T} \right) \right]. \end{aligned} \quad (12)$$

One advantage of the above non-parametric variance estimators is that their computation does not require *a priori* knowledge of the spatial arrangement of cross sectional units. As we shall see later in the paper, mis-specification of the spatial weights matrix may lead to substantial size distortions in tests based on the ML or quasi-ML



estimators of  $\beta_i$  (or  $\beta$ ). Another advantage of using the above approach over standard spatial techniques is that, while allowing for serially correlated errors, it does not entail information on the time series processes underlying  $\varepsilon_{it}$ , so long as these processes are covariance stationary.

Under the special case of homogeneous slopes, with  $\beta_i = \beta$  for all  $i$ , to obtain non-degenerate asymptotic distributions, the MG and pooled estimators should now be multiplied by  $\sqrt{NT}$ , rather than by  $\sqrt{N}$ . In this case, we have

$$\sqrt{NT} \left( \hat{\beta}_{MG} - \beta \right) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \left[ (T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{e}_i \right], \quad (13)$$

$$\sqrt{NT} \left( \hat{\beta}_P - \beta \right) = \left[ N^{-1} \sum_{i=1}^N (T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i) \right]^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_D \mathbf{e}_i. \quad (14)$$

Using results (A.15) and (A.16) in the appendix, the asymptotic distributions of (13) and (14) can be easily derived. These are set out in the following theorem.

**Theorem 3 (MG estimator - Homogeneous slopes, spatial corr. and no common factors)** *Consider the panel data model (1)-(2). Suppose that Assumptions 1-4 and 5(a) hold, that  $\gamma_i = \mathbf{0}$  for  $i = 1, 2, \dots, N$ , and  $\Omega_{\mathbf{v}} = \mathbf{0}$ . Then for the mean group estimator,  $\hat{\beta}_{MG}$ , given by (3), as  $N$  and/or  $T \rightarrow \infty$  we have*

$$\sqrt{NT} \left( \hat{\beta}_{MG} - \beta \right) \xrightarrow{d} N(\mathbf{0}, \Sigma_{MG}),$$

where

$$\Sigma_{MG} = \lim_{\mathfrak{M} \rightarrow \infty} \left( \frac{1}{\mathfrak{M}} \mathbf{H}' \Omega_{\varepsilon\varepsilon} \mathbf{H} \right), \quad (15)$$

with  $\mathfrak{M} = NT$ ,  $\mathbf{H}' = (\mathbf{W}'_{.1} \mathbf{R}_1, \mathbf{W}'_{.2} \mathbf{R}_2, \dots, \mathbf{W}'_{.T} \mathbf{R}_T)$ ,  $\mathbf{W}'_{.t} = (\mathbf{w}_{1t}, \mathbf{w}_{2t}, \dots, \mathbf{w}_{Nt})$ , and  $\mathbf{w}_{it}$  is the  $t^{\text{th}}$  column of  $\mathbf{W}'_{.i} = (T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{M}_D$ .

**Theorem 4 (Pooled estimator - Homogeneous slopes, spatial corr. and no common factors)** *Consider the panel data model (1)-(2). Suppose that Assumptions 1-4 and 5(b) hold, and that  $\gamma_i = \mathbf{0}$  for  $i = 1, 2, \dots, N$ , and  $\Omega_{\mathbf{v}} = \mathbf{0}$ . Then for the pooled estimator,  $\hat{\beta}_P$ , given by (5), as  $N$  and/or  $T \rightarrow \infty$  we have*

$$\sqrt{NT} \left( \hat{\beta}_P - \beta \right) \xrightarrow{d} N(\mathbf{0}, \Sigma_P),$$

where

$$\Sigma_P = \mathbf{Q}^{-1} \Psi_P \mathbf{Q}^{-1}, \quad (16)$$

with

$$\begin{aligned}\mathbf{Q} &= \lim_{\mathfrak{M} \rightarrow \infty} \left( \frac{1}{\mathfrak{M}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i \right), \\ \Psi_P &= \lim_{\mathfrak{M} \rightarrow \infty} \left( \frac{1}{\mathfrak{M}} \mathbf{P}' \Omega_{\varepsilon\varepsilon} \mathbf{P} \right),\end{aligned}$$

$\mathfrak{M} = NT$ ,  $\mathbf{P}' = (\tilde{\mathbf{X}}'_1 \mathbf{R}_1, \tilde{\mathbf{X}}'_2 \mathbf{R}_2, \dots, \tilde{\mathbf{X}}'_T \mathbf{R}_T)$ ,  $\tilde{\mathbf{X}}_t = (\tilde{\mathbf{x}}_{1t}, \tilde{\mathbf{x}}_{2t}, \dots, \tilde{\mathbf{x}}_{Nt})'$ , and  $\tilde{\mathbf{x}}_{it}$  is the  $t^{\text{th}}$  column of  $\tilde{\mathbf{X}}'_i = \mathbf{X}'_i \mathbf{M}_D$ . ■

The asymptotic variances  $\Sigma_{MG}$  and  $\Sigma_P$  depend on the particular specifications of  $\mathbf{R}_t$ ,  $t = 1, 2, \dots, T$ , and on  $\Omega_{\varepsilon\varepsilon}$ . One important question is to determine whether the robust variance estimators introduced above can still be used under the case of homogeneous slopes. To investigate this issue, one possibility would be to check whether the individual estimators  $\sqrt{T}(\hat{\beta}_i - \beta)$ , for  $i = 1, 2, \dots, N$  (see formula 4), are asymptotically independent and normal across  $i$ . Under these conditions, using results in Ibragimov and Müller (2010), it is easy to show that it is still valid to base inference on  $\hat{\beta}_{MG}$  and its robust variance estimator given by (10). These authors prove that the type I error using a  $t$ -test based on  $\hat{\beta}_{MG}$  is not greater than the level of significance chosen for this test,  $\alpha$ , under the condition that  $\alpha < 0.083$  (see also Bakirov and Székely (2006)). Note that

$$\sqrt{T}(\hat{\beta}_i - \beta) = (T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i)^{-1} \frac{1}{\sqrt{T}} \mathbf{X}'_i \mathbf{M}_D \mathbf{e}_i,$$

and the covariance between  $\sqrt{T}(\hat{\beta}_i - \beta)$  and  $\sqrt{T}(\hat{\beta}_j - \beta)$  for  $i \neq j$ , is given by

$$\begin{aligned}T \cdot E \left[ (\hat{\beta}_i - \beta) (\hat{\beta}_j - \beta)' \right] &= \frac{1}{T} \mathbf{W}'_i E(\mathbf{e}_i \mathbf{e}'_j) \mathbf{W}_j = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T [\mathbf{w}_{it} \mathbf{w}'_{js} E(e_{it} e_{js})] \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T [\mathbf{w}_{it} \mathbf{w}'_{js} E(\mathbf{r}'_{i,t} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_s \mathbf{r}_{j,s})] \\ &= \frac{1}{T} \sum_{t,s=1}^T \left[ (\mathbf{w}_{it} \mathbf{w}'_{js}) \sum_{h=1}^N E(r_{ih,t} r_{jh,s} \varepsilon_{ht} \varepsilon_{hs}) \right].\end{aligned}$$

The above covariance is zero for all  $i \neq j$  only under certain conditions. For example, it is zero if the idiosyncratic errors are cross sectionally independent (though possibly serially correlated), or if the elements,  $r_{ih,t}$ , are random and satisfy  $E(r_{ih,t} r_{jh,s}) = 0$ , for for all  $i \neq j$ ,  $h = 1, 2, \dots, N$  and  $t, s = 1, 2, \dots, T$ . We observe that condition  $E(r_{ih,t} r_{jh,s}) = 0$  holds if the entries in the  $i^{\text{th}}$  row in  $\mathbf{R}_t$  are independently distributed of the entries in the  $j^{\text{th}}$  row of  $\mathbf{R}_t$ , at all time periods. However, these restrictions are unlikely to hold in general.

To obtain robust estimates of the asymptotic variances in the general case one possibility would be to consider a generalized version of the Newey-West procedure that allows for the spatial effects. For purely spatial error processes, heteroskedasticity, spatially consistent (HSC) estimators have been proposed by Conley (1999) (see also early contributions by Driscoll and Kraay (1998), and the method suggested in Pinkse, Slade, and Brett

(2002)). More recently, Kelejian and Prucha (2007) have proposed a new HSC estimator that approximate the true covariance matrix with a weighted average of cross products of regression errors, where each element is weighted by a function of (possible multiple) distances between cross section units. Bester, Conley, and Hansen (2009), using results taken from Ibragimov and Müller (2010), propose to divide the sample in groups so that group-level averages are approximately independent, and accordingly suggest a HSC estimator based on a discrete group-membership metric. However, the validity of this approach relies on the ability of the researcher to construct groups whose averages are approximately independent. In contrast, the Kelejian and Prucha (2007) approach stands out as a flexible and robust method, as it does not entail high level assumptions, allows for multiple distance measures, and is robust to some measurement errors in the specification of the distance matrix.

Note that (15) and (16) can be written as

$$\begin{aligned}\boldsymbol{\Sigma}_{MG} &= \frac{1}{NT} \sum_{i,j=1}^N \sum_{t,s=1}^T \mathbf{w}_{it} \mathbf{w}'_{js} E(e_{it} e_{js}), \\ \boldsymbol{\Sigma}_P &= \mathbf{Q}^{-1} \left[ \frac{1}{NT} \sum_{i,j=1}^N \sum_{t,s=1}^T \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}'_{js} E(e_{it} e_{js}) \right] \mathbf{Q}^{-1}.\end{aligned}$$

Following Kelejian and Prucha (2007), assume that there exists a “meaningful”, time-invariant, measure of distance between cross sectional units, summarized in the  $N \times N$  matrix,  $\mathbf{D}$  with elements  $d_{ij} \geq 0$ . Let  $\hat{e}_{it} = y_{it} - \hat{\boldsymbol{\alpha}}'_i \mathbf{d}_t - \hat{\boldsymbol{\beta}}'_{MG} \mathbf{x}_{it}$  (see Section 5 for estimation of  $\boldsymbol{\alpha}_i$ ). A Newey-West SHAC estimator of the variance of  $\hat{\boldsymbol{\beta}}_{MG}$  can be computed as

$$\widehat{Asy.Var}(\hat{\boldsymbol{\beta}}_{MG}) = \frac{1}{(NT)^2} \sum_{i,j=1}^N \sum_{t,s=1}^T K\left(\frac{d_{ij}}{d_N}, \frac{|t-s|}{p+1}\right) \mathbf{w}_{it} \mathbf{w}'_{js} \hat{e}_{it} \hat{e}_{js}, \quad (17)$$

where  $d_N > 0$  is an arbitrary scalar function of  $N$ ,  $p$  is the window size for the time series dimension, and  $K(\cdot)$  is a kernel function that we set equal to

$$K\left(\frac{d_{ij}}{d_N}, \frac{|t-s|}{p+1}\right) = K_1\left(\frac{d_{ij}}{d_N}\right) K_2\left(\frac{|t-s|}{p+1}\right),$$

where  $K_1(\cdot)$  and  $K_2(\cdot)$  satisfy a set of regularity conditions (see, in particular, Assumption 7 in Kelejian and Prucha (2007)). Note that  $K_1\left(\frac{d_{ij}}{d_N}\right) = 0$  for  $d_{ij} > d_N$ , and  $K_2\left(\frac{|t-s|}{p+1}\right) = 0$  for  $|t-s| > p+1$ , and that  $K_1(0) = K_2(0) = 1$ . Similarly, a Newey-West SHAC estimator of the variance of  $\hat{\boldsymbol{\beta}}_P$  is

$$\widehat{Asy.Var}(\hat{\boldsymbol{\beta}}_P) = \mathbf{Q}_{NT}^{-1} \left[ \frac{1}{(NT)^2} \sum_{i,j=1}^N \sum_{t,s=1}^T K\left(\frac{d_{ij}}{d_N}, \frac{|t-s|}{p+1}\right) \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}'_{js} \hat{e}_{it} \hat{e}_{js} \right] \mathbf{Q}_{NT}^{-1}, \quad (18)$$

where  $\mathbf{Q}_{NT}$  is given by (12), and  $\hat{e}_{it}$  is now given by  $\hat{e}_{it} = y_{it} - \hat{\boldsymbol{\alpha}}'_i \mathbf{d}_t - \hat{\boldsymbol{\beta}}'_P \mathbf{x}_{it}$ . The rest of the notations are as above. We observe that the above estimators require knowledge of the exact relative position of units across space, although as argued in Kelejian and Prucha (2007), the estimator remains valid under certain mis-specifications of the distance metric.

## 4 Estimating panels with unobserved common factors and spatial error correlation

We now turn to the estimation of the slope coefficients in the context of panels with both common factors and spatial error dependence. We restrict our attention to the CCE approach since, as compared to other existing methods, it is simple to apply and has been shown to be robust to the choice of  $m$  (the number of common factors), the temporal dynamics of unobserved common factors, and the idiosyncratic error. The idea underlying this approach is that, as far as estimation of the slope coefficients are concerned, the unobservable common factors can be well approximated by the cross section averages of the dependent variable  $\bar{y}_t = N^{-1} \sum_{i=1}^N y_{it}$  and individual-specific regressors,  $\bar{\mathbf{x}}_t = N^{-1} \sum_{i=1}^N \mathbf{x}_{it}$ . Hence, estimation can be carried out by least squares applied to auxiliary regressions where the observed regressors are augmented with these cross section averages plus the observed common factors,  $\mathbf{d}_t$ .

To model the correlation between the individual-specific regressors,  $\mathbf{x}_{it}$ , and the common factors, it is supposed that

$$\mathbf{x}_{it} = \mathbf{A}_i' \mathbf{d}_t + \mathbf{\Gamma}_i' \mathbf{f}_t + \mathbf{v}_{it}, \quad (19)$$

where  $\mathbf{A}_i$  and  $\mathbf{\Gamma}_i$  are  $n \times k$  and  $m \times k$  factor loading matrices with fixed components, and  $\mathbf{v}_{it}$  is the individual-specific component of  $\mathbf{x}_{it}$ .

Let  $\bar{\mathbf{M}}$  be defined by

$$\bar{\mathbf{M}} = \mathbf{I}_T - \bar{\mathbf{H}}(\bar{\mathbf{H}}'\bar{\mathbf{H}})^{-1}\bar{\mathbf{H}}', \quad (20)$$

$\bar{\mathbf{H}} = (\mathbf{D}, \bar{\mathbf{Z}})$ , where  $\mathbf{D}$  and  $\bar{\mathbf{Z}}$  are, respectively, the matrices of observations on  $\mathbf{d}_t$  and  $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{\mathbf{x}}_t')$ . We make the following assumptions on the common factors and their loadings and on the individual, or unit-specific, errors:

**ASSUMPTION 6** *The  $(n+m) \times 1$  vector  $\mathbf{g}_t = (\mathbf{d}_t', \mathbf{f}_t')$  is a covariance stationary process, with absolute summable autocovariances, distributed independently of  $e_{is}$  and  $\mathbf{v}_{is}$  for all  $i, t, s$ .<sup>1</sup>*

**ASSUMPTION 7** *The unobserved factor loadings,  $\gamma_i$  and  $\mathbf{\Gamma}_i$  are bounded, i.e.  $\|\gamma_i\|_2 < K$  and  $\|\mathbf{\Gamma}_i\|_2 < K$ , for all  $i$ . Further, it is assumed that the random deviations,  $\mathbf{v}_i$ , for the slope coefficients are independently distributed of  $\gamma_i$  and  $\mathbf{\Gamma}_i$ .*

**ASSUMPTION 8** *The individual-specific errors  $e_{it}$  and  $\mathbf{v}_{js}$  are distributed independently for all  $i, j, t$  and  $s$ , and for each  $i$ ,  $\mathbf{v}_{it}$  follows a linear stationary process with absolute summable autocovariances given by*

$$\mathbf{v}_{it} = \sum_{\ell=0}^{\infty} \mathbf{\Pi}_{i\ell} \zeta_{i,t-\ell},$$

where for each  $i$ ,  $\zeta_{it}$  is a  $k \times 1$  vector of serially uncorrelated random variables with mean zero, the variance matrix

---

<sup>1</sup>This assumption can be relaxed to allow for unit roots in the common factors, along the lines set out in Kapetanios, Pesaran, and Yagamata (2009).

$\mathbf{I}_k$ , and finite fourth-order cumulants. For each  $i$ , the coefficient matrices  $\mathbf{\Pi}_{i\ell}$  satisfy the condition

$$\text{Var}(\mathbf{v}_{it}) = \sum_{\ell=0}^{\infty} \mathbf{\Pi}_{i\ell} \mathbf{\Pi}'_{i\ell} = \mathbf{\Sigma}_{v_i},$$

where  $\mathbf{\Sigma}_{v_i}$  is a positive definite matrix, such that  $\sup_i \|\mathbf{\Sigma}_{v_i}\|_2 < K$ .

**ASSUMPTION 9** Let  $\tilde{\mathbf{\Gamma}} = E(\gamma_i, \mathbf{\Gamma}_i) = (\gamma, \mathbf{\Gamma})$ . We assume that  $\text{Rank}(\tilde{\mathbf{\Gamma}}) = m$ .

**ASSUMPTION 10** Consider the cross section averages of the individual-specific variables,  $\mathbf{z}_{it} = (y_{it}, \mathbf{x}'_{it})'$  defined by  $\bar{\mathbf{z}}_{.t} = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_{it}$ , and let  $\bar{\mathbf{M}}$  be defined by (20). Then the following conditions hold:

- (a) The matrix  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbf{\Sigma}_{v_i}$  exists and is non-singular.
- (b) There exists  $T_0$  and  $N_0$ , such that for all  $T \geq T_0$  and  $N \geq N_0$ , the  $k \times k$  matrices  $(T^{-1} \mathbf{X}'_{i.} \bar{\mathbf{M}} \mathbf{X}_{i.})$ , and  $(T^{-1} \mathbf{X}'_{i.} \mathbf{M}_g \mathbf{X}_{i.})$ , where  $\mathbf{M}_g = \mathbf{I}_T - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$ , with  $\mathbf{G} = (\mathbf{D}, \mathbf{F})$ , exist and are non-singular for all  $i$ , and  $\sup_i E \left\| \frac{\mathbf{X}'_{i.} \mathbf{M}_g \mathbf{X}_{i.}}{T} \right\|^2 < K < \infty$ .

**Remark 1** Note that Assumption 10 provides extensions of Assumption 5 to the case where individual-specific regressors,  $\mathbf{x}_{it}$ , are random and correlated with the common factors. These conditions ensure the existence of the probability limits involved in the derivation of the asymptotic distribution of the CCE estimators as  $(N, T) \xrightarrow{j} \infty$ .

Following Pesaran (2006), the mean group and pooled estimators for  $\beta$  in a panel with spatial correlation and common factors are given by (3) and (5), applied to a regression equation where the observed regressors are augmented with the cross section averages of the dependent variable,  $\bar{y}_{.t}$ , and of the regressors,  $\bar{\mathbf{x}}_{.t}$ . Specifically, the CCE mean group estimator is

$$\hat{\beta}_{CCEMG} = N^{-1} \sum_{i=1}^N \hat{\beta}_{CCE,i}, \quad (21)$$

where

$$\hat{\beta}_{CCE,i} = (\mathbf{X}'_{i.} \bar{\mathbf{M}} \mathbf{X}_{i.})^{-1} \mathbf{X}'_{i.} \bar{\mathbf{M}} \mathbf{y}_{i.}, \quad (22)$$

and the CCE pooled estimator is

$$\hat{\beta}_{CCEP} = \left( \sum_{i=1}^N \mathbf{X}'_{i.} \bar{\mathbf{M}} \mathbf{X}_{i.} \right)^{-1} \sum_{i=1}^N \mathbf{X}'_{i.} \bar{\mathbf{M}} \mathbf{y}_{i.}, \quad (23)$$

The following theorems apply to the above estimators (proofs are provided in the Appendix).

**Theorem 5 (CCE MG estimator - Heterogeneous slopes, spatial corr. and common factors)** Consider the panel data model given by equations (1), (2), and (19). Suppose that Assumptions 1-3 and 6-10 hold. Then for the common correlated effects mean group estimator  $\hat{\beta}_{CCEMG}$  given by (21), as  $(N, T) \xrightarrow{j} \infty$  we have

$$\sqrt{N} \left( \hat{\beta}_{CCEMG} - \beta \right) \rightarrow N(\mathbf{0}, \mathbf{\Sigma}_{CCEMG}), \quad (24)$$

where  $\Sigma_{CCEMG} = \Omega_v$ .

**Theorem 6 (CCE Pooled estimator - Heterogeneous slopes, spatial corr. and common factors)** Consider the panel data model given by equations (1), (2) and (19). Suppose that Assumptions 1-3 and 6-10 hold. Then for the common correlated effects pooled estimator  $\hat{\beta}_{CCEP}$  given by (23), as  $(N, T) \xrightarrow{j} \infty$ , we have

$$\sqrt{N} \left( \hat{\beta}_{CCEP} - \beta \right) \rightarrow N(\mathbf{0}, \Sigma_{CCEP}),$$

where

$$\Sigma_{CCEP} = \Psi^{*-1} \mathbf{R}^* \Psi^{*-1},$$

with

$$\begin{aligned} \Psi^* &= \lim_{N \rightarrow \infty} \left( \frac{1}{N} \sum_{i=1}^N \Sigma_{v_i} \right), \\ \mathbf{R}^* &= \lim_{N \rightarrow \infty} \left[ \frac{1}{N} \sum_{i=1}^N \Sigma_{v_i} \Omega_v \Sigma_{v_i} \right]. \end{aligned}$$

■

Consistent estimators for the asymptotic variances of  $\hat{\beta}_{CCEP}$  and  $\hat{\beta}_{CCEMG}$  are (see also Section 3 above)

$$\widehat{Asy.Var} \left( \hat{\beta}_{CCEMG} \right) = \frac{1}{N(N-1)} \sum_{i=1}^N \left( \hat{\beta}_{CCE,i} - \hat{\beta}_{CCEMG} \right) \left( \hat{\beta}_{CCE,i} - \hat{\beta}_{CCEMG} \right)', \quad (25)$$

$$\widehat{Asy.Var} \left( \hat{\beta}_{CCEP} \right) = \frac{1}{N} \hat{\Psi}^{*-1} \hat{\mathbf{R}}^* \hat{\Psi}^{*-1}. \quad (26)$$

with

$$\begin{aligned} \hat{\Psi}^* &= \frac{1}{N} \sum_{i=1}^N \left( \frac{\mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i}{T} \right), \\ \hat{\mathbf{R}}^* &= \frac{1}{N-1} \sum_{i=1}^N \left[ \left( \frac{\mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i}{T} \right) \left( \hat{\beta}_{CCE,i} - \hat{\beta}_{CCEMG} \right) \left( \hat{\beta}_{CCE,i} - \hat{\beta}_{CCEMG} \right)' \left( \frac{\mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i}{T} \right) \right]. \end{aligned}$$

As in the pure spatial case, if the slope coefficients  $\beta_i$  are homogeneous the CCE estimators must be multiplied by  $\sqrt{NT}$ , rather than by  $\sqrt{N}$ , to obtain non-degenerate asymptotic distributions, namely

$$\sqrt{NT} \left( \hat{\beta}_{CCEMG} - \beta \right) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \left[ (T^{-1} \mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i)^{-1} \mathbf{X}'_i \bar{\mathbf{M}} (\mathbf{F} \gamma_i + \mathbf{e}_i) \right], \quad (27)$$

$$\sqrt{NT} \left( \hat{\beta}_{CCEP} - \beta \right) = \left[ N^{-1} \sum_{i=1}^N (T^{-1} \mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i) \right]^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \bar{\mathbf{M}} (\mathbf{F} \gamma_i + \mathbf{e}_i). \quad (28)$$

Using results (A.17)-(A.18) and (A.19)-(A.20) in the Appendix, it follows that  $\hat{\beta}_{CCEMG}$  and  $\hat{\beta}_{CCEP}$  continue to be consistent for  $\beta$  as  $N \rightarrow \infty$ , for  $T$  fixed or  $T \rightarrow \infty$ , although their asymptotic distributions will generally depend on nuisance parameters. Following similar lines of reasoning as in the pure spatial case, we now investigate whether  $\hat{\beta}_{CCEMG}$  together with the non-parametric variance estimator (10) can still be used in the homogeneous slopes case. First note that, using results (A.12)-(A.14) in the Appendix, we have

$$\sqrt{T} \left( \hat{\beta}_{CCE,i} - \beta \right) = \hat{\Psi}_{iT}^{-1} \frac{\mathbf{X}'_i \bar{\mathbf{M}} (\mathbf{F} \gamma_i + \mathbf{e}_i)}{\sqrt{T}} = \frac{1}{\sqrt{T}} \mathbf{W}'_i \mathbf{e}_i + O_p \left( \frac{\sqrt{T}}{N} \right) + O_p \left( \frac{1}{\sqrt{N}} \right),$$

where  $\mathbf{W}'_i = (T^{-1} \mathbf{X}'_i \mathbf{M}_g \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{M}_g$ . Further, for large  $N$  and  $T$  with  $\sqrt{T}/N \rightarrow 0$ ,

$$\begin{aligned} T \cdot \left[ \left( \hat{\beta}_{CCE,i} - \beta \right) \left( \hat{\beta}_{CCE,j} - \beta \right)' \right] &= \hat{\Psi}_{iT}^{-1} \left[ T^{-1} \mathbf{X}'_i \bar{\mathbf{M}} (\mathbf{F} \gamma_i + \mathbf{e}_i) (\gamma'_j \mathbf{F}' + \mathbf{e}'_j) \bar{\mathbf{M}} \mathbf{X}_j \right] \hat{\Psi}_{jT}^{-1} \\ &\approx T^{-1} \mathbf{W}'_i \mathbf{e}_i \mathbf{e}'_j \mathbf{W}_j. \end{aligned}$$

As in the pure spatial case, the above expression is asymptotically zero only under certain conditions, for example when the idiosyncratic errors are cross sectionally independent, or if the entries of the matrix  $\mathbf{R}_t$ , for  $t = 1, 2, \dots, T$ , are random and independently distributed across rows. Later in the paper, we will investigate the small sample properties of tests based on robust variances (25) and (26) both under heterogenous and homogeneous slopes.

**Remark 2** *The CCE continues to be applicable even if the rank condition outlined in Assumption 9 is not satisfied. Failure of the rank condition can occur if there is an unobserved factor for which the average of the loadings in the  $y_{it}$  and  $\mathbf{x}_{it}$  equations tends to a zero vector (Pesaran and Tosetti (2009)). This could happen if, for example, such a factor carries weak cross section dependence. Another possible reason for failure of Assumption 9 is if the number of unobservable factors,  $m$ , is larger than  $k + 1$ , where  $k$  is the number of regressors. In these cases, common factors cannot be estimated from cross section averages. However, it is possible to show that the cross sectional means of the slope coefficients,  $\beta_i$ , can still be consistently estimated, under the additional assumption that the unobserved factor loadings,  $\gamma_i$ , in equation (1) are independently and identically distributed across  $i$ , and of  $e_{jt}$ ,  $\mathbf{v}_{jt}$ , and  $\mathbf{g}_t = (\mathbf{d}'_t, \mathbf{f}'_t)'$  for all  $i, j$  and  $t$ . No assumptions (other than Assumption 7) are required on the loadings attached to the regressors,  $\mathbf{x}_{it}$ . The proofs of consistency and asymptotic normality of the CCE estimator in the rank deficiency case are straightforward extensions of the results provided in Pesaran (2006).*

**Remark 3** *We observe that the CCE estimator does not entail any assumptions on the cumulative effect of factors on cross section units. This is in contrast to the use of principal components that require errors to display a strong factor structure, namely that  $\sum_{i=1}^N \gamma_{i\ell}^2 > \sqrt{c} \sigma^2$ , for  $\ell = 1, 2, \dots, m$ , where  $c$  is such that  $\frac{N}{T} - c = o(N^{-1/2})$ , and  $\sigma^2$  is the variance of the idiosyncratic error. In the absence of this condition the principal components estimates of the factors would be inconsistent. See, for example, Onatski (2009), and Paul (2007).*

**Remark 4** *Kapetanios, Pesaran, and Yagamata (2009) considered the case where the unobservable common factors follow unit root processes and could be cointegrated. They showed that the asymptotic distribution of panel estimators in the case of  $I(1)$  factors is similar to that in the stationary case, reported in Theorems 5 and 6 above.*

## 5 Residuals from CCE regression

We now consider the consistent estimation of regression errors  $u_{it} = y_{it} - \boldsymbol{\alpha}'_i \mathbf{d}_t - \boldsymbol{\beta}'_i \mathbf{x}_{it}$  in model (1). Estimation of  $u_{it}$  is needed for computing tests of error cross section independence, or when the objects of interest are the coefficients of the spatial process,  $e_{it}$ . Before continuing, without loss of generality, we specify some further assumptions on the observed and unobserved common factors. In particular:

**ASSUMPTION 11**  $E(\mathbf{f}_t) = \mathbf{0}$ , for  $t = 1, \dots, T$ , and the  $n \times 1$  vector of observed common factors,  $\mathbf{d}_t$ , is distributed independently of  $\mathbf{f}_{t'}$ , for all  $t$  and  $t'$ , such that

$$\frac{\mathbf{D}'\mathbf{F}}{T} = O_p\left(\frac{1}{\sqrt{T}}\right). \quad (29)$$

This is an identification condition that allows to separate the effects of observed and unobserved common effects in  $u_{it}$ . Note that the cross section averages,  $\bar{\mathbf{z}}_{.t}$ , contain information not only on the unobserved factors,  $\mathbf{f}_t$ , but also on the observed factors,  $\mathbf{d}_t$ . Given that the number, nature and the source of the unobserved common factors are unknown, without Assumption 11 it would not be possible to separate the effects of these two sets of common variables. However, since  $\mathbf{f}_t$  is unobserved this assumption can be easily accommodated by a suitable re-definition of  $\mathbf{f}_t$  and the associated factor loadings.

Consider the OLS estimates

$$\hat{\boldsymbol{\alpha}}_i = (\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}' \left( \mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{CCE,i} \right), \quad (30)$$

where  $\hat{\boldsymbol{\beta}}_{CCE,i}$  is given by (22). Under Assumptions 1-3 and 6-10, and given (A.12)-(A.14), we have<sup>2</sup>

$$\hat{\boldsymbol{\beta}}_{CCE,i} - \boldsymbol{\beta}_i = (T^{-1} \mathbf{X}'_i \bar{\mathbf{M}} \mathbf{X}_i)^{-1} \frac{\mathbf{X}'_i \bar{\mathbf{M}} (\mathbf{F} \boldsymbol{\gamma}_i + \mathbf{e}_i)}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i &= (T^{-1} \mathbf{D}'\mathbf{D})^{-1} \left( \frac{\mathbf{D}'\mathbf{X}_i}{T} \right) (\boldsymbol{\beta}_i - \hat{\boldsymbol{\beta}}_{CCE,i}) + (T^{-1} \mathbf{D}'\mathbf{D})^{-1} \left( \frac{\mathbf{D}'\mathbf{F}}{T} \right) \boldsymbol{\gamma}_i \\ &\quad + (T^{-1} \mathbf{D}'\mathbf{D})^{-1} \left( \frac{\mathbf{D}'\mathbf{e}_i}{T} \right), \end{aligned}$$

and hence

$$\hat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \quad (31)$$

Note that, unlike in the case of a simple panel data model with fixed effects and no unobserved common factors, consistency of  $\hat{\boldsymbol{\alpha}}_i$  requires both  $N$  and  $T$  going to infinity, due to the additional  $O_p(N^{-1})$  term in (31). This

---

<sup>2</sup>Note that under our assumptions  $T^{-1} \mathbf{D}'\mathbf{e}_i = O_p(T^{-1/2})$ . Indeed,  $E(T^{-1} \mathbf{D}'\mathbf{e}_i) = \mathbf{0}$ , since under Assumption 6,  $\mathbf{D}$  and  $\mathbf{e}_i$  are independently distributed. Further, under Assumption 6 and from Proposition A.1 (see result (A.2)), the largest eigenvalue of  $E(\mathbf{e}_i \mathbf{e}'_i)$  is bounded. It follows that

$$Var(T^{-1} \mathbf{D}'\mathbf{e}_i) = E(T^{-2} \mathbf{D}'\mathbf{e}_i \mathbf{e}'_i \mathbf{D}) \leq K \cdot E(T^{-2} \mathbf{D}'\mathbf{D}) = O(T^{-1}).$$



term arises since the unobserved common factors are approximated by cross section averages. Now, consider the residuals  $\hat{u}_{it} = y_{it} - \hat{\beta}_{CCE,i}\mathbf{x}_{it} - \hat{\alpha}'_i\mathbf{d}_t$ . Given the consistency of  $\hat{\beta}_{CCE,i}$  and  $\hat{\alpha}_i$ , it follows that

$$u_{it} = \hat{u}_{it} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

Similarly, in the homogenous case, adopting the CCEP estimator, under Assumptions 1-3 and 6-10, from (A.20) we have

$$\begin{aligned}\hat{\alpha}_i - \alpha_i &= (T^{-1}\mathbf{D}'\mathbf{D})^{-1}\left(\frac{\mathbf{D}'\mathbf{X}_i}{T}\right)(\beta - \hat{\beta}_{CCEP}) + (T^{-1}\mathbf{D}'\mathbf{D})^{-1}\left(\frac{\mathbf{D}'\mathbf{F}}{T}\right)\gamma_i + (T^{-1}\mathbf{D}'\mathbf{D})^{-1}\left(\frac{\mathbf{D}'\mathbf{e}_i}{T}\right) \\ &= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).\end{aligned}\quad (32)$$

Let  $\hat{u}_{it} = y_{it} - \hat{\beta}'_{CCEP}\mathbf{x}_{it} - \hat{\alpha}'_i\mathbf{d}_t$ . Given (32) and (A.20), we obtain

$$u_{it} = \hat{u}_{it} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Principal components analysis can be applied to the above residuals,  $\hat{u}_{it}$ , to estimate the common factors,  $\mathbf{f}_t$ , and their loadings,  $\gamma_i$ . Note that these residuals continue to be consistent, as  $(N, T) \xrightarrow{j} \infty$ , even when the loadings attached to the unobserved factors are set to zero, namely, when the data generating process is (1)-(2). In this case, the parameters of the spatial process can be recovered by applying standard spatial econometric techniques to  $\hat{u}_{it}$ .

## 6 Monte Carlo experiments

### 6.1 Monte Carlo design

This section provides Monte Carlo evidence on the small sample properties of our estimators, under a range of assumptions on the stochastic process generating the error terms. The study is comprised of three sets of experiments. In the first set, we consider a panel where the error term is generated by a SAR process and with no common factors. In the second set, we assume that the error process is the orthogonal sum of a factor structure and a spatial process, and allow the dependent variable and the individual-specific regressors to be correlated with the unobserved common factors. In the third set of experiments, we make a number of robustness checks, to see the extent to which our estimators are effective in dealing with various special conditions, such as when errors are serially correlated, there are sizeable spatial error correlations, or when the pattern of cross section dependence varies over time.

For all experiments we considered the following data generating process

$$\begin{aligned}y_{it} &= \alpha_i d_{1t} + \beta_{i1} x_{1it} + \beta_{i2} x_{2it} + \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + e_{it}, \\ x_{ijt} &= a_{ij1} d_{1t} + a_{ij2} d_{2t} + \gamma_{ij1} f_{1t} + \gamma_{ij3} f_{3t} + v_{ijt}, \quad j = 1, 2,\end{aligned}$$

for  $i = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ . In the above equations,  $d_{1t}$  and  $d_{2t}$  are observed common factors,  $f_{1t}$ ,  $f_{2t}$ , and  $f_{3t}$  are unobserved common effects, and  $e_{it}$  are idiosyncratic errors. We adopt the following data generating processes:

$$\begin{aligned} d_{1t} &= 1, & d_{2t} &= \rho_d d_{2,t-1} + v_{dt}, & t &= -49, \dots, 0, 1, \dots, T, \\ v_{dt} &\sim IIDN(0, 1 - \rho_d^2), & \rho_d &= 0.5, & d_{2,-50} &= 0, \end{aligned}$$

$$\begin{aligned} f_{\ell t} &= \rho_{f_\ell} f_{\ell,t-1} + v_{f_{\ell t}}, & \ell &= 1, 2, 3; & t &= -49, \dots, 0, 1, \dots, T, \\ v_{f_{\ell t}} &\sim IIDN(0, 1 - \rho_{f_\ell}^2), & \rho_{f_\ell} &= 0.5, & f_{\ell,-50} &= 0, \end{aligned}$$

and

$$\begin{aligned} v_{ijt} &= \rho_{v_{ij}} v_{ij,t-1} + \vartheta_{ijt}, & t &= -49, \dots, 0, 1, \dots, T, \\ \vartheta_{ijt} &\sim N(0, 1 - \rho_{\vartheta_{ij}}^2), & v_{ij,-50} &= 0, \\ \rho_{\vartheta_{ij}} &\sim IIDU(0.05, 0.95) & \text{for } j &= 1, 2. \end{aligned}$$

The first 50 observations are discarded. The factor loadings of the observed common effects do not change across replications and are generated as

$$\begin{aligned} \alpha_i &\sim IIDN(1, 1), & i &= 1, 2, \dots, N, \\ (a_{i11}, a_{i21}, a_{i12}, a_{i22}) &\sim IIDN(0.5\boldsymbol{\tau}_4, 0.5\mathbf{I}_4), \end{aligned}$$

where  $\boldsymbol{\tau}_4 = (1, 1, 1, 1)'$  and  $\mathbf{I}_4$  is a  $4 \times 4$  identity matrix.

We consider two alternative sets of experiments, that involve different hypotheses on the data generating process for the loadings of the unobserved common factors, and the way the idiosyncratic errors  $e_{it}$  are generated:

A The factor loadings of the unobserved common effects are set to zero,  $\gamma_{i11} = \gamma_{i13} = \gamma_{i21} = \gamma_{i23} = \gamma_{i1} = \gamma_{i2} = 0$ , and the individual-specific errors,  $e_{it}$ , are generated according to the SAR process

$$e_{it} = \delta_t \sum_{j=1}^N s_{ij} e_{jt} + \varepsilon_{it}, \quad \text{for } i = 1, 2, \dots, N, t = 1, 2, \dots, T, \quad (33)$$

$$\varepsilon_{it} \sim N(0, \sigma_i^2), \quad \sigma_i^2 \sim IIDU(0.5, 1.5), \quad \text{for } i = 1, 2, \dots, N, \quad (34)$$

where  $\delta_t$  is the time-varying spatial autoregressive coefficient, that we set  $\delta_t = \delta = 0.4$ .  $s_{ij}$ , for  $i, j = 1, 2, \dots, N$ , are elements of a spatial weights matrix  $\mathbf{S}$ , assumed to be time-invariant. We follow Kelejian and Prucha (2007) and assume that units are located on a rectangular grid at locations  $(r, s)$ , for  $r = 1, \dots, m_1; s = 1, 2, \dots, m_2$ , such that  $N = m_1 m_2$ .<sup>3</sup> The distance  $d_{ij}$  between units is given by the Euclidean distance, and

---

<sup>3</sup>For a given value of  $N = m_1 m_2$ , we set  $m_1$  and  $m_2$  such that these are integer numbers and  $|m_1 - m_2|$  is minimized. In particular,

$\mathbf{S}$  is taken to be a rook-type matrix where two units are neighbors if their Euclidean distance is less than or equal to one. The weights matrix is normalized such that the weights in each row sum to one.

B The parameters of the unobserved common effects in the  $\mathbf{x}_{it}$  and  $y_{it}$  equations are generated as

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

and

$$\gamma_{i1} \sim IIDN(1, 0.2), \gamma_{i2} \sim IIDN(1, 0.2), \gamma_{i3} = 0,$$

and the individual-specific errors,  $e_{it}$ , are generated as in (33) and (34), with  $\delta_t = \delta = 0.4$ . This set of experiments aims at investigating the extent to which the CCE estimators capture the effects of local as well global cross section dependence.

For each case, we consider two alternative assumptions on the slope coefficients:

- (i) The case of heterogeneous slopes where  $\beta_{ij} = \beta_j + \eta_{ij}$ , with  $\beta_j = 1$ , and  $\eta_{ij} \sim IIDN(0, 0.04)$ , for  $i = 1, 2, \dots, N$  and  $j = 1, 2$ , varying across replications.
- (ii) The case of homogeneous slopes where  $\beta_{ij} = 1$ , for  $i = 1, 2, \dots, N$  and  $j = 1, 2$ .

### Experiment C: robustness checks

The aim of this set of experiments is to investigate the extent to which the use of robust standard errors are effective in dealing with serially correlated errors, high spatial error correlation, and time variations in the degree and source of error cross section dependence:

1. Serially correlated errors. We allow errors  $\varepsilon_{it}$  in (33) to be serially correlated. In particular,  $\varepsilon_{it}$  are generated as stationary AR(1) processes for half of the cross-section units, and as MA(1) processes for the remaining cross-section units:

$$\begin{aligned} \varepsilon_{it} &= \rho_{i\varepsilon} \varepsilon_{i,t-1} + \sigma_i (1 - \rho_{i\varepsilon}^2)^{1/2} \zeta_{it}, & i = 1, \dots, \lfloor N/2 \rfloor, \\ \varepsilon_{it} &= \sigma_i (1 + \theta_{i\varepsilon}^2)^{-1/2} (\zeta_{it} + \theta_{i\varepsilon} \zeta_{i,t-1}), & i = \lfloor N/2 \rfloor + 1, \dots, N. \end{aligned}$$

where  $\zeta_{it} \sim IIDN(0, 1)$ ,  $\sigma_i^2 \sim IIDU(0.5, 1.5)$ ,  $\rho_{i\varepsilon} \sim IIDU(0.05, 0.95)$ , and  $\theta_{i\varepsilon} \sim IIDU(0, 0.8)$ . For this sub-experiment we set  $\delta_t = \delta = 0.4$ , no unobserved common factors, and consider both cases of heterogeneous and homogeneous slopes.

2. High spatial error correlation. For this sub-experiment we set  $\delta_t = \delta = 0.8$ . Further, we assume that there are no unobserved common factors and the slope coefficients are heterogeneous (i.e. as in case (i)).

---

for  $N = 20$  we set  $m_1 = 5$ ,  $m_2 = 4$ , for  $N = 30$  we set  $m_1 = 6$ ,  $m_2 = 5$ , for  $N = 50$ , we set  $m_1 = 10$ ,  $m_2 = 5$ , and for  $N = 100$  we set  $m_1 = m_2 = 10$ .

3. Time-varying spatial correlation. The spatial autoregressive coefficients are generated as  $\delta_t \sim IIDU(0, 0.8)$ , for  $t = 1, 2, \dots, T$ , and fixed across replications. For this sub-experiment we assume no unobserved common factors and heterogeneous slopes (i.e. as in case (i)).
4. Time-varying cross section dependence. We allow the cross dependence to change from weak to strong and back to weak. Specifically, for  $t = 1, 2, \dots, \lfloor T/3 \rfloor$  we assume parameters of the unobserved common effects in the  $\mathbf{x}_{it}$  and in the  $y_{it}$  equations and the errors  $e_{it}$  are generated as in Experiment A, with  $\delta_t = \delta = 0.8$ . For  $t = \lfloor T/3 \rfloor + 1, \dots, \lfloor 2T/3 \rfloor$  parameters of the unobserved common effects in the  $\mathbf{x}_{it}$  and in the  $y_{it}$  equations and the errors  $e_{it}$  are generated as in Experiment B, with  $\delta_t = \delta = 0$  (i.e., error processes include common factors only). For  $t = \lfloor 2T/3 \rfloor + 1, \dots, T$  parameters of the unobserved common effects in the  $\mathbf{x}_{it}$  and in the  $y_{it}$  equations and the errors  $e_{it}$  are generated as in Experiment A, with  $\delta_t = \delta = 0.8$ . For this sub-experiment we assume heterogeneous slopes (i.e. as in case (i)). The aim of this set of experiments is to investigate the robustness of our estimators to the possible time variations in the degree of cross section dependence.

Each experiment was replicated 2,000 times for the  $(N, T)$  pairs with  $N, T = 20, 30, 50, 100$ . We report the small sample properties for a number of estimators of the slope coefficients. In particular, we computed the mean group estimator (3), both with robust variance (10) and with SHAC variance (17), the pooled estimator (5) both with robust variance (11) and with SHAC variance (18). The SHAC variance estimators have been computed using both the true distance matrix  $\mathbf{D}$  and a mis-specified version of the distance matrix obtained by incorrectly assuming that units are ordered on a line, rather than on a rectangular grid. Following Kelejian and Prucha (2007), we have set the parameter  $d_N$  in (17) and (18) equal to  $N^{1/4}$ , and have fixed the window size for the time series part equal to  $2T^{1/2}$ . We also computed the ML estimator for a panel containing fixed effects, with spatially correlated and heteroskedastic errors. The likelihood function of this model for a given spatial matrix,  $\mathbf{S}$ , is derived in a supplement which is available on request. Related derivations are also provided in Anselin (1988) and Lee (2004). We refer to this estimator as the ML-SAR estimator. The ML-SAR estimator is computed for two different spatial weights matrices; a correctly specified one, and a mis-specified version, where units with Euclidean distance less than or equal to two are incorrectly taken as neighbors. This is done with the intent to check the effect of mis-specification of  $\mathbf{S}$  on the ML-SAR estimator. Finally, we report results for the CCE mean group estimator (21) with variance (25), and the CCE pooled estimator (23) with variance (26).

## 6.2 Monte Carlo results

Results for Experiment A are summarized in Tables A1-A2, for Experiment B in Tables B1-B2, and for Experiment C in Tables C1-C4. Each table provides estimates of bias, root mean squared errors (RMSE), size, and power. The nominal size is set to 5 per cent, while the power of the various tests is computed under the alternative  $H_1 : \beta_1 = 0.95$ . In what follows we focus on estimation of  $\beta_1$ ; results for  $\beta_2$  are very similar and are not reported.

Tables A1-A2 summarize the results for the case where the errors are generated by a spatial autoregressive process without any common factors, under heterogeneous slopes (Table A1) and homogeneous slopes (Table A2). We first note that, for these experiments, the mean group and pooled procedures provide unbiased estimators for the mean of the slope coefficients,  $\beta$ . Accordingly, these estimators display very small biases and their RMSEs

decline steadily with increases in  $N$  and/or  $T$ . Considering the empirical sizes of the tests, the ones based on the robust variance estimators (10) and (11) display rejection frequencies that are close to the nominal size under heterogeneous slopes, while they slightly over-reject the null under slope homogeneity, namely when  $\beta_i = \beta$ , for all  $i$ .

Indeed, as noted in Section 3, the robust standard errors given by (10) and (11) are not applicable under slope homogeneity, and consistent estimation of the variance of the pooled and mean group estimators in general requires knowledge of the spatial arrangement of the cross section units. In contrast, the tests based on SHAC variances severely over-reject the null hypothesis in all experiments with heterogeneous slopes, while they do have the correct size when data are generated under  $\beta_i = \beta$ , and when  $N$  and  $T$  are sufficiently large. Indeed, when  $N$  and  $T$  are smaller than 50, the SHAC based tests are slightly over-sized. That the use of Newey-West robust standard errors lead to an over-rejection of the null hypothesis in small samples is well known within the time series literature (see, for example, the Monte Carlo study reported in Smith and McAleer (1994)). Our results seem to indicate that adopting the Newey-West procedure jointly with the Kelejian and Prucha (2007) variance estimator in a panel data framework may also lead to over-rejection of the null hypothesis in small samples. We note that Kelejian and Prucha, in their Monte Carlo experiments, only report results when  $N$  is relatively large (they focus on a single cross section,  $T = 1$ , with  $N = 400$  and  $1024$ ). Also, they do not report sizes of tests based on their proposed variance estimator (see also the Monte Carlo study reported in Fingleton and Le Gallo (2008)). To further investigate this issue, we have run some additional experiments using Kelejian and Prucha (2007) Monte Carlo design, with  $T = 1$ ,  $N = 100, 200, 400, 1024$ . The results show that tests based on the non-parametric standard errors proposed by Kelejian and Prucha have empirical sizes close to the 5 per cent nominal size if  $N \geq 400$ , but tend to over-reject for smaller values of  $N$ .<sup>4</sup> However, the results in Table A2 show that errors in the measurement of the distance between cross section units does not seem to have much affect on the properties of SHAC estimators, which is in line with the theoretical results obtained in Kelejian and Prucha.

A number of other interesting findings also emerge from the results reported in Table A2. We can see that under the ideal conditions that the spatial process generating the error term and the spatial arrangement of units are both known, the ML estimator has the correct size for large  $T$ , and a high power. Also tests based on CCE Mean Group and CCE Pooled estimators have empirical sizes that are very close to the nominal size, under both cases of heterogeneous and homogeneous slopes. These findings suggest that augmenting the panel regressions with cross section averages even in the absence of common factors can help deal with spatial error spillover effects. The attraction of the CCE type estimators in these contexts lies in the fact that they do not require a quantification of the exact relative position of the units in space, which is required by the SHAC type estimators. But, not surprisingly a comparison of the power of the CCE type tests with the tests based on the ML-SAR in Table A2 shows that not using information on the spatial ordering of cross section units can result in some loss of power.

In Experiments B (Tables B1-B2), the combination of common factors and spatial correlation in the error term leads to large distortions in the pooled and mean group estimators. The bias and RMSE of the ML-SAR estimator are smaller than those of the pooled estimator, although they remain substantial even for large values of  $N$  and  $T$ . Further, tests based on the ML-SAR estimator substantially over-reject the null hypothesis. However,

---

<sup>4</sup>To save space, we have not reported these results. However, they are available upon request.

the combination of common factors and spatial correlation in the errors does not affect the empirical size of CCE estimators, which is close to the nominal size of 5 per cent.

Turning to results in Tables C1-C5, we observe that serial correlation in errors does not seem to affect the properties of mean group and pooled estimators with robust variances (10) and (11), and of CCE estimators under both heterogeneous and homogeneous slopes (see Tables C1-C2). Also the over-rejection tendency of the SHAC estimator is much more pronounced in the presence of residual serial correlation (Table C2). Turning to the experiments with a high value of the spatial coefficient ( $\delta = 0.8$ ), we see from Table C3 that tests based on mean group and pooled estimators that use robust standard errors tend to over-reject, in some cases significantly. But it is interesting that the CCE estimators continue to have the correct size even with such high degrees of spatial dependence, although there is some evidence of a loss in power. One explanation for this result is that, when the degree of spatial correlation is high, an unobserved factor structure may better approximate the process generating cross section dependence, and the CCE type estimators that allow for a factor error structure might be more appropriate.

Finally, results reported in Table C5 suggest that CCE estimators are also robust to possible time variations in the degree of cross section dependence. This important property of CCE type estimators is not necessarily shared by estimation methods that use principal components (see Bai (2009)), since time variation in the degree of cross section dependence can yield inconsistent estimates of the principal components. We refer to Chudik, Pesaran, and Tosetti (2010) for a comparison of the CCE method with the principal components approach in the estimation of panel regression models subject to common factors.

## 7 Concluding remarks

The main aim of this paper has been to consider estimation of a panel regression model under a number of different specifications of cross section error correlations, such as spatial and/or common factor models. We have derived the asymptotic distributions of the mean group and pooled estimators for a panel regression model where the source of error cross section dependence is purely spatial or results from omitted unobserved factors, or both. In each case we have distinguished between panels when the slopes are homogeneous across the cross section units and when they are not. Our main conclusion (based on theoretical and Monte Carlo results) is that the augmentation of panel regressions with cross section averages together with non-parametric variance estimators associated with the CCE estimators, seem to be most effective in dealing with error cross section dependencies, irrespective of whether they arise from are spatial spillovers or are due to the presence of unobserved common factors. The CCE type estimators also seem to be robust to possible serial correlations in the errors and time variations in the degree and nature of cross section error dependence. Our Monte Carlo results also document the tendency of the tests based on the SHAC type standard errors to over-reject the null hypothesis in small samples even in the case of error cross section dependence which is purely spatial in nature.

## A Appendix: proof of Theorems 1–2 and 5–6

The following two Lemmas establish a few key results used in the proofs of Theorems 1-6.

**Lemma A.1** Consider the process (2), where  $\boldsymbol{\varepsilon}_{.t} = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$  satisfies Assumption 1, and  $\mathbf{R}_t$  satisfies Assumption 2. Then for all  $t$

$$E(\bar{\varepsilon}_{.t}^2) = O(N^{-1}), \text{ and } Var(\bar{\varepsilon}_{.t}^2) = O(N^{-2}), \quad (\text{A.1})$$

where  $\bar{\varepsilon}_{.t} = \frac{1}{N} \sum_{i=1}^N \varepsilon_{it}$  and

$$\lambda_1 [E(\mathbf{e}_i \mathbf{e}_j')] = O(1), \text{ for all } i \text{ and } j, \quad (\text{A.2})$$

where  $\mathbf{e}_i = (e_{i1}, e_{i2}, \dots, e_{iT})'$ .

**Proof.** First note that, under Assumption 1,  $\varepsilon_{it}$  has mean zero, finite variances  $0 < \sigma_i^2 < \sigma_{\max}^2 < \infty$ , and finite fourth-order moments,  $E(\varepsilon_{it}^4) = \mu'_{i4} < K < \infty$ . To prove (A.1), note that

$$E(\bar{\varepsilon}_{.t}^2) = \frac{1}{N^2} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\Lambda}_\varepsilon \mathbf{R}_t' \boldsymbol{\tau} \leq \frac{1}{N^2} \sigma_{\max}^2 (\boldsymbol{\tau}' \boldsymbol{\tau}) \lambda_1(\mathbf{R}_t \mathbf{R}_t').$$

where  $\boldsymbol{\tau}$  is an  $N$ -dimensional vector of ones. But since  $\mathbf{R}_t$  has bounded row and column norms,  $\lambda_1(\mathbf{R}_t \mathbf{R}_t')$  is bounded, and we have

$$E(\bar{\varepsilon}_{.t}^2) = O(N^{-1}).$$

Let  $\mathbf{M}_t = \frac{1}{N^2} \boldsymbol{\Lambda}_\varepsilon^{1/2} \mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\Lambda}_\varepsilon^{1/2}$ , with elements  $m_{ij,t}$ , for  $i, j = 1, \dots, N$ . The diagonal elements of  $\mathbf{M}_t$  satisfy (denoting the  $i^{\text{th}}$  column of  $\mathbf{R}_t$  by  $\mathbf{r}_{.i,t}$ )

$$\begin{aligned} m_{ii,t} &= \frac{1}{N^2} \sigma_{ii}^2 \mathbf{r}_{.i,t}' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{r}_{.i,t} = \frac{1}{N^2} \sigma_{ii}^2 \left( \sum_{j=1}^N r_{ji,t} \right)^2 \\ &\leq \frac{1}{N^2} \sigma_{ii}^2 \left( \sum_{j=1}^N |r_{ji,t}| \right)^2. \end{aligned}$$

But by assumption  $\sum_{j=1}^N |r_{ji,t}| = O(1)$ , then  $m_{ii,t} = O(N^{-2})$  for all  $i$  and  $t$ . Using results on moments for quadratic forms established in the literature (see Ullah (2004)), we have

$$\begin{aligned} Var(\bar{\varepsilon}_{.t}^2) &= Var \left[ \left( \frac{1}{N} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\varepsilon}_{.t} \right)^2 \right] = E \left[ \left( \frac{1}{N^2} \boldsymbol{\varepsilon}_{.t}' \mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\varepsilon}_{.t} \right)^2 \right] - \left[ E \left( \frac{1}{N^2} \boldsymbol{\varepsilon}_{.t}' \mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\varepsilon}_{.t} \right) \right]^2 \\ &= \left[ Tr \left( \frac{1}{N^2} \mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\Lambda}_\varepsilon \right) \right]^2 + 2Tr \left[ \left( \frac{1}{N^2} \mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\Lambda}_\varepsilon \right)^2 \right] \\ &\quad + \sum_{i=1}^N (\mu'_{i4} - 3\sigma_i^4) m_{ii,t}^2 - \left[ Tr \left( \frac{1}{N^2} \mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\Lambda}_\varepsilon \right) \right]^2 \\ &= 2Tr \left[ \left( \frac{1}{N^2} \mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\Lambda}_\varepsilon \right)^2 \right] + \sum_{i=1}^N (\mu'_{i4} - 3\sigma_i^4) m_{ii,t}^2. \end{aligned}$$

But  $|\mu'_{i4} - 3\sigma_i^4| < \mu'_{i4} + 3\sigma_i^4 < K$ , and therefore

$$\begin{aligned} Var(\bar{\varepsilon}_{.t}^2) &\leq 2 \frac{1}{N^4} Tr(\mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t \boldsymbol{\Lambda}_\varepsilon^2 \mathbf{R}_t' \boldsymbol{\tau} \boldsymbol{\tau}' \mathbf{R}_t) + K \sum_{i=1}^N m_{ii,t}^2 \\ &\leq 2 \frac{1}{N^4} \sigma_{\max}^4 (\boldsymbol{\tau}' \mathbf{R}_t \mathbf{R}_t' \boldsymbol{\tau})^2 + O(N^{-3}) \\ &\leq 2 \frac{1}{N^4} \sigma_{\max}^4 [\lambda_1(\mathbf{R}_t \mathbf{R}_t')]^2 (\boldsymbol{\tau}' \boldsymbol{\tau})^2 + O(N^{-3}), \end{aligned}$$

which establishes the second result in (A.1). To establish (A.2) consider the  $(t, s)^{th}$  element of the  $T \times T$  matrix  $E(\mathbf{e}_i \mathbf{e}'_j)$  and note that, since for  $h \neq q$ ,  $E(\varepsilon_{ht} \varepsilon_{qs}) = 0$ ,

$$E(e_{it} e_{js}) = E(\mathbf{r}'_{i.,t} \boldsymbol{\varepsilon}_{.t} \boldsymbol{\varepsilon}'_{.s} \mathbf{r}_{j.,t}) = \sum_{h=1}^N \sum_{q=1}^N r_{ih,t} r_{jq,s} E(\varepsilon_{ht} \varepsilon_{qs}) = \sum_{q=1}^N r_{iq,t} r_{jq,s} E(\varepsilon_{qt} \varepsilon_{qs}),$$

and the largest eigenvalue of  $E(\mathbf{e}_i \mathbf{e}'_j)$  satisfies (using the result that  $\lambda_1(\mathbf{A}) \leq \|\mathbf{A}\|_\infty$ )

$$\begin{aligned} \lambda_1 [E(\mathbf{e}_i \mathbf{e}'_j)] &\leq \max_{1 \leq s \leq T} \sum_{t=1}^T \sum_{q=1}^N |r_{iq,t}| |r_{jq,s}| |E(\varepsilon_{qt} \varepsilon_{qs})| \\ &\leq q \max_{1 \leq s \leq T} \sum_{q=1}^N |r_{iq,s}| \sum_{t=1}^T |E(\varepsilon_{qt} \varepsilon_{qs})| = O(1), \end{aligned}$$

given that, by Assumption 1,  $\sum_{t=1}^T |E(\varepsilon_{qt} \varepsilon_{qs})| = O(1)$ , and by Assumption 2,  $|r_{iq,t}| = O(1)$ ,  $\sum_{q=1}^N |r_{iq,s}| = O(1)$ , for all  $s$ . ■

Therefore, for any process of form (2) with  $\mathbf{R}_t$  having bounded row and column norms,  $\bar{\mathbf{e}}_t^2$  converges to zero in quadratic mean as  $N \rightarrow \infty$ , and the degree of cross section dependence of  $\mathbf{e}_i$  will be bounded in  $N$ .

**Lemma A.2** Consider the general process  $\mathbf{e}_t = \mathbf{R}_t \boldsymbol{\varepsilon}_t$ . Then under Assumptions 1-8 we have

$$\frac{\bar{\mathbf{e}}' \bar{\mathbf{e}}}{T} = O_p\left(\frac{1}{N}\right), \quad (\text{A.3})$$

$$\frac{\mathbf{F}' \bar{\mathbf{e}}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \frac{\mathbf{D}' \bar{\mathbf{e}}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{A.4})$$

$$\frac{\mathbf{V}'_i \bar{\mathbf{e}}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \frac{\mathbf{e}'_i \bar{\mathbf{e}}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{A.5})$$

where  $\bar{\mathbf{e}} = (\bar{e}_{.1}, \dots, \bar{e}_{.T})'$ ,  $\bar{e}_{.t}^2 = N^{-1} \sum_{i=1}^N e_{it}$ ,  $\mathbf{D}$  and  $\mathbf{F}$  are  $T \times n$  and  $T \times m$  matrices on observed and unobserved common factors, and  $\mathbf{V}_i = (\mathbf{v}_{i1}, \dots, \mathbf{v}_{iT})'$ .

**Proof.** Note that  $T^{-1} \bar{\mathbf{e}}' \bar{\mathbf{e}} = T^{-1} \sum_{t=1}^T \bar{e}_{.t}^2$ . From the Markov inequality we have, for every  $\epsilon > 0$ ,

$$P\left(\left|\frac{\bar{\mathbf{e}}' \bar{\mathbf{e}}}{T}\right| \geq \epsilon\right) \leq \frac{1}{\epsilon} E\left|\frac{\bar{\mathbf{e}}' \bar{\mathbf{e}}}{T}\right| = \frac{1}{\epsilon} T^{-1} \sum_{t=1}^T E(\bar{e}_{.t}^2) = O\left(\frac{1}{N}\right). \quad (\text{A.6})$$

which proves (A.3). As for (A.4), consider the  $\ell^{th}$  row of  $T^{-1} \mathbf{F}' \bar{\mathbf{e}}$  and note that it can be written as  $T^{-1} \sum_{t=1}^T f_{\ell t} \bar{e}_{.t}$ , where  $f_{\ell t}$  and  $\bar{e}_{.t}$  are distributed independently of each other. Then, given (A.1),  $T^{-1} \sum_{t=1}^T f_{\ell t} \bar{e}_{.t}$  has zero mean and variance

$$\begin{aligned} \text{Var}\left(T^{-1} \sum_{t=1}^T f_{\ell t} \bar{e}_{.t}\right) &= T^{-2} \sum_{t=1}^T \sum_{t'=1}^T [E(f_{\ell t} f_{\ell t'}) E(\bar{e}_{.t} \bar{e}_{.t'})] \\ &\leq \sigma_{\max}^2 O\left(\frac{1}{N}\right) \left[T^{-2} \sum_{t=1}^T \sum_{t'=1}^T E(f_{\ell t} f_{\ell t'})\right] = O\left(\frac{1}{NT}\right). \end{aligned}$$

This establishes (A.4). The second result in (A.4) and the first result in (A.5) follow similarly. As for the second result in (A.5), note that

$$\begin{aligned} T^{-1} \mathbf{e}'_i \bar{\mathbf{e}} &= T^{-1} \sum_{t=1}^T e_{it} \bar{e}_{.t} = \frac{1}{NT} \sum_{t=1}^T \sum_{j=1}^N (e_{it} e_{jt}) \\ &= \frac{1}{NT} \sum_{t=1}^T \sum_{h=1}^N \sum_{j=1}^N \sum_{q=1}^N (r_{ih,t} r_{jq,t} \varepsilon_{ht} \varepsilon_{qt}) = \frac{1}{NT} \sum_{t=1}^T \sum_{h=1}^N \sum_{q=1}^N (r_{ih,t} r_{.q,t} \varepsilon_{ht} \varepsilon_{qt}), \end{aligned}$$



where  $r_{.q,t} = \sum_{j=1}^N r_{jq,t}$ . Its mean is

$$\begin{aligned} E(T^{-1} \mathbf{e}'_i \bar{\mathbf{e}}) &= \frac{1}{NT} \sum_{t=1}^T \sum_{h=1}^N \sum_{q=1}^N [r_{ih,t} r_{.q,t} E(\varepsilon_{ht} \varepsilon_{qt})] \\ &= \frac{1}{NT} \sum_{t=1}^T \sum_{q=1}^N (r_{iq,t} r_{.q,t} \sigma_q^2) \leq K \frac{1}{NT} \sum_{t=1}^T \sum_{q=1}^N r_{iq,t} = O\left(\frac{1}{N}\right), \end{aligned}$$

since  $\sum_{q=1}^N r_{iq,t} = O(1)$ . Also, under Assumptions 1 and 2, we have

$$\begin{aligned} E\left[(T^{-1} \mathbf{e}'_i \bar{\mathbf{e}})^2\right] &= \frac{1}{N^2 T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{h=1}^N \sum_{\ell=1}^N \sum_{q=1}^N \sum_{p=1}^N [r_{ih,t} r_{ih,s} r_{.q,t} r_{.p,s} E(\varepsilon_{ht} \varepsilon_{\ell s} \varepsilon_{qt} \varepsilon_{ps})] \\ &= \frac{1}{N^2 T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{h=1}^N \sum_{\ell=1}^N [r_{ih,t} r_{ih,s} r_{.h,t} r_{.h,s} E(\varepsilon_{ht}^2 \varepsilon_{\ell s}^2)] \\ &\quad + \frac{1}{N^2 T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{h=1}^N \sum_{\ell=1, \ell \neq h}^N [r_{ih,t} r_{ih,s} r_{.t} r_{.s} E(\varepsilon_{ht} \varepsilon_{hs} \varepsilon_{\ell s} \varepsilon_{\ell t})] \\ &\quad + \frac{1}{N^2 T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{h=1}^N \sum_{\ell=1, \ell \neq h}^N [r_{ih,t} r_{ih,s} r_{.t} r_{.h,s} E(\varepsilon_{ht} \varepsilon_{hs} \varepsilon_{\ell s} \varepsilon_{\ell t})] \\ &\leq K \frac{1}{N^2 T^2} \sum_{t=1}^T \sum_{s=1}^T \sum_{h=1}^N \sum_{\ell=1}^N (|r_{ih,t} r_{ih,s} r_{.h,t} r_{.h,s}| + |r_{ih,t} r_{ih,s} r_{.t} r_{.s}| + |r_{ih,t} r_{ih,s} r_{.t} r_{.h,s}|) \\ &= O\left(\frac{1}{N^2}\right) + O\left(\frac{1}{NT}\right), \end{aligned}$$

since the first and the third terms of the above are  $O\left(\frac{1}{N^2}\right)$ , while, denoting  $r_{.t}^* = \max_{1 \leq \ell \leq T} \{r_{.t,\ell}\} = O(1)$ , the second term satisfies

$$\begin{aligned} \frac{1}{N^2 T^2} \sum_{h=1}^N \sum_{\ell=1}^N \sum_{t=1}^T \sum_{s=1}^T |r_{ih,t} r_{ih,s} r_{.t} r_{.s}| &\leq \frac{1}{N^2 T^2} \sum_{\ell=1}^N |r_{.t}^*| \sum_{h=1}^N \sum_{t=1}^T \sum_{s=1}^T (|r_{ih,t}| |r_{ih,s}| |r_{.t,s}|) \\ &\leq \frac{K}{N^2 T^2} \sum_{\ell=1}^N |r_{.t}^*| \sum_{h=1}^N \sum_{s=1}^T (|r_{ih,s}| |r_{.t,s}|) \\ &\leq \frac{K}{N^2 T^2} \sum_{\ell=1}^N |r_{.t}^*| \sum_{h=1}^N \sum_{s=1}^T |r_{ih,s}| \\ &\leq \frac{K}{N^2 T} \sum_{\ell=1}^N |r_{.t}^*| = O\left(\frac{1}{NT}\right). \end{aligned}$$

since  $N^{-2} \sum_{\ell=1}^N |r_{.t}^*| = O(N^{-1})$ . It follows that  $T^{-1} \mathbf{e}'_i \bar{\mathbf{e}} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$ . ■

The above results can be used to prove further results that are helpful in deriving the asymptotic distribution of CCE estimators. Rewrite equations (1) and (19) more compactly as

$$\mathbf{z}_{it} = \begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \mathbf{B}'_i \mathbf{d}_t + \mathbf{C}'_i \mathbf{f}_t + \boldsymbol{\xi}_{it}, \quad (\text{A.7})$$

where

$$\begin{aligned} \mathbf{B}_i &= \begin{pmatrix} \boldsymbol{\alpha}_i & \mathbf{A}_i \end{pmatrix} \mathbf{D}_i, \quad \mathbf{C}_i = \begin{pmatrix} \gamma_i & \boldsymbol{\Gamma}_i \end{pmatrix} \mathbf{D}_i, \\ \mathbf{D}_i &= \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{\beta}_i & \mathbf{I}_k \end{pmatrix}, \quad \boldsymbol{\xi}_{it} = \begin{pmatrix} e_{it} + \boldsymbol{\beta}'_i \mathbf{v}_{it} \\ \mathbf{v}_{it} \end{pmatrix}. \end{aligned}$$

From Lemma A.2 it follows that (see also Lemmas 2 and 3 in Pesaran (2006))

$$\frac{\bar{\xi}'\bar{\xi}}{T} = O_p\left(\frac{1}{N}\right), \quad (\text{A.8})$$

$$\frac{\mathbf{F}'\bar{\xi}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \frac{\mathbf{D}'\bar{\xi}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{A.9})$$

$$\frac{\mathbf{V}'_i\bar{\xi}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad \frac{\mathbf{e}'_i\bar{\xi}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{A.10})$$

$$\frac{\mathbf{X}'_i\bar{\xi}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \quad (\text{A.11})$$

Under Assumption 9, the above results in turn yield:

$$\frac{\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{F}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{A.12})$$

$$\frac{\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{X}_i}{T} = \frac{\mathbf{X}'_i\mathbf{M}_g\mathbf{X}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{A.13})$$

$$\frac{\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{e}_i}{T} = \frac{\mathbf{X}'_i\mathbf{M}_g\mathbf{e}_i}{T} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \quad (\text{A.14})$$

where  $\mathbf{M}_g = \mathbf{I}_T - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$ . Note that (A.12)-(A.14) are identical to relations (40), (43) and (44) in Pesaran (2006), and will be used to derive the asymptotic distribution of CCE Pooled and CCE Mean Group estimators.

In what follows we sketch the proofs of Theorems 1-2 and 5-6.

**Proof of Theorem 1.** Consider (6), and rewrite it as

$$\sqrt{N}\left(\hat{\beta}_{MG} - \beta\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbf{v}_i + \frac{1}{\sqrt{T}}\mathbf{h}_{NT},$$

where

$$\begin{aligned} \mathbf{h}_{NT} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N (T^{-1}\mathbf{X}'_i\mathbf{M}_D\mathbf{X}_i)^{-1}\mathbf{X}'_i\mathbf{M}_D\mathbf{e}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{W}'_i\mathbf{e}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \mathbf{w}_{it}e_{it} \\ &= \frac{1}{\sqrt{NT}}\sum_{t=1}^T \mathbf{W}'_{.t}\mathbf{e}_{.t} = \frac{1}{\sqrt{NT}}\mathbf{H}'\boldsymbol{\varepsilon}, \end{aligned}$$

with  $\mathbf{W}'_{.t} = (\mathbf{w}_{1t}, \mathbf{w}_{2t}, \dots, \mathbf{w}_{Nt})$ . Under Assumption 1, the  $NT \times 1$  vector  $\boldsymbol{\varepsilon}$  is a zero mean covariance stationary process, with covariance matrix  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega}_{\varepsilon\varepsilon}$ . Since the elements of  $\mathbf{H}$  are uniformly bounded, using standard results on double array central limit theorem for stationary processes (see, for example, Chung (2001), Chapter 7), it follows that  $\mathbf{h}_{NT}$  is asymptotically normally distributed if its variance exists and is positive semi-definite. Note that  $\boldsymbol{\Omega}_{\varepsilon\varepsilon}$  is made of  $T^2$  blocks of  $N \times N$  diagonal matrices with elements

$$\gamma_i(s) = \gamma_i(-s) = \sum_{j=0}^{\infty} a_{ij}a_{i,j+|s|}, \quad \text{for } i = 1, 2, \dots, N, \quad s = 0, 1, 2, \dots,$$

that are absolute summable, namely  $\sum_{s=0}^{\infty} |\gamma_i(s)| < K$  for all  $i$ . It follows that  $\boldsymbol{\Omega}_{\varepsilon\varepsilon}$  has bounded row and column norms and the variance of  $\mathbf{h}_{NT}$  satisfies<sup>5</sup>

$$\begin{aligned} \text{Var}(\mathbf{h}_{NT}) &= \frac{1}{NT} \mathbf{H}' \boldsymbol{\Omega}_{\varepsilon\varepsilon} \mathbf{H} \leq \frac{1}{NT} \lambda_1(\boldsymbol{\Omega}_{\varepsilon\varepsilon}) (\mathbf{H}' \mathbf{H}) \leq \frac{1}{NT} K_1 (\mathbf{H}' \mathbf{H}) \\ &= \frac{1}{NT} K_1 \sum_{t=1}^T (\mathbf{W}'_{\cdot t} \mathbf{R}_t \mathbf{R}'_t \mathbf{W}_{\cdot t}) \leq \frac{1}{NT} K_1 \sum_{t=1}^T (\mathbf{W}'_{\cdot t} \mathbf{W}_{\cdot t}) \lambda_1(\mathbf{R}_t \mathbf{R}'_t) \\ &\leq K_1 K_2 \frac{1}{N} \sum_{i=1}^N (T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i)^{-1}, \end{aligned} \tag{A.15}$$

which, under Assumption 5(a), tends to a non-singular matrix with finite elements. Hence, we have

$$\sqrt{N} (\hat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + \frac{1}{\sqrt{T}} \mathbf{h}_{NT} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + O_p\left(\frac{1}{\sqrt{T}}\right),$$

which proves the theorem. ■

**Proof of Theorem 2.** Consider (7), and let

$$\begin{aligned} \mathbf{q}_{NT} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_D \mathbf{e}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \tilde{\mathbf{X}}'_i \mathbf{e}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \tilde{\mathbf{x}}_{it} e_{it} \\ &= \frac{1}{\sqrt{NT}} \sum_{t=1}^T \tilde{\mathbf{X}}'_{\cdot t} \mathbf{R}_t \boldsymbol{\varepsilon}_t = \frac{1}{\sqrt{NT}} \mathbf{P}' \boldsymbol{\varepsilon}, \end{aligned}$$

where  $\tilde{\mathbf{X}}_{\cdot t} = (\tilde{\mathbf{x}}_{1t}, \tilde{\mathbf{x}}_{2t}, \dots, \tilde{\mathbf{x}}_{Nt})'$ , and  $\mathbf{P}' = (\tilde{\mathbf{X}}'_{\cdot 1} \mathbf{R}_1, \tilde{\mathbf{X}}'_{\cdot 2} \mathbf{R}_2, \dots, \tilde{\mathbf{X}}'_{\cdot T} \mathbf{R}_T)$ . Following similar lines of reasoning as for (A.15), the variance of  $\mathbf{q}_{NT}$  satisfies

$$\begin{aligned} \text{Var}(\mathbf{q}_{NT}) &= \left( \frac{1}{NT} \mathbf{P}' \boldsymbol{\Omega}_{\varepsilon\varepsilon} \mathbf{P} \right) \leq K_1 \left( \frac{\mathbf{P}' \mathbf{P}}{NT} \right) \\ &= K_1 \left( \frac{1}{NT} \sum_{t=1}^T \tilde{\mathbf{X}}'_{\cdot t} \mathbf{R}_t \mathbf{R}'_t \tilde{\mathbf{X}}_{\cdot t} \right) \leq K_1 K_2 \frac{1}{T} \sum_{t=1}^T \left( \frac{\tilde{\mathbf{X}}'_{\cdot t} \tilde{\mathbf{X}}_{\cdot t}}{N} \right) \\ &= K_1 K_2 \frac{1}{NT} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i, \end{aligned} \tag{A.16}$$

which tends, under Assumption 5(b), to a finite, positive definite matrix. It follows that

$$\begin{aligned} \sqrt{N} (\hat{\boldsymbol{\beta}}_P - \boldsymbol{\beta}) &= \left( N^{-1} \sum_{i=1}^N T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i \right)^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^N T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i \mathbf{v}_i + \frac{1}{\sqrt{T}} \mathbf{q}_{NT} \right) \\ &= \left( N^{-1} \sum_{i=1}^N T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N T^{-1} \mathbf{X}'_i \mathbf{M}_D \mathbf{X}_i \mathbf{v}_i + O_p\left(\frac{1}{\sqrt{T}}\right), \end{aligned}$$

which proves the theorem. ■

**Proof of Theorem 5.** Consider

$$\sqrt{N} (\hat{\boldsymbol{\beta}}_{CCEMG} - \boldsymbol{\beta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{iT}^{-1} \left( \frac{\sqrt{N} \mathbf{X}'_i \bar{\mathbf{M}} \mathbf{F}}{T} \right) \gamma_i + \frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{\boldsymbol{\Psi}}_{iT}^{-1} \left( \frac{\mathbf{X}'_i \bar{\mathbf{M}} \mathbf{e}_i}{T} \right),$$

<sup>5</sup>We make use of the following result. Let  $\mathbf{A}$  be an  $n \times n$  symmetric matrix, and  $\mathbf{B}$  be an  $n \times m$  matrix. Then  $(\mathbf{B}' \mathbf{B}) \lambda_1(\mathbf{A}) - \mathbf{B}' \mathbf{A} \mathbf{B}$  is a positive semi-definite matrix (see Bernstein (2005), pp. 264 and 271).

where, by Assumption 10(b),  $\hat{\Psi}_{iT}^{-1} = (T^{-1}\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{X}_i)^{-1}$  exists for all  $i$ . First note that, using (A.12), and since, by Assumption 7, factor loadings are bounded, it follows that (see also Pesaran (2006), p. 983)

$$N^{-1} \sum_{i=1}^N \hat{\Psi}_{iT}^{-1} \left( \frac{\sqrt{N}\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{F}}{T} \right) \gamma_i = O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right), \quad \text{as } N \rightarrow \infty. \quad (\text{A.17})$$

Further, given (A.13)-(A.14) we have

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{\Psi}_{iT}^{-1} \left( \frac{\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{e}_i}{T} \right) &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left( \frac{\mathbf{V}'_i\mathbf{M}_g\mathbf{V}_i}{T} \right)^{-1} \left( \frac{\mathbf{V}'_i\mathbf{M}_g\mathbf{e}_i}{T} \right) + O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right) \\ &= \frac{1}{\sqrt{T}} \mathbf{h}_{NT} + O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right), \end{aligned}$$

where

$$\begin{aligned} \mathbf{h}_{NT} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \left( \frac{\mathbf{V}'_i\mathbf{M}_g\mathbf{V}_i}{T} \right)^{-1} (\mathbf{V}'_i\mathbf{M}_g\mathbf{e}_i) \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{w}_{it} e_{it} = \frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbf{W}'_{.t} \mathbf{e}_{.t} = \frac{1}{\sqrt{NT}} \mathbf{H}' \boldsymbol{\varepsilon}, \end{aligned}$$

where  $\mathbf{H}' = (\mathbf{W}'_{.1}\mathbf{R}_1, \mathbf{W}'_{.2}\mathbf{R}_2, \dots, \mathbf{W}'_{.T}\mathbf{R}_T)$ ,  $\mathbf{W}'_{.t} = (\mathbf{w}_{1t}, \mathbf{w}_{2t}, \dots, \mathbf{w}_{Nt})$ , and  $\mathbf{w}_{it}$  is the  $t^{\text{th}}$  column of  $\mathbf{W}'_i = (T^{-1}\mathbf{V}'_i\mathbf{M}_g\mathbf{V}_i)^{-1} \mathbf{V}'_i\mathbf{M}_g$ . Using similar lines of reasoning as in the proof of Theorem 1,  $\mathbf{h}_{NT}$  has zero mean and its variance satisfies

$$\begin{aligned} \text{Var}(\mathbf{h}_{NT}) &= \frac{1}{NT} E(\mathbf{H}'\boldsymbol{\Omega}_{\varepsilon\varepsilon}\mathbf{H}) \leq K_1K_2 \frac{1}{NT} \sum_{i=1}^N E(\mathbf{W}'_i\mathbf{W}_i) \\ &= K_1K_2 \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{v_i}, \end{aligned}$$

which, under Assumption 10(a) tends to a finite, positive definite matrix. Therefore, we have

$$\begin{aligned} \sqrt{N}(\hat{\beta}_{CCEMG} - \beta) &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + \frac{1}{\sqrt{T}} \mathbf{h}_{NT} + O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{v}_i + O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right), \end{aligned} \quad (\text{A.18})$$

which proves the theorem. ■

**Proof of Theorem 6.** Consider

$$\sqrt{N}(\hat{\beta}_{CCCEP} - \beta) = \left( \sum_{i=1}^N \frac{1}{N} \frac{\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{X}_i}{T} \right)^{-1} \sum_{i=1}^N \frac{1}{\sqrt{N}} \frac{\mathbf{X}'_i\bar{\mathbf{M}}(\mathbf{X}_i\mathbf{v}_i + \mathbf{F}\gamma_i + \mathbf{e}_i)}{T},$$

Using (A.12)-(A.14) we have

$$\left( \sum_{i=1}^N \frac{1}{N} \frac{\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{X}_i}{T} \right)^{-1} \sum_{i=1}^N \frac{1}{\sqrt{N}} \frac{\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{F}}{T} \gamma_i = O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right). \quad (\text{A.19})$$

Further,

$$\sum_{i=1}^N \frac{1}{\sqrt{N}} \frac{\mathbf{X}'_i\bar{\mathbf{M}}\mathbf{e}_i}{T} = \sum_{i=1}^N \frac{1}{\sqrt{N}} \frac{\mathbf{V}'_i\mathbf{M}_g\mathbf{e}_i}{T} + O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{\sqrt{T}} \right).$$

Let

$$\mathbf{q}_{NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \tilde{\mathbf{v}}_i' \mathbf{e}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \tilde{v}_{it} e_{it} = \frac{1}{\sqrt{NT}} \sum_{t=1}^T \tilde{\mathbf{v}}_t' \mathbf{R}_t \boldsymbol{\varepsilon}_t = \frac{1}{\sqrt{NT}} \mathbf{P}' \boldsymbol{\varepsilon},$$

where  $\tilde{\mathbf{V}}_i = \mathbf{M}_g \mathbf{V}_i$ ,  $\tilde{\mathbf{V}}_t = (\tilde{v}_{1t}, \tilde{v}_{2t}, \dots, \tilde{v}_{Nt})'$ ,  $\mathbf{P}' = (\tilde{\mathbf{V}}_1' \mathbf{R}_1, \tilde{\mathbf{V}}_2' \mathbf{R}_2, \dots, \tilde{\mathbf{V}}_T' \mathbf{R}_T)$ . Following similar lines of reasoning developed in the proofs to Theorem 2,  $\mathbf{q}_{NT}$  has mean zero and its variance satisfies

$$\begin{aligned} \text{Var}(\mathbf{q}_{NT}) &= \frac{1}{NT} E(\mathbf{P}' \boldsymbol{\Omega}_{\varepsilon\varepsilon} \mathbf{P}) \leq \frac{1}{NT} K_1 K_2 \sum_{i=1}^N E(\tilde{\mathbf{V}}_i' \tilde{\mathbf{V}}_i) \\ &= K_1 K_2 \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{v_i}, \end{aligned}$$

which, by Assumption 10(a), tends to a finite, positive definite matrix. It follows that

$$\begin{aligned} \sqrt{N}(\hat{\boldsymbol{\beta}}_{CCEP} - \boldsymbol{\beta}) &= \left( \sum_{i=1}^N \frac{1}{N} \frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i}{T} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i \mathbf{v}_i}{T} \\ &\quad + \left( \sum_{i=1}^N \frac{1}{N} \frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i}{T} \right)^{-1} \frac{1}{\sqrt{T}} \mathbf{q}_{NT} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &= \left( \sum_{i=1}^N \frac{1}{N} \frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i}{T} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\mathbf{X}_i' \mathbf{M}_g \mathbf{X}_i \mathbf{v}_i}{T} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned} \quad (\text{A.20})$$

which proves the theorem. ■

## References

- Andrews, D. (2005). Cross section regression with common shocks. *Econometrica* 73, 1551–1585.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.
- Bakirov, N. K. and G. J. Székely (2006). Student's t-test for gaussian scale mixtures. *Journal of Mathematical Sciences* 139, 6497–6505.
- Baltagi, B., S. Song, and W. Koh (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117, 123–150.
- Baltagi, B. H., P. Egger, and M. Pfaffermayr (2009). A generalized spatial panel data model with random effects. Center for Policy Research Working Paper No. 113.
- Bernstein, D. S. (2005). *Matrix mathematics: theory, facts, and formulas with application to linear systems theory*. Princeton University Press.
- Bester, C. A., T. G. Conley, and C. B. Hansen (2009). Inference with dependent data using cluster covariance estimators. Mimeo, University of Chicago.
- Chudik, A., M. H. Pesaran, and E. Tosetti (2010). Weak and strong cross section dependence and estimation of large panels. Mimeo, University of Cambridge.
- Chung, K. L. (2001). *A Course in Probability Theory*. Academic Press.
- Coakley, J., A. M. Fuertes, and R. Smith (2002). A principal components approach to cross-section dependence in panels. Birkbeck College Discussion Paper 01/2002.
- Coakley, J., A. M. Fuertes, and R. Smith (2006). Unobserved heterogeneity in panel time series. *Computational Statistics and Data Analysis* 50, 2361–2380.
- Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics* 92, 1–45.

- Conley, T. G. and G. Topa (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics* 17, 303–327.
- Cooper, R. and J. Haltiwanger (1996). Evidence on macroeconomic complementarities. *The Review of Economics and Statistics* 78, 78–93.
- Cowan, R., W. Cowan, and G. M. P. Swann (2004). Waves in consumption with interdependence among consumers. *Canadian Journal of Economics* 37, 149–177.
- Driscoll, J. C. and A. C. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics* 80, 549–560.
- Egger, P., M. Pfaffermayr, and H. Winner (2005). An unbalanced spatial panel data approach to us state tax competition. *Economics Letters* 88, 329–335.
- Fingleton, B. (2008). A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spatial Economic Analysis* 3, 27–44.
- Fingleton, B. and J. Le Gallo (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science* 87, 319–339.
- Ibragimov, R. and U. K. Müller (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics* Forthcoming.
- Kapetanios, G. and M. H. Pesaran (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: Cambridge University Press.
- Kapetanios, G., M. H. Pesaran, and T. Yagamata (2009). Panels with nonstationary multifactor error structures. CESifo Working Paper No. 1788, revised April 2010.
- Kapoor, M., H. H. Kelejian, and I. Prucha (2007). Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97–130.
- Kelejian, H. H. and I. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–533.
- Kelejian, H. H. and I. Prucha (2009). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. forthcoming, *Journal of Econometrics*.
- Kelejian, H. H. and I. R. Prucha (2007). HAC estimation in a spatial framework. *Journal of Econometrics* 140, 131–154.
- Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899–1925.
- Lee, L. F. (2007). Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489–514.
- Lee, L. F. and J. Yu (2008). Estimation of spatial autoregressive panel data models with fixed effects. Mimeo, Ohio State University.
- Mardia, K. V. and R. J. Marshall (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* 71, 135–146.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Onatski, A. (2009). Asymptotic distribution of the principal components estimator of large factor models when factors are relatively weak. Mimeo, Columbia University.
- Paul, D. (2007). Asymptotics of the leading sample eigenvalues for a spiked covariance model. *Statistica Sinica* 17, 1617–1642.
- Pesaran, M. H. (2006). Estimation and inference in large heterogenous panels with multifactor error structure. *Econometrica* 74, 967–1012.
- Pesaran, M. H., T. Schuermann, and S. Weiner (2004). Modelling regional interdependencies using a global error-correcting macroeconomic model. *Journal of Business and Economic Statistics* 22, 129–162.
- Pesaran, M. H. and R. P. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68, 79–113.

- Pesaran, M. H. and E. Tosetti (2009). Large panels with common factors and spatial correlation. CESifo Working Paper Series No. 2103.
- Phillips, P. C. B. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal* 6, 217–259.
- Pinkse, J., M. Slade, and C. Brett (2002). Spatial price competition: a semiparametric approach. *Econometrica* 70, 1111–1153.
- Robertson, D. and J. Symons (2000). Factor residuals in SUR regressions: Estimating panels allowing for cross sectional correlation,. Mimeo, University of Cambridge.
- Robertson, D. and J. Symons (2007). Maximum likelihood factor analysis with rank-deficient sample covariance matrices. *Journal of Multivariate Analysis* 98, 813–828.
- Smith, J. and M. McAleer (1994). Newey-west covariance matrix estimates for models with generated regressors. *Applied Economics* 26, 635–640.
- Ullah, A. (2004). *Finite Sample Econometrics*. Oxford: Oxford University Press.
- Whittle, P. (1954). On stationary processes on the plane. *Biometrika* 41, 434–449.
- Yu, J., R. de Jong, and L. F. Lee (2007). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large: A nonstationary case. Mimeo, Ohio State University.
- Yu, J., R. de Jong, and L. F. Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large. *Journal of Econometrics* 146, 118–137. Mimeo, The Ohio State University.
- Yu, J. and L. F. Lee (2007). Estimation of unit root spatial dynamic panel data models. Mimeo, Ohio State University.

Table A1: Small sample properties of estimators for panels with spatially correlated errors ( $\delta = 0.4$ ) and no unobserved common factors, under slope heterogeneity.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	-0.07	0.10	-0.15	-0.38	8.40	6.59	5.72	5.14	0.07	-0.11	0.01	-0.36	7.78	6.61	5.92	5.65
30	-0.07	-0.29	0.09	-0.07	6.40	5.44	4.52	4.03	-0.06	-0.13	0.06	-0.13	6.02	5.51	4.82	4.38
50	-0.12	0.09	0.10	0.16	4.94	4.36	3.54	3.09	-0.26	0.15	0.05	0.23	4.74	4.29	3.91	3.46
100	0.04	0.09	0.01	0.03	3.38	2.89	2.48	2.31	-0.05	0.01	-0.04	-0.01	3.35	3.02	2.74	2.65
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	0.06	0.02	0.06	-0.27	7.70	6.57	5.91	5.67	0.05	0.09	0.02	-0.31	7.89	6.74	5.94	5.63
30	-0.18	-0.15	0.05	-0.11	5.95	5.42	4.75	4.40	-0.10	-0.17	0.01	-0.08	6.01	5.44	4.71	4.40
50	-0.23	0.32	0.13	0.22	4.57	4.25	3.81	3.38	-0.23	0.30	0.13	0.24	4.71	4.28	3.90	3.43
100	0.00	0.02	-0.01	-0.03	3.32	2.90	2.70	2.65	-0.02	0.01	-0.01	-0.04	3.38	2.90	2.75	2.66
	CCE Mean group								CCE Pooled							
20	0.26	0.15	-0.12	-0.32	10.24	7.61	6.12	5.33	0.40	0.20	-0.23	-0.27	8.92	7.29	6.07	5.40
30	-0.31	-0.23	0.09	-0.07	8.35	6.05	4.98	4.22	-0.24	-0.19	0.10	0.00	7.19	5.86	5.00	4.30
50	-0.33	0.27	0.12	0.20	6.24	5.02	3.81	3.25	-0.38	0.25	0.07	0.22	5.42	4.68	3.80	3.30
100	0.01	0.03	0.02	0.02	4.44	3.44	2.66	2.41	-0.04	-0.01	-0.01	0.01	3.89	3.25	2.67	2.46
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	7.70	6.70	6.50	7.60	13.60	14.40	16.10	16.40	8.90	6.30	7.10	7.10	13.80	14.30	17.20	17.40
30	7.40	6.60	6.00	6.80	15.40	20.10	20.50	21.00	6.40	6.50	5.00	6.40	15.10	18.80	21.90	22.50
50	5.50	4.90	5.40	5.20	17.30	24.30	27.40	31.00	4.90	6.70	5.10	4.70	16.60	25.70	29.00	36.60
100	5.70	5.60	5.50	5.20	33.20	38.80	43.10	51.80	5.10	6.20	5.10	5.30	31.10	42.20	51.10	60.40
	SHAC Mean group								SHAC Pooled							
20	11.10	11.90	17.80	27.30	18.40	23.20	31.80	46.10	15.40	17.70	20.90	32.40	22.10	26.40	35.10	46.80
30	10.70	11.80	16.50	23.70	20.10	27.20	39.10	54.80	13.00	17.20	20.30	27.70	24.50	33.00	43.40	53.30
50	7.20	11.90	15.30	22.50	23.70	36.00	52.50	68.30	12.00	16.00	20.90	26.80	28.20	41.40	52.50	68.00
100	8.80	11.10	14.60	25.90	40.00	56.40	72.60	85.10	12.10	14.40	17.80	31.60	46.50	57.90	71.20	81.00
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	11.60	11.90	18.40	27.30	17.20	22.60	31.60	45.60	15.10	18.00	21.40	31.20	21.40	26.00	35.60	47.20
30	11.00	11.90	15.70	23.70	19.40	27.60	38.90	53.20	13.70	16.80	19.30	28.00	24.30	32.90	42.90	52.80
50	7.90	11.60	14.30	22.00	23.20	35.40	52.10	67.80	11.70	15.80	19.80	26.70	27.60	42.10	51.60	68.00
100	8.80	10.40	14.60	26.00	40.50	56.20	73.10	84.90	11.90	14.10	17.70	31.00	47.70	58.30	71.90	81.10
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	27.50	30.80	39.30	52.50	39.10	42.80	52.90	65.80	25.60	29.90	35.40	50.20	34.80	41.00	50.10	64.00
30	25.00	31.40	36.90	48.40	36.90	48.50	59.70	70.40	21.80	29.40	33.90	46.30	36.70	43.60	57.00	67.80
50	22.30	31.30	36.40	48.00	45.00	59.50	69.20	82.00	23.40	30.30	37.40	47.90	43.20	58.50	68.20	81.00
100	25.50	30.90	37.40	55.00	65.60	74.20	85.00	91.10	25.10	27.80	36.00	52.80	63.80	72.90	84.00	89.60
	CCE Mean group								CCE Pooled							
20	7.60	7.00	7.80	8.80	9.60	12.80	16.00	18.90	7.50	8.40	7.70	8.50	11.60	13.30	15.50	20.50
30	6.90	5.30	5.90	6.70	11.80	13.90	19.50	23.50	5.90	5.10	5.50	7.10	12.50	14.80	19.50	23.10
50	5.10	7.50	5.40	5.10	11.80	23.20	27.70	37.60	5.20	5.60	5.30	5.20	13.70	24.90	26.40	36.20
100	5.10	5.70	4.70	5.60	20.50	32.60	44.70	58.10	5.70	5.30	5.40	5.50	24.60	34.70	43.90	58.00

Notes: Mean Group, Pooled, CCE Mean group and CCE Pooled are (3), (5), (21), (23), respectively. Robust variances of Mean group and Pooled estimators are (10) and (11). SHAC variances of Mean group and Pooled estimators are given by (17) and (18).

Variances of CCE Mean group and CCE Pooled estimators are given by (25) and (26).

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.



Table A2: Small sample properties of estimators for panels with spatially correlated errors ( $\delta = 0.4$ ) and no unobserved common factors, under slope homogeneity.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	-0.10	0.04	-0.03	-0.05	6.25	4.81	3.43	2.30	0.01	0.06	0.02	-0.04	4.98	3.92	2.87	2.03
30	0.19	0.05	-0.08	-0.03	5.25	3.86	2.87	1.89	0.02	0.03	-0.08	0.00	4.36	3.32	2.59	1.74
50	-0.26	0.04	-0.01	-0.03	4.31	3.15	2.27	1.48	-0.11	-0.01	0.02	-0.07	3.21	2.65	1.92	1.32
100	-0.02	0.11	-0.02	-0.04	2.71	2.19	1.47	1.01	-0.05	0.03	-0.03	-0.03	2.16	1.85	1.30	0.92
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	0.06	0.11	0.06	0.03	4.39	3.37	2.47	1.76	-0.04	0.09	0.08	0.01	4.68	3.57	2.63	1.84
30	-0.01	0.04	-0.09	-0.06	3.73	2.85	2.19	1.45	0.04	0.04	-0.09	-0.06	3.94	3.01	2.28	1.54
50	-0.03	0.09	0.01	-0.03	2.90	2.34	1.69	1.11	0.02	0.07	0.02	-0.03	3.04	2.45	1.75	1.16
100	-0.03	-0.02	0.00	-0.04	1.87	1.56	1.11	0.76	-0.03	-0.01	-0.02	-0.04	2.00	1.68	1.18	0.80
	CCE Mean group								CCE Pooled							
20	-0.25	0.15	0.05	-0.03	8.70	6.19	4.23	2.55	0.05	0.11	0.11	-0.02	7.21	5.38	3.88	2.44
30	0.46	0.19	-0.08	-0.09	7.19	5.07	3.50	2.24	0.29	0.04	-0.08	-0.12	5.75	4.26	3.19	2.18
50	-0.18	0.20	-0.03	0.01	5.85	4.00	2.70	1.72	-0.25	0.17	-0.03	0.00	4.49	3.44	2.45	1.63
100	0.00	0.13	-0.10	-0.04	3.85	2.72	1.80	1.19	-0.03	0.11	-0.10	-0.04	3.03	2.39	1.67	1.13
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	6.10	5.40	5.20	6.80	20.50	27.40	43.50	69.80	6.00	6.00	7.70	5.90	15.80	21.20	33.50	59.50
30	9.30	7.20	9.00	7.70	28.60	37.90	57.50	84.30	7.60	6.50	6.70	7.70	20.30	30.70	46.10	76.40
50	5.80	6.30	6.40	6.70	35.00	52.70	75.10	96.40	5.80	6.50	6.70	5.90	22.40	39.10	62.30	91.10
100	5.50	8.80	6.60	7.30	64.70	83.60	97.00	100.00	5.00	7.40	4.40	5.30	45.10	68.60	90.90	99.90
	SHAC Mean group								SHAC Pooled							
20	7.50	6.80	7.00	5.20	14.80	22.00	33.80	61.30	7.00	6.10	5.20	5.70	21.90	29.40	44.40	70.90
30	8.30	6.40	7.30	6.90	20.80	30.70	46.60	76.70	9.40	6.60	7.70	7.10	27.60	39.10	56.60	82.60
50	7.10	6.50	7.40	6.50	23.60	40.70	64.40	92.40	7.30	7.10	6.20	6.20	35.70	56.00	75.60	96.70
100	4.80	7.60	5.00	4.70	47.30	69.40	91.80	99.90	6.60	8.60	5.50	5.70	64.80	82.20	96.50	100.00
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	6.40	6.80	7.00	4.70	15.70	22.20	34.00	61.20	6.80	6.10	5.50	5.20	22.40	30.10	44.60	71.40
30	7.90	6.80	7.60	7.00	20.60	32.30	47.30	77.50	9.50	7.00	7.90	6.80	29.90	40.30	57.80	83.80
50	7.00	6.50	6.70	6.30	23.90	40.70	63.80	92.30	7.10	7.40	6.50	5.80	36.60	56.00	76.00	96.90
100	4.70	7.80	4.40	5.50	47.20	70.30	92.20	99.90	6.70	8.40	5.90	6.40	66.80	83.30	97.00	100.00
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	8.00	6.20	6.10	5.90	26.50	37.20	56.20	84.70	7.30	5.10	5.40	4.70	23.30	32.80	48.30	78.00
30	8.30	5.40	5.90	5.00	34.00	46.90	65.60	93.20	7.50	6.30	5.90	4.90	32.60	42.60	61.10	89.20
50	8.30	8.70	6.30	4.40	46.60	67.10	85.70	99.40	7.70	7.70	6.10	4.70	44.50	63.40	82.90	98.80
100	6.20	7.10	5.70	4.80	80.00	91.50	99.60	100.00	7.30	8.30	5.90	4.90	75.50	88.80	98.40	100.00
	CCE Mean group								CCE Pooled							
20	6.40	6.70	7.30	6.10	10.80	18.30	26.60	48.80	6.50	7.00	6.60	5.90	13.20	20.20	30.00	52.40
30	5.90	6.30	6.10	7.80	14.40	19.80	33.70	63.60	6.40	5.70	6.30	8.10	15.90	21.40	37.70	66.10
50	6.90	6.50	5.80	5.80	15.30	27.60	47.70	83.80	6.20	6.30	5.90	5.80	18.80	34.60	53.80	87.10
100	4.40	4.80	5.30	6.20	23.10	46.80	75.60	98.20	4.80	6.20	5.80	6.10	35.40	58.50	82.60	99.30

Notes: see notes to Table A1.

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.

Table B1: Small sample properties of estimators for panels with spatially correlated errors ( $\delta = 0.4$ ) and unobserved common factors, under slope heterogeneity.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	15.38	14.69	14.86	14.12	20.02	18.48	17.80	16.59	17.12	16.83	17.05	16.63	21.37	20.43	19.83	19.03
30	15.26	14.50	15.00	14.32	19.35	17.82	17.52	16.06	17.46	17.11	17.70	17.17	21.60	20.41	20.24	18.98
50	15.86	15.97	15.99	15.75	18.62	17.86	17.48	16.83	18.00	18.13	18.49	18.45	21.00	20.31	20.21	19.60
100	15.23	14.76	15.23	14.77	17.37	16.33	16.25	15.53	16.87	16.60	17.29	17.13	19.48	18.56	18.55	18.01
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	2.23	2.17	1.75	1.17	8.94	7.90	7.08	6.73	1.41	1.36	0.94	0.44	8.62	7.70	6.76	6.42
30	1.95	1.88	1.93	1.63	7.40	6.67	6.28	5.88	1.27	0.84	1.09	0.83	6.99	6.25	5.74	5.46
50	1.56	2.00	1.53	1.80	5.63	5.36	4.86	4.58	0.76	1.12	0.65	0.94	5.29	4.94	4.44	4.17
100	1.56	1.44	1.52	1.43	4.25	3.85	3.59	3.50	0.71	0.60	0.68	0.64	4.00	3.48	3.24	3.19
	CCE Mean group								CCE Pooled							
20	0.54	0.16	-0.10	-0.34	10.09	7.54	6.07	5.44	0.51	0.00	-0.23	-0.34	9.06	7.46	6.17	5.62
30	-0.43	-0.08	0.13	-0.10	8.00	6.05	4.94	4.32	-0.39	-0.17	0.08	-0.06	7.22	5.97	5.03	4.40
50	-0.27	0.15	0.09	0.21	6.10	4.98	3.85	3.21	-0.28	0.23	0.06	0.22	5.35	4.73	3.84	3.28
100	0.00	0.07	-0.03	0.00	4.43	3.39	2.67	2.41	-0.02	0.05	-0.05	-0.01	3.92	3.21	2.71	2.46
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	42.90	43.50	50.50	52.20	57.90	60.00	69.70	72.10	32.80	34.50	42.90	43.90	47.40	50.90	61.00	64.20
30	54.00	58.10	63.20	70.00	69.90	74.20	82.40	87.50	44.40	46.90	54.40	59.00	59.60	66.40	75.90	83.50
50	68.90	78.10	82.40	89.70	85.00	91.50	95.40	98.20	56.40	69.60	75.50	84.20	78.00	87.40	92.60	97.10
100	80.60	87.00	93.80	96.60	93.10	95.60	99.00	99.30	76.20	82.80	92.40	94.40	91.60	95.60	99.30	99.70
	SHAC Mean group								SHAC Pooled							
20	34.30	40.40	53.70	66.40	52.20	61.80	75.60	85.50	45.80	52.70	65.30	78.30	65.70	73.40	84.90	91.30
30	40.80	49.00	63.60	79.20	60.90	70.20	86.30	94.00	52.50	59.50	74.00	87.00	73.20	80.30	90.70	96.40
50	57.70	73.70	84.50	94.70	81.00	90.70	96.70	99.60	71.60	81.80	90.70	97.60	87.90	95.40	98.60	99.70
100	76.60	84.80	95.90	97.50	92.30	96.40	99.80	100.00	81.40	88.50	96.50	98.20	95.10	97.20	99.70	100.00
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	35.40	42.70	55.20	67.70	53.40	63.20	76.40	85.90	48.30	54.40	66.50	79.40	66.00	72.80	85.80	91.70
30	46.30	53.40	67.60	81.70	65.00	73.50	87.50	94.60	59.20	65.00	77.70	88.50	76.50	83.00	91.60	96.60
50	61.60	76.20	85.90	95.30	83.10	91.80	97.70	99.70	74.70	84.60	91.80	98.00	89.30	95.50	98.70	99.70
100	78.90	87.00	96.50	98.10	92.70	96.50	99.90	100.00	84.10	90.30	97.00	98.30	95.50	97.40	99.80	100.00
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	40.90	43.20	51.10	62.40	54.30	57.90	65.40	71.40	35.90	38.60	45.90	60.90	48.40	53.80	61.50	69.50
30	38.80	43.60	53.00	62.30	54.90	63.70	73.60	82.00	35.20	39.40	47.50	59.00	49.50	56.70	67.60	76.00
50	37.60	46.30	52.30	65.30	62.70	72.60	80.00	87.40	34.20	41.90	49.40	60.60	59.30	67.80	75.30	84.10
100	41.80	47.30	53.90	68.10	79.40	84.80	92.70	94.40	36.40	41.50	49.70	63.80	70.90	79.70	86.80	91.30
	CCE Mean group								CCE Pooled							
20	7.80	6.90	7.50	8.40	12.60	12.90	14.80	17.70	8.50	7.80	7.60	8.60	13.50	13.80	14.20	17.30
30	5.70	5.80	6.10	7.80	12.10	13.50	19.00	22.30	5.70	6.10	6.10	7.80	12.50	14.30	20.00	23.10
50	4.80	6.90	5.80	5.40	11.80	21.10	27.70	36.60	4.70	6.00	5.60	5.80	14.30	24.60	27.90	35.30
100	5.20	5.50	4.60	5.20	21.10	34.00	45.00	57.00	5.10	5.60	4.60	5.20	25.40	34.90	43.20	57.50

Notes: see notes to Table A1.

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.

Table B2: Small sample properties of estimators for panels with spatially correlated errors ( $\delta = 0.4$ ) and unobserved common factors, under slope homogeneity.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	15.36	14.54	15.05	14.85	19.38	17.68	17.37	16.46	17.33	16.81	17.39	17.37	20.95	19.69	19.59	18.82
30	14.46	14.73	14.43	14.58	18.61	17.53	16.42	15.88	16.62	16.97	17.08	17.28	20.75	19.85	19.07	18.56
50	16.45	16.13	15.80	15.44	18.85	17.82	17.10	16.29	18.38	18.43	18.24	18.19	20.82	20.18	19.58	19.05
100	15.43	15.22	14.93	14.77	17.52	16.71	15.93	15.37	17.39	17.17	17.16	17.22	20.11	19.00	18.38	17.89
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	2.14	1.89	1.66	1.65	6.17	5.10	4.32	3.78	1.33	1.16	0.97	1.07	5.82	4.82	4.09	3.51
30	1.78	1.92	1.57	1.65	5.03	4.53	3.67	3.44	1.00	1.17	0.84	0.85	4.68	4.10	3.32	3.04
50	1.62	1.82	1.62	1.53	4.14	3.60	3.06	2.68	0.94	0.95	0.80	0.75	3.76	3.08	2.48	2.16
100	1.77	1.42	1.43	1.32	3.15	2.60	2.36	1.99	0.99	0.71	0.68	0.58	2.68	2.18	1.92	1.56
	CCE Mean group								CCE Pooled							
20	-0.19	0.05	0.06	0.03	8.29	5.90	4.39	2.87	-0.03	-0.01	0.09	0.07	6.93	5.31	4.13	2.94
30	0.20	0.07	-0.08	-0.07	7.11	5.06	3.48	2.35	0.00	-0.04	-0.08	-0.10	5.76	4.45	3.20	2.32
50	-0.15	0.03	-0.03	-0.01	5.57	3.92	2.68	1.72	-0.10	0.08	-0.04	0.00	4.37	3.46	2.44	1.65
100	0.05	0.18	-0.10	-0.04	3.77	2.79	1.77	1.18	0.03	0.14	-0.09	-0.04	3.08	2.44	1.65	1.13
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	51.20	57.60	68.00	73.70	69.60	75.30	84.20	91.30	36.90	40.00	50.90	59.70	51.70	62.30	73.40	81.90
30	60.10	68.50	78.40	86.30	76.80	84.10	91.70	96.80	44.90	55.60	64.60	75.20	63.30	75.70	84.90	92.40
50	80.00	87.50	93.10	97.50	94.00	98.00	98.90	99.90	65.80	76.80	85.10	92.30	84.50	92.00	97.00	98.90
100	87.40	91.50	97.90	99.70	95.40	98.80	99.90	100.00	80.10	87.40	95.60	99.20	93.70	97.90	99.80	100.00
	SHAC Mean group								SHAC Pooled							
20	32.10	44.00	62.10	78.40	54.90	68.20	84.80	95.50	50.20	59.00	77.80	91.40	71.10	81.90	93.40	97.90
30	41.20	54.90	69.10	86.10	62.80	78.20	90.60	98.10	55.70	68.80	81.40	94.20	76.30	88.00	94.90	99.10
50	65.00	79.40	89.70	97.70	87.10	95.50	98.50	100.00	79.10	90.30	94.90	99.80	95.00	98.80	99.40	100.00
100	80.00	87.70	96.90	99.70	94.40	97.90	99.90	100.00	86.60	91.60	98.30	99.90	95.60	98.70	100.00	100.00
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	35.20	44.80	62.40	79.30	56.60	69.20	85.00	95.40	54.60	62.10	78.80	91.20	71.60	82.10	93.50	98.10
30	46.50	58.90	72.80	87.70	67.30	80.20	91.50	98.20	61.50	72.60	85.30	95.10	80.70	89.10	95.20	99.30
50	68.20	82.00	90.80	97.60	87.80	95.50	98.70	100.00	82.50	91.40	95.70	99.70	95.50	99.00	99.30	100.00
100	82.00	89.70	97.40	99.80	94.80	98.20	99.90	100.00	88.30	93.10	98.70	99.90	96.00	99.00	100.00	100.00
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	24.90	25.40	30.50	41.50	52.00	60.20	74.40	86.40	19.90	21.20	26.60	32.50	45.10	51.60	64.70	80.80
30	23.90	27.30	31.20	45.20	57.80	71.10	81.50	91.40	19.10	21.30	24.70	37.60	48.70	61.60	74.20	86.00
50	25.00	30.60	36.30	44.00	70.90	85.00	92.90	98.10	20.20	22.20	25.70	34.30	63.60	78.40	88.70	96.40
100	30.70	34.50	41.00	51.10	92.40	96.90	99.20	100.00	22.50	22.20	27.80	35.80	86.80	94.10	97.70	99.90
	CCE Mean group								CCE Pooled							
20	5.10	6.20	7.00	6.10	11.40	17.00	25.50	45.40	6.10	7.00	7.10	6.80	12.60	18.40	29.20	46.50
30	6.20	5.80	5.90	7.40	12.90	19.70	32.70	58.60	6.20	6.30	6.60	7.30	15.20	22.10	34.80	60.30
50	6.70	6.10	6.00	4.90	15.10	25.90	46.40	82.40	4.80	6.00	5.70	5.30	19.50	35.40	52.90	85.30
100	4.60	6.30	4.40	4.50	24.80	48.50	77.00	98.70	4.60	6.10	4.80	4.50	35.00	57.50	83.00	99.40

Notes: see notes to Table A1.

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.

Table C1: Small sample properties of estimators for panels with spatially correlated errors ( $\delta = 0.4$ ) no unobserved common factors, under slope heterogeneity and serial correlation.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	0.00	0.12	0.15	-0.05	9.00	7.40	6.32	5.35	0.25	0.16	-0.01	0.01	8.24	7.61	6.69	5.72
30	0.57	0.06	0.07	0.14	7.41	6.21	5.16	4.33	0.39	-0.01	-0.10	0.18	6.96	6.14	5.37	4.56
50	-0.22	-0.08	-0.03	-0.07	5.81	4.70	3.94	3.40	-0.31	-0.01	-0.09	-0.05	5.40	4.67	4.20	3.67
100	-0.01	-0.08	0.16	-0.20	3.91	3.30	2.85	2.42	-0.05	-0.08	0.09	-0.15	3.66	3.31	3.07	2.66
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	0.20	-0.08	0.02	-0.06	8.04	7.74	6.90	6.01	0.24	-0.10	0.04	-0.04	8.24	7.90	6.90	6.00
30	0.35	0.11	0.09	0.17	6.65	6.29	5.45	4.94	0.27	0.16	0.06	0.16	6.79	6.35	5.50	4.96
50	-0.10	-0.22	-0.19	0.04	5.23	4.85	4.30	3.80	-0.07	-0.20	-0.18	0.03	5.35	4.85	4.38	3.84
100	-0.01	0.02	0.10	-0.18	3.66	3.31	3.00	2.70	-0.06	0.00	0.09	-0.18	3.71	3.39	3.02	2.72
	CCE Mean group								CCE Pooled							
20	-0.24	0.01	0.11	-0.11	11.02	8.47	6.77	5.64	0.18	0.18	0.04	-0.06	9.30	8.13	6.78	5.68
30	0.51	0.05	0.14	0.10	9.15	6.88	5.77	4.60	0.40	0.08	0.11	0.12	8.01	6.42	5.55	4.69
50	-0.26	-0.18	0.08	-0.13	7.01	5.24	4.23	3.57	-0.22	-0.09	0.04	-0.09	6.09	4.99	4.24	3.60
100	0.12	-0.02	0.15	-0.22	4.94	3.85	3.08	2.53	0.13	-0.04	0.13	-0.20	4.29	3.62	3.10	2.53
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	7.00	8.60	8.10	7.50	10.70	15.20	14.70	15.70	6.70	6.80	7.20	6.40	10.80	12.80	15.80	17.70
30	8.20	7.70	6.70	6.70	14.90	17.70	19.30	22.10	7.50	7.30	6.50	6.60	14.40	16.60	18.90	23.70
50	5.20	5.50	5.70	5.80	16.00	20.40	24.20	27.40	5.60	5.10	5.60	5.50	15.30	19.50	23.90	31.30
100	5.40	6.30	5.70	5.40	29.30	34.70	42.40	44.90	6.00	4.80	5.70	5.60	25.40	33.10	45.90	52.70
	SHAC Mean group								SHAC Pooled							
20	11.00	11.90	15.40	22.00	16.50	20.90	28.30	38.50	12.70	16.80	19.20	27.00	19.20	26.30	31.90	42.30
30	11.60	14.00	14.10	21.20	20.60	25.70	32.20	47.30	12.40	15.30	17.80	24.80	24.30	29.10	37.40	49.80
50	10.30	9.90	13.20	18.90	21.40	29.50	40.10	56.90	11.60	14.70	19.90	23.50	27.90	32.90	44.00	57.30
100	9.20	10.30	16.00	20.70	35.30	44.90	63.30	77.50	11.50	14.70	20.90	25.60	41.10	51.90	64.30	77.10
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	11.10	10.90	15.30	22.50	16.00	20.60	28.00	37.90	13.30	16.60	19.00	26.10	18.80	26.10	30.90	42.00
30	11.00	13.40	14.00	20.20	18.80	25.10	32.70	47.80	13.70	15.00	17.50	24.40	25.20	29.50	37.70	49.30
50	9.30	9.90	13.20	18.90	20.60	29.20	39.50	56.50	11.80	13.50	18.80	23.20	27.70	33.40	43.50	56.20
100	8.80	10.10	15.30	20.50	33.60	44.30	62.80	77.10	11.50	15.10	21.50	25.60	40.80	51.10	64.60	76.90
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	38.40	45.20	48.80	57.30	47.00	53.70	59.70	70.20	36.00	44.00	45.10	54.70	43.10	51.10	58.80	66.00
30	37.50	43.70	46.00	57.40	51.10	55.00	64.10	73.80	34.10	43.00	44.60	52.70	48.20	53.10	62.40	71.80
50	35.70	43.10	46.10	55.30	54.70	60.10	69.90	79.80	33.90	40.00	46.00	55.30	53.30	59.60	68.40	77.70
100	38.80	41.60	48.20	58.80	69.20	77.20	84.10	90.20	36.20	39.10	47.80	56.00	65.90	75.60	82.40	89.40
	CCE Mean group								CCE Pooled							
20	7.70	6.30	7.30	7.70	11.30	11.70	15.60	18.30	6.70	6.70	8.10	8.30	10.80	12.30	15.20	17.80
30	6.30	6.20	6.30	6.70	12.50	13.40	17.40	23.00	7.30	6.10	6.30	6.20	12.30	14.40	18.10	23.10
50	6.00	5.60	6.00	5.60	13.10	14.90	21.80	31.60	6.90	4.50	6.20	6.10	13.20	17.20	21.80	30.70
100	6.50	4.90	4.50	5.40	20.50	27.00	39.00	49.00	6.00	5.30	6.20	4.90	23.80	28.00	40.60	48.50

Notes: see notes to Table A1.

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.

Table C2: Small sample properties of estimators for panels with spatially correlated errors ( $\delta = 0.4$ ) no unobserved common factors, under slope homogeneity and serial correlation.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	0.25	-0.15	0.11	0.01	7.83	6.16	4.40	3.10	0.16	-0.11	0.01	0.03	6.23	5.15	3.78	2.72
30	0.05	0.07	0.17	0.12	6.12	4.54	3.66	2.50	0.08	0.03	0.09	0.14	5.09	4.11	3.27	2.36
50	-0.11	-0.01	-0.09	0.06	5.20	3.94	2.82	1.90	-0.09	-0.06	-0.09	0.05	4.15	3.32	2.46	1.75
100	-0.05	-0.08	-0.01	0.03	3.46	2.60	2.01	1.53	-0.02	-0.05	-0.02	0.04	2.78	2.13	1.70	1.10
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	0.09	-0.03	0.00	-0.04	4.65	3.93	3.03	2.22	0.17	-0.08	-0.02	-0.02	5.05	4.22	3.23	2.31
30	0.14	0.06	0.10	0.02	4.02	3.23	2.50	1.85	0.14	-0.02	0.12	0.02	4.20	3.34	2.67	1.92
50	-0.04	0.00	-0.07	0.04	3.19	2.62	1.99	1.51	-0.03	0.03	-0.05	0.03	3.29	2.78	2.08	1.55
100	-0.01	0.05	0.02	0.03	2.07	1.70	1.32	0.94	0.01	0.06	0.01	0.02	2.18	1.76	1.40	0.97
	CCE Mean group								CCE Pooled							
20	0.66	-0.38	-0.02	0.01	9.55	7.37	5.07	3.46	0.40	-0.20	-0.08	0.03	8.22	6.43	4.72	3.30
30	0.13	-0.12	-0.02	0.03	8.07	5.82	4.15	2.81	0.39	-0.08	-0.03	0.07	6.72	5.14	3.80	2.72
50	-0.15	-0.02	-0.01	0.05	6.39	4.64	3.32	2.18	-0.24	-0.13	-0.05	0.06	5.27	4.05	3.01	2.07
100	0.04	-0.04	0.05	0.00	4.41	3.30	2.38	1.84	-0.02	-0.08	0.02	-0.01	3.66	2.85	2.15	1.78
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	7.30	8.90	7.20	6.70	16.80	19.50	29.50	51.40	6.90	8.10	6.00	7.00	14.70	17.10	26.50	43.10
30	6.70	7.30	7.30	9.20	23.40	28.10	43.00	67.10	6.10	6.30	7.60	7.90	17.90	20.80	35.40	58.20
50	6.10	6.80	5.60	5.90	26.50	36.60	53.40	83.00	7.80	7.40	5.30	4.70	18.90	29.10	42.30	75.00
100	8.30	5.90	6.60	5.60	51.30	65.00	85.80	96.60	6.90	5.80	6.40	5.60	34.00	46.70	73.30	84.20
	SHAC Mean group								SHAC Pooled							
20	8.30	8.50	6.20	7.40	15.00	18.60	28.10	43.10	8.30	8.30	7.10	6.50	18.80	22.70	32.40	51.50
30	7.00	7.60	7.60	7.40	19.40	23.40	36.90	59.90	6.70	7.00	7.10	7.40	24.70	28.80	41.20	65.20
50	9.50	8.50	6.00	5.30	23.10	32.20	45.30	77.20	8.50	7.80	6.90	6.00	30.60	39.90	56.50	83.60
100	8.90	6.00	8.00	7.70	38.20	51.70	75.60	95.30	9.30	7.50	6.50	7.60	52.90	67.60	86.80	99.60
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	7.80	8.20	6.70	6.60	15.60	18.60	26.90	42.70	8.30	8.80	6.90	6.40	19.30	21.50	31.70	51.30
30	8.20	5.60	8.40	8.70	19.70	22.40	37.40	59.90	8.30	7.30	7.30	8.10	25.60	29.30	43.10	67.30
50	9.70	8.40	6.60	5.40	22.00	32.10	45.40	76.80	8.20	8.00	6.90	6.00	31.10	38.80	57.20	84.10
100	8.20	6.80	7.60	7.60	37.80	51.10	75.60	95.00	9.30	7.40	7.30	7.20	54.80	68.10	86.90	100.0
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	17.10	18.00	17.20	16.40	40.60	46.80	60.20	79.80	16.60	17.20	16.10	13.90	37.40	43.40	56.20	75.60
30	19.00	17.70	14.70	16.80	49.90	56.50	73.20	90.90	17.70	15.40	14.50	13.70	45.30	52.00	68.70	88.00
50	16.40	17.80	16.70	17.50	60.50	71.70	85.30	98.20	16.50	17.10	14.20	17.60	57.20	69.00	83.30	96.50
100	17.80	15.80	15.40	5.60	85.40	93.80	99.10	100.0	18.00	15.30	14.60	5.20	83.00	92.30	98.50	100.0
	CCE Mean group								CCE Pooled							
20	6.70	7.50	5.40	7.10	11.20	13.40	20.80	35.30	6.70	7.00	6.70	6.30	11.80	14.70	21.20	37.30
30	6.00	6.40	5.20	5.90	13.30	15.20	26.80	47.80	6.30	6.00	5.90	6.80	16.10	18.10	29.10	50.70
50	6.20	5.40	6.50	5.60	13.70	21.30	34.10	65.00	6.60	7.00	6.60	5.80	17.20	25.30	39.50	67.90
100	4.60	6.10	5.70	5.00	22.20	33.40	58.30	83.50	5.00	6.40	6.20	5.0	30.60	41.40	66.10	88.30

Notes: see notes to Table A1.

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.

Table C3: Small sample properties of estimators for panels with spatially correlated errors ( $\delta = 0.8$ ) no unobserved common factors, under slope heterogeneity.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	-0.29	0.00	-0.22	-0.46	12.72	9.64	7.61	6.18	-0.15	-0.30	-0.03	-0.42	11.33	9.15	7.48	6.59
30	-0.12	-0.38	0.21	-0.06	10.57	8.47	6.51	5.02	-0.11	-0.25	0.19	-0.13	10.18	8.67	6.94	5.44
50	0.10	-0.01	0.09	0.16	7.15	5.75	4.50	3.60	-0.13	0.04	0.06	0.25	6.49	5.50	4.76	3.91
100	0.01	0.11	0.05	0.05	4.93	4.07	3.19	2.68	-0.08	0.03	-0.02	-0.01	4.94	4.27	3.46	3.04
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	0.09	0.06	0.10	-0.24	7.57	6.51	5.95	5.75	0.07	0.20	0.02	-0.33	8.13	6.97	6.10	5.72
30	-0.21	-0.08	0.04	-0.12	6.02	5.48	4.86	4.49	-0.04	-0.12	-0.02	-0.06	6.29	5.59	4.83	4.48
50	-0.24	0.33	0.14	0.19	4.52	4.27	3.77	3.40	-0.25	0.29	0.13	0.23	4.89	4.40	3.99	3.51
100	0.01	0.01	-0.03	-0.04	3.35	2.93	2.71	2.69	-0.01	-0.01	-0.03	-0.06	3.51	2.96	2.83	2.70
	CCE Mean group								CCE Pooled							
20	0.10	0.23	-0.11	-0.30	12.97	9.36	7.10	5.79	0.23	0.28	-0.26	-0.24	10.90	8.75	6.94	5.81
30	-0.32	-0.19	0.16	-0.08	10.83	7.74	6.00	4.72	-0.28	-0.18	0.16	0.00	9.12	7.31	5.84	4.75
50	-0.20	0.32	0.14	0.22	8.66	6.54	4.72	3.73	-0.31	0.30	0.06	0.23	7.09	5.85	4.59	3.72
100	-0.01	0.01	0.04	0.05	6.02	4.53	3.27	2.70	-0.05	-0.05	0.01	0.04	5.02	4.13	3.20	2.71
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	10.90	10.90	8.50	10.00	13.80	14.30	16.90	15.70	10.80	8.80	8.50	9.10	11.10	13.10	14.70	15.50
30	16.00	16.40	12.70	9.30	20.90	22.50	22.70	21.40	12.40	12.70	10.10	8.20	16.60	18.90	21.70	22.30
50	7.40	5.80	7.70	6.90	15.70	19.50	21.80	28.80	5.20	6.20	6.00	6.10	12.00	16.10	22.20	30.80
100	10.80	10.60	9.30	8.70	28.50	33.10	38.80	46.30	7.10	8.60	6.70	7.00	21.10	30.10	40.20	52.70
	SHAC Mean group								SHAC Pooled							
20	10.90	11.10	12.90	17.80	12.10	16.30	20.30	29.80	13.20	15.70	16.20	24.70	17.20	19.80	25.20	35.20
30	13.00	13.50	15.10	18.30	16.30	20.80	28.20	37.00	16.00	18.10	18.50	21.80	19.80	25.10	32.10	40.30
50	7.30	9.80	11.20	15.80	15.40	21.20	33.90	53.70	9.90	12.00	15.90	22.00	20.50	26.80	36.80	55.70
100	8.00	11.50	12.10	18.90	23.90	35.70	51.90	71.20	12.60	14.50	15.60	25.60	30.10	39.90	52.30	68.90
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	11.30	11.50	13.40	18.40	12.30	15.60	20.50	29.40	13.20	15.30	17.10	24.10	16.80	19.80	25.90	35.20
30	14.00	15.20	15.00	19.10	17.50	22.00	28.60	37.60	18.60	21.40	21.10	24.40	22.90	28.20	34.80	43.40
50	7.40	9.00	11.20	15.10	15.80	21.00	34.80	52.80	11.60	11.50	16.90	21.40	20.50	28.20	38.50	55.30
100	9.30	11.40	12.80	19.40	25.50	37.00	53.20	72.70	15.40	16.80	17.10	27.80	33.70	42.60	56.00	71.40
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	29.00	31.50	42.60	53.80	41.40	45.30	54.40	68.20	21.90	24.70	30.50	46.50	31.10	36.10	45.70	60.00
30	27.10	32.00	38.30	51.20	38.60	50.00	60.80	69.80	18.90	24.00	28.80	42.60	31.60	39.00	52.00	63.50
50	22.70	33.50	37.50	49.40	46.20	60.10	70.00	81.60	21.30	27.80	33.40	45.40	37.90	54.50	65.00	78.80
100	27.60	33.20	40.30	57.00	66.10	74.40	85.50	91.20	22.60	24.30	32.90	48.90	57.60	69.00	79.80	88.10
	CCE Mean group								CCE Pooled							
20	7.50	6.60	7.00	7.60	9.80	11.60	13.40	17.30	8.00	7.50	7.30	7.90	9.30	13.90	13.50	18.10
30	6.40	6.50	5.80	6.50	8.70	12.10	16.20	19.00	6.40	5.70	5.80	6.90	10.20	13.30	15.70	19.80
50	4.90	6.60	6.40	5.80	9.80	16.90	20.60	31.80	5.40	6.10	6.10	5.90	9.90	18.30	20.70	31.90
100	5.30	6.90	4.60	6.00	14.40	21.70	34.60	48.60	5.10	6.60	5.40	5.30	16.00	23.70	35.50	50.30

Notes: see notes to Table A1.

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.

Table C4: Small sample properties of estimators for panels with spatially correlated errors ( $\delta_t \sim IIDU(0, 0.8)$ ) no unobserved common factors, under slope heterogeneity.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	0.20	0.19	-0.13	-0.43	9.31	6.92	6.02	5.27	0.23	-0.14	0.09	-0.37	8.51	6.92	6.21	5.76
30	-0.21	-0.25	0.05	-0.08	7.37	5.99	4.75	4.13	-0.08	-0.19	0.07	-0.12	6.86	5.96	5.01	4.50
50	-0.05	0.07	0.17	0.14	5.45	4.50	3.62	3.14	-0.19	0.07	0.11	0.24	5.16	4.48	3.95	3.51
100	0.08	0.07	0.03	0.03	3.88	3.11	2.58	2.39	-0.07	-0.03	-0.04	0.01	3.68	3.20	2.86	2.70
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	0.31	0.03	0.13	-0.25	7.66	6.62	6.10	5.71	0.35	0.05	0.10	-0.28	8.02	6.77	6.17	5.70
30	-0.23	-0.10	0.06	-0.17	6.01	5.55	4.78	4.46	-0.14	-0.13	0.01	-0.10	6.07	5.57	4.78	4.45
50	-0.21	0.22	0.12	0.14	4.68	4.33	3.74	3.41	-0.17	0.18	0.13	0.16	4.90	4.37	3.85	3.47
100	-0.02	-0.03	-0.06	0.01	3.30	2.97	2.72	2.67	-0.05	-0.04	-0.08	0.00	3.37	2.97	2.76	2.68
	CCE Mean group								CCE Pooled							
20	0.56	0.19	-0.12	-0.42	11.05	7.92	6.26	5.40	0.47	0.12	-0.17	-0.30	9.70	7.62	6.23	5.43
30	-0.38	-0.21	0.03	-0.04	8.99	6.52	5.05	4.30	-0.41	-0.19	0.08	0.04	7.56	6.18	5.05	4.37
50	-0.20	0.21	0.20	0.18	6.90	5.19	3.92	3.27	-0.23	0.25	0.16	0.22	5.81	4.85	3.89	3.33
100	0.09	0.06	0.03	0.02	4.84	3.59	2.77	2.46	0.01	0.00	-0.02	0.03	4.08	3.35	2.77	2.49
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	9.00	6.70	8.10	7.80	14.50	14.40	16.60	15.10	8.60	6.50	8.10	7.10	12.80	14.60	16.50	15.50
30	8.50	8.10	6.80	7.30	16.60	20.00	20.40	22.00	8.10	8.30	5.80	6.60	14.50	17.40	21.30	22.80
50	6.00	6.00	5.90	5.50	17.30	21.90	26.50	30.60	5.40	5.70	4.70	4.30	15.00	22.50	30.60	36.20
100	6.90	5.60	6.20	5.40	30.10	38.30	43.00	51.50	6.20	6.60	6.00	5.80	27.80	40.50	48.80	59.40
	SHAC Mean group								SHAC Pooled							
20	10.60	11.00	17.90	26.50	15.90	21.40	30.10	42.20	13.70	15.90	20.80	31.40	20.40	23.90	33.00	44.60
30	9.90	13.40	14.60	22.30	17.20	24.40	38.10	51.90	13.60	15.70	19.80	26.60	21.60	30.20	40.40	50.50
50	7.70	12.00	13.00	21.40	19.70	32.00	49.30	64.80	11.70	15.00	18.20	27.40	26.40	37.90	51.00	66.80
100	9.00	11.50	14.70	25.20	33.90	51.10	70.70	83.80	10.60	14.70	18.80	30.80	40.60	54.80	68.20	79.40
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	11.50	10.90	17.80	25.60	16.60	22.10	30.10	41.50	14.60	14.50	20.40	30.30	21.00	25.00	34.30	44.00
30	10.50	14.40	14.20	22.00	17.80	24.30	37.20	51.10	14.00	18.70	20.30	26.90	22.30	29.70	40.60	51.10
50	8.20	11.00	12.10	21.30	20.00	32.10	48.80	64.60	11.70	15.00	17.80	25.90	25.20	37.70	49.90	65.90
100	9.30	11.50	14.60	24.70	34.40	50.70	70.60	83.40	12.60	15.60	19.30	31.10	42.70	55.80	68.90	79.30
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	28.20	29.60	39.10	52.50	38.90	41.60	55.10	66.20	25.00	27.10	34.50	50.20	34.10	39.10	50.60	63.60
30	23.90	30.10	34.50	49.30	36.30	46.00	59.10	69.60	19.60	27.80	31.60	44.80	33.30	42.20	55.10	67.40
50	24.10	29.80	34.70	46.60	44.40	56.80	69.80	80.20	24.40	28.80	35.50	46.90	43.70	56.30	67.60	80.20
100	26.60	31.30	36.20	51.30	64.00	73.00	83.90	90.30	23.60	28.20	36.20	49.70	59.90	71.60	81.80	89.50
	CCE Mean group								CCE Pooled							
20	8.10	6.90	7.20	8.80	11.30	13.70	16.20	17.90	8.40	7.90	7.80	8.40	12.50	14.60	14.90	19.00
30	6.90	7.00	6.80	6.80	10.20	13.00	18.10	22.90	5.10	6.90	6.20	6.20	10.00	14.70	18.00	22.40
50	5.90	6.20	5.10	5.10	11.90	20.40	29.30	35.80	5.00	5.90	5.80	6.00	12.20	21.80	28.40	35.30
100	4.90	6.00	4.70	6.70	18.60	30.50	44.70	56.50	4.80	6.40	4.80	6.40	20.70	32.90	43.40	56.40

Notes: see notes to Table A1.

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.

Table C5: Small sample properties of estimators for panels with cross section dependence switching from spatial processes ( $\delta = 0.4$ ) to unobserved common factors, under slope heterogeneity.

N\T	Bias ( $\times 100$ )				RMSE ( $\times 100$ )				Bias ( $\times 100$ )				RMSE ( $\times 100$ )			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Mean group								Pooled							
20	8.42	7.94	8.07	7.17	13.88	11.91	11.21	9.87	8.60	7.85	8.29	7.45	14.08	12.24	11.78	10.40
30	8.02	7.55	7.95	7.33	12.87	11.15	10.81	9.28	8.36	7.96	8.51	7.81	13.67	11.96	11.85	10.12
50	8.75	8.38	8.83	8.49	11.99	10.84	10.59	9.74	9.04	8.63	9.16	8.92	12.70	11.42	11.28	10.47
100	8.49	7.85	8.20	7.70	11.01	9.80	9.55	8.61	8.42	7.76	8.25	7.80	11.73	10.32	10.04	9.00
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	2.12	1.92	1.88	1.31	8.56	7.29	6.50	6.10	1.28	1.29	1.09	0.61	8.19	7.11	6.24	5.87
30	1.89	1.88	2.01	1.79	6.82	6.21	5.67	5.15	1.12	0.94	1.10	0.97	6.33	5.77	5.23	4.81
50	1.70	2.05	1.85	2.02	5.42	5.03	4.46	4.21	0.80	1.18	0.91	1.11	5.02	4.61	4.10	3.78
100	1.64	1.58	1.67	1.63	4.10	3.67	3.41	3.27	0.78	0.70	0.77	0.75	3.68	3.18	2.98	2.87
	CCE Mean group								CCE Pooled							
20	0.20	0.20	-0.15	-0.37	9.84	7.46	6.01	5.34	0.31	0.14	-0.30	-0.35	8.94	7.34	6.00	5.43
30	-0.22	-0.13	0.06	-0.10	7.70	5.98	4.88	4.21	-0.20	-0.20	0.03	-0.05	6.86	5.76	4.97	4.28
50	-0.22	0.14	0.11	0.22	5.95	4.81	3.80	3.19	-0.21	0.22	0.07	0.22	5.31	4.69	3.79	3.25
100	-0.01	-0.01	-0.04	0.02	4.22	3.28	2.64	2.37	-0.05	-0.03	-0.06	0.01	3.93	3.15	2.69	2.42
	Size ( $\times 100$ )				Power ( $\times 100$ )				Size ( $\times 100$ )				Power ( $\times 100$ )			
	Robust Mean Group								Robust Pooled							
20	23.90	23.10	26.40	26.10	40.80	41.40	48.60	50.80	20.40	20.30	25.00	24.90	36.40	41.00	48.30	52.20
30	32.10	34.00	36.90	36.00	50.00	54.20	63.10	68.00	27.00	28.90	33.80	32.50	45.40	51.50	60.80	68.90
50	39.70	43.00	49.90	54.70	63.40	71.60	79.70	88.00	33.30	40.80	51.60	55.70	60.90	70.60	83.20	91.10
100	53.40	55.60	65.00	69.90	78.10	84.00	90.30	94.30	53.80	56.60	69.70	75.00	79.10	87.20	94.00	98.00
	SHAC Mean group								SHAC Pooled							
20	19.50	21.10	29.50	39.60	37.50	46.10	57.70	71.10	25.20	25.70	33.20	41.60	43.50	48.30	61.70	70.30
30	23.00	27.50	36.20	46.60	43.80	54.40	70.50	83.60	27.80	33.60	41.90	50.50	51.40	58.90	71.40	82.50
50	30.40	40.10	53.90	69.50	62.40	75.30	87.90	96.30	36.50	45.00	55.80	68.90	68.70	77.50	87.80	95.30
100	51.30	55.80	72.30	81.10	80.80	90.00	96.30	99.00	53.60	57.00	70.20	80.40	81.00	88.20	94.90	97.90
	SHAC Mean group, $D$ mis-specified <sup>(2)</sup>								SHAC Pooled, $D$ mis-specified <sup>(2)</sup>							
20	20.20	23.10	31.90	40.90	39.00	48.00	59.60	70.90	26.40	27.30	35.20	43.10	43.80	49.40	62.60	71.00
30	27.50	31.50	41.70	51.10	47.70	58.60	72.10	84.80	32.20	38.80	47.30	54.20	55.10	60.60	74.10	84.10
50	34.20	44.50	59.60	72.00	64.60	77.90	89.40	96.60	43.10	50.90	61.30	72.40	69.10	78.50	89.10	95.70
100	55.50	60.60	76.90	83.60	81.90	91.20	96.60	99.40	56.80	61.50	74.00	82.50	81.90	89.80	95.20	98.30
	ML-SAR								ML-SAR, $S$ mis-specified <sup>(1)</sup>							
20	34.20	35.30	41.00	55.90	49.40	53.20	63.20	73.20	29.70	33.10	38.90	54.40	42.10	48.20	57.00	67.70
30	30.30	36.00	44.60	55.60	51.00	58.30	70.20	80.30	24.70	32.30	40.60	51.10	44.10	52.70	65.10	74.70
50	30.70	40.00	43.70	56.70	60.60	71.20	81.20	90.30	28.20	35.40	40.50	52.60	53.40	66.60	75.00	85.10
100	34.40	40.90	49.20	64.10	79.00	86.30	93.10	96.50	28.90	34.40	43.50	57.50	71.70	80.00	88.50	93.60
	CCE Mean group								CCE Pooled							
20	7.50	7.10	7.50	8.30	12.30	14.10	15.00	18.80	7.80	7.80	7.20	8.00	13.70	14.20	15.00	19.50
30	6.60	4.80	6.20	6.70	10.60	14.70	18.40	23.70	4.60	5.30	6.70	7.30	11.60	15.10	19.40	23.00
50	4.90	6.70	6.20	5.50	12.70	22.20	29.30	36.70	5.70	6.70	5.80	5.20	15.40	25.70	28.60	36.10
100	5.30	4.80	5.10	5.10	20.80	33.00	45.80	58.20	5.50	6.40	5.10	5.60	25.90	35.50	45.10	57.90

Notes: see notes to Table A1.

<sup>(1)</sup> - This estimator is computed under the incorrect assumption that cross section units are neighbours if their Euclidean distance is less than or equal to 2 while in the true  $S$  matrix cross section units are neighbours if their Euclidean distance is equal to 1. <sup>(2)</sup> - This estimator is computed under the incorrect assumption that cross section units are ordered on a line, rather than on a rectangular grid.