Estimation and inference in heterogeneous spatial panels with a multifactor error structure

Estimation and Inference in Heterogeneous Spatial Panels with a Multifactor Error Structure*

Jia Chen† Yongcheol Shin ‡ Chaowen Zheng §

This Version: May 7, 2021

Abstract

We develop a unifying econometric framework for the analysis of heterogeneous panel data models that can account for both spatial dependence and common factors. To tackle the challenging issues of endogeneity due to the spatial lagged term and the correlation between the regressors and factors, we propose the CCEX-IV estimation procedure that approximates factors by the cross-section averages of regressors and deals with the spatial endogeneity using the internal instrumental variables. We develop the individual and Mean Group estimators, and establish their consistency and asymptotic normality. By contrast, the Pooled estimator is shown to be inconsistent in the presence of parameter heterogeneity. Monte Carlo simulations confirm that the finite sample performance of the proposed estimators is quite satisfactory. We demonstrate the usefulness of our approach with an application to the house price growth for Local Authority Districts in the UK over 1997Q1-2016Q4.

JEL Classification: C13, C15, C23, R30.

Key Words: Spatial Dependence and Heterogeneity, Unobserved Common Factors, CCEX-IV Estimation, the UK House Price Growth, GCM Analysis.

---

*We are mostly grateful for the insightful comments by the editor, the associated editor and two anonymous referees as well as In Choi, David Kang, Young Hoon Lee, Rui Lin, Laura Serlenga, Stefano Soccorsi, Michael Thornton, Takashi Yamagata, and the seminar participants at Lancaster University Management School and University of York for helpful comments. Partial financial support from the ESRC (Grant Reference: ES/T01573X/1) is gratefully acknowledged. The usual disclaimer applies.

†Department of Economics and Related Studies, University of York, York, YO10 5DD, UK; E-mail address: jia.chen@york.ac.uk.

‡The corresponding author, Department of Economics and Related Studies, University of York, York, YO10 5DD, UK; E-mail address: yongcheol.shin@york.ac.uk.

§Department of Economics and Related Studies, University of York, York, YO10 5DD, UK; E-mail address: cz1113@york.ac.uk.

1 Introduction

The increasingly globalised nature of the world economy and pervasive evidence of cross-section dependence (CSD) found in empirical studies (e.g., Pesaran (2015); Mastromarco et al. (2016)) have driven considerable interests in developing panel data models that can explain observed patterns of comovement among economic and financial variables. As the interdependence among different economic agents could arise from strategic interaction or a common third factor, such as common technological shocks and regulatory changes, two strands of work have been proposed accordingly. The first uses a spatial-based approach by assuming that the behaviour of each spatial unit is related to the behaviour of its neighbours. The second uses a factor-based approach which allows each individual to be affected by unobserved factors with different intensities. While these two methods are mostly developed separately, researchers are now paying more efforts to combine them and develop a unified characterisation of CSD.

To date, some progress has been made in the literature. Pesaran and Tosetti (2011) allow the error components to be spatially dependent and contain unobserved factors, and propose the common correlated effects (CCE) estimator advanced by Pesaran (2006). Bailey et al. (2016a) develop a multi-step estimation procedure that can distinguish CSD that is purely spatial from the one that is due to common factors. Mastromarco et al. (2016) propose a technique for modelling stochastic frontier panels by combining the exogenously driven factor-based approach and an endogenous threshold regime selection mechanism. Bai and Li (2014) consider a homogeneous spatial panel data model with common shocks, and develop the quasi maximum likelihood (QML) estimation. Using a similar model structure, Yang (2021) develops consistent estimators that combine CCE and instrumental variable (IV)/generalised method of moments (GMM) estimation. See also Bai and Li (2021), Shi and Lee (2017) and Lu (2017).

The aforementioned studies have developed the joint analysis of the spatial and factor dependence under the assumption that the slope parameters are homogeneous. In a data-rich environment, slope homogeneity is a restrictive assumption, as the strength and direction of spatial dependence between entities may vary over space. In the spatial literature, only recently, Aquaro et al. (2021) and Shin and Thornton (2020) have explicitly allowed the slope parameters to be heterogeneous and develop the QML and the control function-based estimators, see also LeSage and Chih (2018).

Following this research trend, our primary objective is to develop a unifying econometric framework for the estimation of heterogeneous panel data models that can jointly accommodate the spatial and factor dependence. To tackle the challenging issues for developing consistent estimation in the presence of spatial heterogeneity and endogeneity caused by the spatial lagged term and the correlation between the regressors and unobserved factors, we propose to combine the CCE and IV methods. Notice, however, in the panel data literature with interactive effects that most studies impose some restrictions on the factor structure affecting the dependent and independent variables (Pesaran and Tosetti (2011), Bai and Li (2014), Bai and Li (2021), and Yang (2021)). Following the forecasting literature (Bai and Ng (2008)), we propose a more general factor structure by explicitly allowing the factors affecting the dependent and independent variables to be arbitrarily different, so far as the rank condition is satisfied (see Assumption 3 in Section 2). We may apply the standard CCE approach to approximate unobserved common factors by employing the cross-section averages (CSA) of dependent and independent variables, denoted $\bar{y}_t$ and $\bar{\boldsymbol{x}}_t$, and then deal with the spatial endogeneity via the IV method. This is referred to as the CCE-IV estimation. The conventional use of $\bar{y}_t$ as factor proxy leads to the small sample bias, due to its correlation with idiosyncratic errors unless N is large. This issue will be more complicated in the presence of spatial dependence. We conjecture that such a bias will be non-negligible, especially if the spatial weighting matrix remains relatively dense for all the sample sizes. In this regard, we propose the simple and robust approach of using $\bar{\boldsymbol{x}}_t$ only as factor proxies that can directly control the factors correlated with the regressors, which is referred to as the CCEX-IV estimation. The factors specific to the regression residuals may be left unaccounted for, but this is shown not to affect consistency of the CCEX-IV estimator. Hence, our approach is in line with the robust estimation, which is widely employed to avoid uncertainty in (consistently) estimating nuisance parameters for potential efficiency gain.

The CCEX-IV estimation has several advantages over existing methods developed for homogeneous spatial panel data models with interactive effects. The QML/PC methods developed by Bai and Li (2014, 2021) are computationally demanding, especially if N is large, and require the consistent estimation of the number of unobserved factors, which is a challenging task. Kuersteiner and Prucha (2020) propose a quasi-differencing transformation to eliminate the individual factor loadings and treat the factors as estimands in homogeneous dynamic spatial panel models. Their GMM method is developed for fixed T panels while our approach is for large T panels. Moreover, the GMM procedure is more complicated in the presence of the multiple factors, which also requires the number of factors to be estimated consistently. More importantly, an extension of these methods to the heterogeneous spatial panels with interactive effects is currently unavailable. Furthermore, our method can be relatively easily extended to the development of nonlinear/quantile heterogeneous panel data models with both spatial dependence and common factors, e.g., Boneva and Linton (2017).

We show that the parameters in the individual regressions can be estimated consistently by applying the de-factored IVs directly to the original regressions. These estimators are $\sqrt{T}$-consistent and follow asymptotic normal distributions. We also establish $\sqrt{N}$-consistency and asymptotic normality for the mean group (MG) estimator. We provide nonparametric variance estimators, robust to residual heteroskedasticity and serial correlation. Moreover, we find that the pooled estimator is no longer consistent in the presence of the heterogeneous parameters. This is mainly due to the remaining correlation between the spatial lagged term and the random spatial coefficients. This is a new finding, extending the seminal work by Pesaran and Smith (1995) who establish that the pooled estimator is inconsistent in the heterogeneous dynamic panels. Only in the special case with the homogeneous parameters, the pooled estimator becomes $\sqrt{NT}$-consistent and follows a normal distribution asymptotically.

Via Monte Carlo simulations, we investigate the finite sample performance of CCEX-IV and

CCE-IV estimators. The biases of the CCEX-IV individual and MG estimators are relatively small in almost all cases. On the other hand, the performance of the CCE-IV estimator depends crucially upon the degree of sparsity of the spatial weighting matrix. If the network is relatively sparse, then its performance is satisfactory. However, if the network is relatively dense, the biases of the CCE-IV spatial coefficients remain substantial at all the sample sizes. In this case RMSEs of the CCE-IV estimator are also significantly higher than the CCEX-IV counterpart even for large N. This is in line with our conjecture that the conventional use of $\bar{y}_t$ as factor proxies may suffer from the remaining endogeneity due to the correlation between $\bar{y}_t$ and idiosyncratic errors, provided that the network remains relatively dense even for large N. Furthermore, we establish that the pooled estimators exhibit substantial biases under the parameter heterogeneity.

We apply our approach to analyse not only the spatial patterns of quarterly real house price growth for Local Authority Districts in the UK over the period 1997Q1-2016Q4, but also the impacts of income and population growth. Our main findings are summarised as follows: (i) The individual spatial coefficients are quite heterogeneous, but overall positive. This is in line with *a priori* expectation that the house price increase in nearby area causes the local house demand to rise. (ii) The MG spatial coefficients at the regional level are all positive and heterogeneous, but tightly clustered around the national mean of 0.57. The impacts of population growth are larger for North East & York, North West and South West while the impacts of income growth are larger in the rest of regions. (iii) Following Greenwood-Nimmo et al. (2021) and Shin and Thornton (2020), we conduct the generalised connectedness measure (GCM) analysis at the regional level and identify London and East Midlands as the most influential transmitters of population and income growth shocks affecting the house price growths in the UK. However, our findings do not provide full support for the London-centric view of ripple effects where house price appreciation begins in South East and London before spreading to the rest of the country. We find that the ripple effects originated from London may have eventually spread to North East & York and South West, but not to North West and Wales.

The rest of the paper is organised as follows. Section 2 describes the model and assumptions. Section 3 develops the asymptotic theory. Section 4 presents the finite sample performance of the proposed estimator. Section 5 provides an empirical application. Section 6 concludes. The mathematical proofs and the additional simulation and empirical results are presented in the Online Supplement.

Notations. C represents a positive constant. For any $N \times N$ real matrix, $\boldsymbol{A} = (a_{ij})$, $\|\boldsymbol{A}\| = \sqrt{tr(\boldsymbol{A}\boldsymbol{A}')}$ denotes the Frobenius norm. The row and column sum norms of $\boldsymbol{A}$ are defined as $\|\boldsymbol{A}\|_\infty = \underset{1\leq i\leq N}{max} \sum_{j=1}^{N} |a_{ij}|$ and $\|\boldsymbol{A}\|_1 = \underset{1\leq j\leq N}{max} \sum_{i=1}^{N} |a_{ij}|$. We denote vec($\boldsymbol{A}$) as the vectorisation operator that stacks the columns of $\boldsymbol{A}$ into a column vector, $\lambda_{\max}(\boldsymbol{A})$ as the largest eigenvalue of $\boldsymbol{A}$, and tr($\boldsymbol{A}$) as the trace of $\boldsymbol{A}$. We use $\otimes$ as the Kronecker product operator, $a \sim b$ representing that a and b are equivalent in the order of magnitude, and $(N, T) \rightarrow \infty$ implying that N and T tend to infinity jointly.

2 The Model and Assumptions

Consider the heterogeneous spatial autoregressive panel data model with unobserved common factors:

$$y_{it} = \rho_i y_{it}^* + \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \boldsymbol{\gamma}_{1i}'\boldsymbol{f}_{1t} + \boldsymbol{\gamma}_{2i}'\boldsymbol{f}_{2t} + \varepsilon_{it} = \rho_i y_{it}^* + \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \boldsymbol{\gamma}_{yi}'\boldsymbol{f}_{yt} + \varepsilon_{it}, \tag{1}$$

where y_{it} is the dependent variable of the i-th spatial unit at time t, $y_{it}^* = \sum_{j=1}^{N} w_{ij} y_{jt}$ is the spatial lagged variable with w_{ij} the (i,j)-th entry of the $N \times N$ spatial weighting matrix, $\boldsymbol{W}$, ρ_i is the heterogeneous spatial autoregressive parameter, $\boldsymbol{x}_{it} = (x_{it,1}, \ldots, x_{it,k})'$ is a $k \times 1$ vector of independent variables and $\boldsymbol{\beta}_i$ is a $k \times 1$ vector of heterogeneous parameters, $\boldsymbol{f}_{1t}$ and $\boldsymbol{f}_{2t}$ are $r_1 \times 1$ and $r_2 \times 1$ vectors of unobserved factors with $\boldsymbol{\gamma}_{1i}$ and $\boldsymbol{\gamma}_{2i}$ being the factor loadings, and ε_{it} is the idiosyncratic disturbance. $\boldsymbol{f}_{yt} = (\boldsymbol{f}_{1t}', \boldsymbol{f}_{2t}')'$ and $\boldsymbol{\gamma}_{yi} = (\boldsymbol{\gamma}_{1i}', \boldsymbol{\gamma}_{2i}')'$ are $r_y \times 1$ vectors of factors and loadings with $r_y = r_1 + r_2$. Next, we consider the following data generating process (DGP) for $\boldsymbol{x}_{it}$:

$$\boldsymbol{x}_{it} = \boldsymbol{\Gamma}_{1i}'\boldsymbol{f}_{1t} + \boldsymbol{\Gamma}_{3i}'\boldsymbol{f}_{3t} + \boldsymbol{v}_{it} = \boldsymbol{\Gamma}_{xi}'\boldsymbol{f}_{xt} + \boldsymbol{v}_{it}, \tag{2}$$

where $\boldsymbol{f}_{3t}$ is an $r_3 \times 1$ vector of unobserved factors, $\boldsymbol{\Gamma}_{1i}$ and $\boldsymbol{\Gamma}_{3i}$ are $r_1 \times k$ and $r_3 \times k$ matrices of factor loadings, and $\boldsymbol{v}_{it} = (v_{it,1}, \ldots, v_{it,k})'$ is a $k \times 1$ vector of idiosyncratic disturbances. Moreover, $\boldsymbol{f}_{xt} = (\boldsymbol{f}_{1t}', \boldsymbol{f}_{3t}')'$ and $\boldsymbol{\Gamma}_{xi} = (\boldsymbol{\Gamma}_{1i}', \boldsymbol{\Gamma}_{3i}')'$ are $r_x \times 1$ vector and $r_x \times k$ matrix with $r_x = r_1 + r_3$.

The model given by (1) and (2) is general and practical. It can accommodate both weak and strong CSD through the spatial lagged term and common factors. Furthermore, it allows the dependent and independent variables to be influenced by different factors (e.g., Bai and Ng (2008)): y_{it} and $\boldsymbol{x}_{it}$ share the common factors, $\boldsymbol{f}_{1t}$, but they are subject to their specific factors, $\boldsymbol{f}_{2t}$ and $\boldsymbol{f}_{3t}$. Importantly, the slope heterogeneity renders the data speaking about the relative susceptibility of each spatial unit to external conditions. In this regard, our model encompasses several existing studies, e.g., Pesaran (2006), Bai (2009), Bai and Li (2014), Aquaro et al. (2021) and Yang (2021).

To develop consistent estimation of the $(k+1) \times 1$ vector of parameters, $\boldsymbol{\theta}_i = (\rho_i, \boldsymbol{\beta}_i')'$, we should address two sources of endogeneity: the regressors, $\boldsymbol{x}_{it}$ are correlated with factors while the spatial lagged term, y_{it}^* is correlated with both factors and idiosyncratic error, ε_{it}. Econometric methods have been developed separately for dealing with the spatial or factor dependence. The spatial endogeneity can be resolved by using QML (Lee (2004)) or IV/GMM estimation (Kelejian and Prucha (1998, 1999)). The common factors can be approximated by the PC estimates (Bai (2003, 2009)) or the CSA of the variables (Pesaran (2006)). Only recently, a number of studies have attempted to combine both approaches. Bai and Li (2014) consider a homogeneous spatial panel data model with common shocks and develop the QML estimator whilst Yang (2021) proposes to combine CCE and IV/GMM analysis. See also Bai and Li (2021), Bailey et al. (2016a), Mastromarco et al. (2016), Shi and Lee (2017) and Kuersteiner and Prucha (2020).

Following this trend, we propose to combine the CCE and IV estimation for consistently estimating $\boldsymbol{\theta}_i$. We first stack (1) over i:

$$\boldsymbol{y}_{.t} = \boldsymbol{\rho}\boldsymbol{W}\boldsymbol{y}_{.t} + \boldsymbol{\mathcal{B}}\boldsymbol{x}_{.t} + \boldsymbol{\gamma}_y\boldsymbol{f}_{yt} + \boldsymbol{\varepsilon}_{.t}, \tag{3}$$

where $\boldsymbol{y}_{.t} = (y_{1t}, \cdots, y_{Nt})'$, $\boldsymbol{x}_{it} = (x_{it,1}, \cdots, x_{it,k})'$, $\boldsymbol{x}_{.t} = (\boldsymbol{x}'_{1t}, \cdots, \boldsymbol{x}'_{Nt})'$, $\boldsymbol{\rho} = diag(\rho_1, \cdots, \rho_N)$, $\boldsymbol{\beta}_i = (\beta_{1i}, \cdots, \beta_{1k})'$, $\boldsymbol{\mathcal{B}} = diag(\boldsymbol{\beta}'_1, \cdots, \boldsymbol{\beta}'_N)$, $\boldsymbol{\gamma}_y = (\boldsymbol{\gamma}_{y1}, \cdots, \boldsymbol{\gamma}_{yN})'$, and $\boldsymbol{\varepsilon}_{.t} = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})'$.

We then stack (3) over t:

$$\boldsymbol{y} = (\boldsymbol{I}_T \otimes \boldsymbol{\rho W})\boldsymbol{y} + (\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})\boldsymbol{x} + (\boldsymbol{I}_T \otimes \boldsymbol{\gamma}_y)\boldsymbol{f}_y + \boldsymbol{\varepsilon}, \tag{4}$$

where $\boldsymbol{y} = (\boldsymbol{y}'_{.1}, \cdots, \boldsymbol{y}'_{.T})'$, $\boldsymbol{x} = (\boldsymbol{x}'_{.1}, \cdots, \boldsymbol{x}'_{.T})'$, $\boldsymbol{f}_y = (\boldsymbol{f}'_{y1}, \cdots, \boldsymbol{f}'_{yT})'$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_{.1}, \cdots, \boldsymbol{\varepsilon}'_{.T})'$. Assume for the moment that $\boldsymbol{f}_{yt}$, $t = 1, \ldots, T$, are observable, and define the idempotent matrices:

$$\tilde{\boldsymbol{M}}_{\boldsymbol{F}_y} = \boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N, \quad \boldsymbol{M}_{\boldsymbol{F}_y} = \boldsymbol{I}_T - \boldsymbol{F}_y(\boldsymbol{F}'_y\boldsymbol{F}_y)^{-1}\boldsymbol{F}'_y,$$

where $\boldsymbol{F}_y = (\boldsymbol{f}_{y1}, \ldots, \ \boldsymbol{f}_{yT})'$ is a $T \times r_y$ matrix. It is then easily seen that

$$\begin{aligned}
&\tilde{\boldsymbol{M}}_{\boldsymbol{F}_y}(\boldsymbol{I}_T \otimes \boldsymbol{\gamma}_y)\boldsymbol{f}_y = (\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)(\boldsymbol{I}_T \otimes \boldsymbol{\gamma}_y)\boldsymbol{f}_y = \text{vec}(\boldsymbol{\gamma}_y\boldsymbol{F}'_y\boldsymbol{M}'_{\boldsymbol{F}_y}) = \boldsymbol{0}, \\
&\tilde{\boldsymbol{M}}_{\boldsymbol{F}_y}(\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})\boldsymbol{x} = (\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)(\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})\boldsymbol{x} = (\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})(\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)\boldsymbol{x}, \\
&\tilde{\boldsymbol{M}}_{\boldsymbol{F}_y}(\boldsymbol{I}_T \otimes \boldsymbol{\rho W})\boldsymbol{y} = (\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)(\boldsymbol{I}_T \otimes \boldsymbol{\rho W})\boldsymbol{y} = (\boldsymbol{I}_T \otimes \boldsymbol{\rho W})(\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)\boldsymbol{y}.
\end{aligned} \tag{5}$$

The equivalence in (5) suggests that $\tilde{\boldsymbol{M}}_{\boldsymbol{F}_y}(\boldsymbol{I}_T \otimes \boldsymbol{\rho W})\boldsymbol{y}$ can be interpreted as the de-factored spatial lagged term of the dependent variable or the spatial lagged term of the de-factored dependent variable. Pre-multiplying (4) by $\tilde{\boldsymbol{M}}_{\boldsymbol{F}_y}$, we obtain:

$$\tilde{\boldsymbol{y}}_0 = (\boldsymbol{I}_T \otimes \boldsymbol{\rho W})\tilde{\boldsymbol{y}}_0 + (\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{\varepsilon}}_0, \tag{6}$$

where $\tilde{\boldsymbol{y}}_0 = (\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)\boldsymbol{y}$ is the de-factored $\boldsymbol{y}$, and $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{\varepsilon}}_0$ are defined similarly. The transformed model, (6) is a heterogeneous spatial panel data (HSPD) model for the de-factored data, to which we can apply the IV/GMM method. Next, to derive the IVs internally, we rewrite (6) as

$$(\boldsymbol{I}_T \otimes (\boldsymbol{I}_N - \boldsymbol{\rho W}))\tilde{\boldsymbol{y}}_0 = (\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)(\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})\boldsymbol{x} + \tilde{\boldsymbol{\varepsilon}}_0.$$

Assuming that $\boldsymbol{I}_N - \boldsymbol{\rho W}$ is invertible (see Assumption 4 below), we obtain:

$$\begin{aligned}
\tilde{\boldsymbol{y}}_0 &= (\boldsymbol{I}_T \otimes (\boldsymbol{I}_N - \boldsymbol{\rho W})^{-1})(\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)(\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})\boldsymbol{x} + (\boldsymbol{I}_T \otimes (\boldsymbol{I}_N - \boldsymbol{\rho W})^{-1})\tilde{\boldsymbol{\varepsilon}}_0 \\
&= (\boldsymbol{M}_{\boldsymbol{F}_y} \otimes \boldsymbol{I}_N)(\boldsymbol{I}_T \otimes (\boldsymbol{I}_N - \boldsymbol{\rho W})^{-1})(\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})\boldsymbol{x} + (\boldsymbol{I}_T \otimes (\boldsymbol{I}_N - \boldsymbol{\rho W})^{-1})\tilde{\boldsymbol{\varepsilon}}_0,
\end{aligned} \tag{7}$$

which implies that valid instruments can be constructed by (higher orders of) spatial lagged terms of the de-factored regressors or the de-factored (higher orders of) spatial lagged terms of regressors. This analysis is not feasible because factors are latent.

Yang (2021) studies a homogeneous spatial panel data model with interactive effects:

$$y_{it} = \rho y^*_{it} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\gamma}'_i\boldsymbol{f}_t + \varepsilon_{it}. \tag{8}$$

Assuming that y_{it} and $\boldsymbol{x}_{it}$ share the same common factors, she develops the CCE estimator with $(\bar{y}_t, \bar{\boldsymbol{x}}_t')'$ as factor proxies. In the presence of spatial dependence, however, the use of $\bar{y}_t$ may result in non-negligible biases unless N is large. To illustrate, consider the homogeneous panel data model with unobserved factors:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\gamma}_i'\boldsymbol{f}_t + \varepsilon_{it},$$

and the augmented model with $(\bar{y}_t, \bar{\boldsymbol{x}}_t')'$ as factor proxies:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\eta}_{1i}'\bar{\boldsymbol{x}}_t + \eta_{2i}\bar{y}_t + \hat{\varepsilon}_{it},$$

where $\bar{y}_t = \sum_{i=1}^N y_{it}/N$, $\bar{\boldsymbol{x}}_t = \sum_{i=1}^N \boldsymbol{x}_{it}/N$, and $\hat{\varepsilon}_{it} = \boldsymbol{\gamma}_i'\boldsymbol{f}_t - \boldsymbol{\eta}_{1i}'\bar{\boldsymbol{x}}_t - \eta_{2i}\bar{y}_t + \varepsilon_{it}$. Assuming that ε_{it} is independent of $\boldsymbol{f}_s$, $\boldsymbol{x}_{js}$ for all j and s, and cross-sectionally independent with uniformly bounded variance, $\sigma_{\varepsilon,i}^2 = var(\varepsilon_{it})$, we have:

$$\mathrm{E}(\bar{y}_t\varepsilon_{it}) = \frac{1}{N}\sum_{j=1}^N \mathrm{E}[(\boldsymbol{x}_{jt}'\boldsymbol{\beta} + \boldsymbol{\gamma}_j'\boldsymbol{f}_t + \varepsilon_{jt})\varepsilon_{it}] = \frac{1}{N}\sum_{j=1}^N \mathrm{E}(\varepsilon_{jt}\varepsilon_{it}) = \frac{\sigma_{\varepsilon,i}^2}{N} \sim \frac{1}{N}.$$

Hence, for small N, the CCE estimator may suffer from finite sample bias due to the "smearing effect" (Greene (2010)). This issue will be more complicated in the presence of spatial dependence. From (8), we obtain:

$$\begin{aligned}
\mathrm{E}(\bar{y}_t\varepsilon_{it}) &= \frac{1}{N}\boldsymbol{\iota}_N'\,\mathrm{E}[(\boldsymbol{I_N} - \rho\boldsymbol{W})^{-1}(\boldsymbol{X}_{.t}\boldsymbol{\beta} + \boldsymbol{\gamma}'\boldsymbol{f}_t + \boldsymbol{\varepsilon}_{.t})\varepsilon_{it}] = \frac{1}{N}\boldsymbol{\iota}_N'\,\mathrm{E}\left[(\boldsymbol{I_N} - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}_{.t}\varepsilon_{it}\right] \\
&= \frac{1}{N}\sum_{r=0}^{\infty}\sum_{j=1}^{N}\rho^r w_{ji}^r\,\mathrm{E}(\varepsilon_{it})^2 = \sum_{r=0}^{\infty}\sum_{j=1}^{N}\rho^r w_{ji}^r\frac{\sigma_{\varepsilon,i}^2}{N} \sim \frac{C(\rho,\boldsymbol{W})}{N},
\end{aligned} \tag{9}$$

where $\boldsymbol{\iota}_N$ is an $N \times 1$ vector of ones, $\boldsymbol{X}_{.t} = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{Nt})'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_N)$, $\boldsymbol{\varepsilon}_{.t} = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, w_{ji}^r is the (j,i)-th element of $\boldsymbol{W}^r$. $C(\rho,\boldsymbol{W})$ is expected to increase with $|\rho|$ and becomes larger as the spatial weighting matrix, $\boldsymbol{W}$ is denser.

In this regard, we propose a simple and robust approach to use $\bar{\boldsymbol{x}}_t$ only as proxies for unobserved factors. This is referred to as the CCEX estimator.[1] The main motivation behind the CCE estimation lies in dealing with endogeneity caused by the correlation between regressors and unobserved factors (see Section 3.3 in Bai (2009)). If so, it will be sufficient to control the factors $\boldsymbol{f}_{xt}$ in (2) that are correlated with $\boldsymbol{f}_{yt}$ in (1), which can be approximated by $\bar{\boldsymbol{x}}_t$. Following the similar logic, Norkuté et al. (2021) propose the IV estimator for dynamic panel data models with exogenous covariates and a multifactor error structure by using the de-factored covariates as instruments. But, they employ the PC approach and do not consider the spatial effects.

[1]Only two studies have used this approach: Boneva and Linton (2017) apply the corresponding CCE estimator in a binary choice model. Chen and Yan (2019) propose a three-step CCE estimator for a homogeneous panel data model with common factors.

We stack (2) over i and vectorise it:

$$\mathrm{vec}(\boldsymbol{X}'_{.t}) = \boldsymbol{\Gamma}'_1 \boldsymbol{f}_{1t} + \boldsymbol{\Gamma}'_3 \boldsymbol{f}_{3t} + \mathrm{vec}(\boldsymbol{V}'_{.t}) = \boldsymbol{\Gamma}'_x \boldsymbol{f}_{xt} + \mathrm{vec}(\boldsymbol{V}'_{.t}), \tag{10}$$

where $\boldsymbol{X}_{.t} = (\boldsymbol{x}_{1t}, \ldots, \boldsymbol{x}_{Nt})'$, $\boldsymbol{\Gamma}_1 = (\boldsymbol{\Gamma}_{11}, \ldots, \boldsymbol{\Gamma}_{1N})$, $\boldsymbol{\Gamma}_3 = (\boldsymbol{\Gamma}_{31}, \ldots, \boldsymbol{\Gamma}_{3N})$, $\boldsymbol{\Gamma}_x = (\boldsymbol{\Gamma}'_1, \boldsymbol{\Gamma}'_3)' = (\boldsymbol{\Gamma}_{x1}, \ldots, \boldsymbol{\Gamma}_{xN})$, and $\boldsymbol{V}_{.t} = (\boldsymbol{v}_{1t}, \ldots, \boldsymbol{v}_{Nt})'$. Next, pre-multiply (10) by $\boldsymbol{\Upsilon} = \boldsymbol{\iota}'_N \otimes \boldsymbol{I}_k / N$:

$$\bar{\boldsymbol{x}}_t = \bar{\boldsymbol{\Gamma}}'_x \boldsymbol{f}_{xt} + \bar{\boldsymbol{v}}_t, \tag{11}$$

where $\bar{\boldsymbol{x}}_t = \boldsymbol{\Upsilon}\,\mathrm{vec}(\boldsymbol{X}'_{.t}) = \sum_{i=1}^N \boldsymbol{x}_{it}/N$, $\bar{\boldsymbol{\Gamma}}_x = (\boldsymbol{\Upsilon}\boldsymbol{\Gamma}'_x)' = \sum_{i=1}^N \boldsymbol{\Gamma}_{xi}/N$, and $\bar{\boldsymbol{v}}_t = \boldsymbol{\Upsilon}\,\mathrm{vec}(\boldsymbol{V}'_{.t}) = \sum_{i=1}^N \boldsymbol{v}_{it}/N$. Assuming that $\bar{\boldsymbol{\Gamma}}_x$ has a full row rank, namely, rank $(\bar{\boldsymbol{\Gamma}}_x) = r_x \leq k$ for all N including $N \to \infty$, we have:

$$\boldsymbol{f}_{xt} = (\bar{\boldsymbol{\Gamma}}_x \bar{\boldsymbol{\Gamma}}'_x)^{-1} \bar{\boldsymbol{\Gamma}}_x (\bar{\boldsymbol{x}}_t - \bar{\boldsymbol{v}}_t). \tag{12}$$

Under Assumption 2, as $N \to \infty$, $\bar{\boldsymbol{v}}_t$ converges to $\boldsymbol{0}$ in quadratic mean (Lemma 1 in the Online Supplement).[2] Hence, under Assumption 3, $\boldsymbol{f}_{xt}$ can be approximated by $\bar{\boldsymbol{x}}_t$:

$$\boldsymbol{f}_{xt} = (\bar{\boldsymbol{\Gamma}}_{x0} \bar{\boldsymbol{\Gamma}}'_{x0})^{-1} \bar{\boldsymbol{\Gamma}}_{x0} \bar{\boldsymbol{x}}_t + o_p(1),$$

where $\bar{\boldsymbol{\Gamma}}_x \xrightarrow{p} \bar{\boldsymbol{\Gamma}}_{x0} = \mathrm{E}(\boldsymbol{\Gamma}_{xi})$ as $N \to \infty$. Augmenting (1) with $\bar{\boldsymbol{x}}_t$ only may leave $\boldsymbol{f}_{2t}$ unaccounted for, in which case we may concern about any efficiency loss. However, if the dimension of $\boldsymbol{f}_{2t}$ is larger than 1, the use of $\bar{y}_t$ does not always achieve efficiency gain because $\bar{y}_t$ may not provide good approximation to $\boldsymbol{f}_{2t}$.[3]

Now, we make the following assumptions:

Assumption 1. *The $r_y \times 1$ vector of unobserved factors, $\boldsymbol{f}_{yt} = (\boldsymbol{f}'_{1t}, \boldsymbol{f}'_{2t})'$ and the $r_x \times 1$ vector of unobserved factors, $\boldsymbol{f}_{xt} = (\boldsymbol{f}'_{1t}, \boldsymbol{f}'_{3t})'$ are covariance stationary with absolutely summable autocovariances, and distributed independently of the idiosyncratic errors, ε_{is} and $\boldsymbol{v}_{is}$, for all i, t and s. Furthermore, $\boldsymbol{F}'_y \boldsymbol{F}_y / T$ and $\boldsymbol{F}'_x \boldsymbol{F}_x / T$ are nonsingular for all T, where $\boldsymbol{F}_y = \left(\boldsymbol{f}_{y1}, \ldots, \boldsymbol{f}_{yT}\right)'$ and $\boldsymbol{F}_x = \left(\boldsymbol{f}_{x1}, \ldots, \boldsymbol{f}_{xT}\right)'$.*

Assumption 2. *The idiosyncratic errors ε_{it} and $\boldsymbol{v}_{js}$ are cross-sectionally uncorrelated and distributed independently of each other, for all i, j, t, s. For each i, $\{\varepsilon_{it}\}_{t=1}^T$ and $\{\boldsymbol{v}_{it}\}_{t=1}^T$ follow stationary linear processes: $\varepsilon_{it} = \sum_{l=0}^\infty a_{il} \epsilon_{i,t-l}$ and $\boldsymbol{v}_{it} = \sum_{l=0}^\infty \boldsymbol{B}_{il} \boldsymbol{\zeta}_{i,t-l}$, where $(\epsilon_{it}, \boldsymbol{\zeta}'_{it})' \sim IID(\boldsymbol{0}_{k+1}, \boldsymbol{I}_{k+1})$, over both i and t, with finite fourth-order moments. Furthermore, $\{\varepsilon_{it}\}_{t=1}^T$ and $\{\boldsymbol{v}_{it}\}_{t=1}^T$ have absolutely summable autocovariances uniformly in i. In particular, let $\sigma^2_{\varepsilon,i} := \mathrm{var}(\varepsilon_{it}) = \sum_{l=0}^\infty a_{il}^2$ and $\boldsymbol{\Omega}_{v,i} := \mathrm{var}(\boldsymbol{v}_{it}) = \sum_{l=0}^\infty \boldsymbol{B}_{il} \boldsymbol{B}'_{il}$. Then, for all i, $\sigma^2_{\varepsilon,i} > 0$ and $\boldsymbol{\Omega}_{v,i}$ is a positive definite matrix, and there exists a positive constant C such that $|\sigma^2_{\varepsilon,i}| \leq C$ and $\|\boldsymbol{\Omega}_{v,i}\| \leq C$.*

Assumption 3. *The factor loadings, $\boldsymbol{\gamma}_{yi}$ and $\boldsymbol{\Gamma}_{xi}$, $1 \leq i \leq N$, are cross-sectionally independently distributed and uniformly bounded in probability and follow the following processes: $\boldsymbol{\gamma}_{yi} \sim$*

[2] This holds with unequal weights if the granular condition in Pesaran (2006) is satisfied.

[3] The rank condition specified in Pesaran (2006) fails in this situation.

$IID(\mathbf{0},\ \boldsymbol{\Omega}_\gamma)$ and $\boldsymbol{\Gamma}_{xi} = \bar{\boldsymbol{\Gamma}}_{x0} + \boldsymbol{\Xi}_{\Gamma,i}$, where $\boldsymbol{\Xi}_{\Gamma,i} \sim IID(\mathbf{0},\ \boldsymbol{\Omega}_\Gamma)$ is independent of $\boldsymbol{\gamma}_{yj}$ for all i and j, $\underset{r_x \times k}{\bar{\boldsymbol{\Gamma}}_{x0}} = (\underset{k\times r_1}{\bar{\boldsymbol{\Gamma}}'_{10}}, \underset{k\times r_3}{\bar{\boldsymbol{\Gamma}}'_{30}})'$ is a matrix with full row rank, and $\boldsymbol{\Omega}_\gamma$ and $\boldsymbol{\Omega}_\Gamma$ are positive definite matrices. Both $\boldsymbol{\gamma}_{yi}$ and $\boldsymbol{\Xi}_{\Gamma,i}$ are independent of $\boldsymbol{\varepsilon}_{jt}$, $\boldsymbol{v}_{jt}$ and $\boldsymbol{f}_{at}$ for all i, j, and t. Moreover, the matrix $\bar{\boldsymbol{\Gamma}}_x = \sum_{i=1}^N \boldsymbol{\Gamma}_{xi}/N$ has full row rank for all N including $N \to \infty$.

Assumption 4. *The $N \times N$ spatial weighting matrix, $\boldsymbol{W} = (w_{ij})$ with $w_{ii} = 0$, has bounded row and column sum norms, i.e., $\|\boldsymbol{W}\|_\infty < C$ and $\|\boldsymbol{W}\|_1 < C$, and*

$$\rho_{sup} := \sup_i |\rho_i| < \max\{1/\|\boldsymbol{W}\|_1, 1/\|\boldsymbol{W}\|_\infty\}. \tag{13}$$

Assumption 5. *The parameters, $\boldsymbol{\theta}_i = (\rho_i, \boldsymbol{\beta}'_i)'$ follow a random coefficient model:*

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\xi}_i, \quad \boldsymbol{\xi}_i \sim\ IID(\mathbf{0}, \boldsymbol{\Omega}_\xi),\ i = 1, \dots, N, \tag{14}$$

where $\boldsymbol{\theta} := \mathrm{E}(\boldsymbol{\theta}_i) = (\mathrm{E}(\rho_i), \mathrm{E}(\boldsymbol{\beta_i})')' = (\rho, \boldsymbol{\beta}')'$ and $\boldsymbol{\Omega}_\xi$ is a positive definite matrix. Moreover, $\boldsymbol{\xi}_i$ is independent of $\boldsymbol{\gamma}_{yj}$, $\boldsymbol{\Gamma}_{xj}$, ε_{jt}, $\boldsymbol{v}_{jt}$, and $\mathbf{f}_{at}$ for all i, j and t.

Remark 1. Assumptions 1–3 and 5 are standard. Assumption 1 allows $\boldsymbol{f}_{1t}$, $\boldsymbol{f}_{2t}$, and $\boldsymbol{f}_{3t}$ to be correlated. If we use $\bar{y}_t$ and $\bar{\boldsymbol{x}}_t$ as factor proxies, then the means of both factor loadings, $\mathrm{E}(\boldsymbol{\gamma}_{yi})$ and $\mathrm{E}(\boldsymbol{\Gamma}_{xi})$, should be nonzero (see Pesaran (2006)). Since we propose using only $\bar{\boldsymbol{x}}_t$ as factor proxies, we require the weaker condition, $\mathrm{E}(\boldsymbol{\Gamma}_{xi}) = \bar{\boldsymbol{\Gamma}}_{x0} \neq \mathbf{0}$. The rank condition, $\mathrm{rank}(\bar{\boldsymbol{\Gamma}}_{x0}) = r_x \leq k$, is no more restrictive than the one imposed by Pesaran (2006) and Yang (2021).[4]

The IVs can be obtained by $\boldsymbol{X}_{.t}$ and their higher order spatial lagged terms such that

$$\underset{NT\times\iota}{\boldsymbol{Q}} = (\boldsymbol{Q}'_{.1}, \dots,\ \boldsymbol{Q}'_{.T})',$$

where $\boldsymbol{Q}_{.t}$ is an $N\times\iota$ ($\iota \geq (k+1)$) matrix consisting of the ι columns of the IV set $(\boldsymbol{X}_{.t}, \dots, \boldsymbol{W}^r\boldsymbol{X}_{.t}, \dots)$ for $r = 0, 1, 2, \cdots$ and for each t. To make the columns of $\boldsymbol{Q}_{.t}$ valid instruments, we first need to de-factorise them (see the discussion below (7)). Using $\bar{\boldsymbol{x}}_t$ as factor proxies, we construct the following de-factorisation matrix:

$$\tilde{\boldsymbol{M}}_{\bar{\boldsymbol{X}}} = \boldsymbol{M}_{\bar{\boldsymbol{X}}} \otimes \boldsymbol{I}_N \ \text{ with } \ \boldsymbol{M}_{\bar{\boldsymbol{X}}} = \boldsymbol{I}_T - \bar{\boldsymbol{X}}(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^+\bar{\boldsymbol{X}}', \tag{15}$$

where $\bar{\boldsymbol{X}} = (\bar{\boldsymbol{x}}_1, \dots, \bar{\boldsymbol{x}}_T)'$ and $(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^+$ is the Moore-Penrose inverse of $\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}$. We construct the $NT\times\iota$ matrix of instrumental variables, denoted $\tilde{\boldsymbol{Q}} = (\boldsymbol{M}_{\bar{\boldsymbol{X}}} \otimes \boldsymbol{I}_N)\boldsymbol{Q}$. Even if the factors, $\boldsymbol{f}_{2t}$ specific to y_{it}, may be left unaccounted for, this does not affect the validity of IVs, because $(\boldsymbol{M}_{\bar{\boldsymbol{X}}} \otimes \boldsymbol{I}_N)\boldsymbol{Q}$ is uncorrelated with $\boldsymbol{f}_{2t}$ under Assumption 1.

We next introduce the identification conditions for the individual coefficient, $\boldsymbol{\theta}_i$ and its mean,

[4] For convenience, we assume: $E(\boldsymbol{\gamma}_{yi}) = 0$ (e.g. Norkutė et al. (2021)). But as pointed out by an anonymous referee, we need only to assume $E(\boldsymbol{\gamma}_{2i}) = 0$. Furthermore, we can also develop our theory based on the assumption $E(\boldsymbol{f}_{2t}) = 0$ instead.

$\boldsymbol{\theta} = \mathrm{E}(\boldsymbol{\theta}_i)$. Define:

$$\underset{T\times\iota}{\tilde{\boldsymbol{Q}}_{i0}} = \boldsymbol{M}_{\boldsymbol{F}_x}(\boldsymbol{I}_T \otimes \boldsymbol{b}_i')\boldsymbol{Q} \quad \text{and} \quad \underset{T\times(k+1)}{\boldsymbol{Z}_{i0.}} = ((\boldsymbol{I}_T \otimes \boldsymbol{b}_i'\boldsymbol{G})(\boldsymbol{I}_T \otimes \boldsymbol{\mathcal{B}})\boldsymbol{x}, \boldsymbol{X}_{i.}),$$

where $\boldsymbol{M}_{\boldsymbol{F}_x} = \boldsymbol{I}_T - \boldsymbol{F}_x(\boldsymbol{F}_x'\boldsymbol{F}_x)^{-1}\boldsymbol{F}_x'$, $\boldsymbol{F}_x = (\boldsymbol{f}_{x1}, \ldots, \boldsymbol{f}_{xT})'$, $\boldsymbol{b}_i$ is the $N \times 1$ column vector with the i-th entry being 1 and 0 otherwise, $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{S}^{-1}$ and $\boldsymbol{X}_{i.} = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})'$.

Assumption 6. *For all i and $T \geq 1$:*

(i) The $\iota \times \iota$ matrix, $\tilde{\boldsymbol{Q}}_{i0}'\tilde{\boldsymbol{Q}}_{i0}/T$ is nonsingular and has bounded second order moment.

(ii) The $\iota \times (k+1)$ matrix, $\tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{Z}_{i0.}/T$ has full column rank and bounded second order moment.

(iii) The $\iota \times 1$ vector, $\tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{F}_2\gamma_{2i}/\sqrt{T}$ has bounded second order moment.

(iv) As $(N, T) \to \infty$, there exist an $\iota \times \iota$ nonsingular matrix $\boldsymbol{\Phi}_i$, an $\iota \times (k+1)$ full-column-rank matrix $\boldsymbol{\Psi}_i$, and an $\iota \times \iota$ positive definite matrix $\boldsymbol{\Sigma}_i$, such that

$$\boldsymbol{\Phi}_i = \operatorname*{plim}_{T\to\infty} \frac{\tilde{\boldsymbol{Q}}_{i0}'\tilde{\boldsymbol{Q}}_{i0}}{T}, \qquad \boldsymbol{\Psi}_i = \operatorname*{plim}_{T\to\infty} \frac{\tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{Z}_{i0.}}{T},$$
$$\boldsymbol{\Sigma}_i = \lim_{T\to\infty} \mathrm{E}\left(\frac{\tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{F}_2\boldsymbol{\Omega}_{\gamma_2}\boldsymbol{F}_2'\tilde{\boldsymbol{Q}}_{i0}}{T} + \frac{\tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{\Omega}_{\varepsilon,i}\tilde{\boldsymbol{Q}}_{i0}}{T}\right), \tag{16}$$

where $\boldsymbol{\Omega}_{\gamma_2} = \mathrm{E}\left(\boldsymbol{\gamma}_{2i}\boldsymbol{\gamma}_{2i}'\right)$ and $\boldsymbol{\Omega}_{\varepsilon,i} = \mathrm{E}\left(\boldsymbol{\varepsilon}_{i.}\boldsymbol{\varepsilon}_{i.}'\right)$.

Remark 2. Assumption 6(i)-(iii) to be held for all $T \geq 1$ is needed to derive consistency of the Mean Group and Pooled estimators for fixed T, see also Pesaran (2006) and Norkutė et al. (2021). If we are only concerned with large T asymptotics, Assumption 6(i)-(iii) needs to hold as $T \to \infty$. Assumption 6(iii) is implied by Assumptions 1-4, but we keep it to facilitate the proofs for Theorems. Assumption 6(iv) ensures the existence of the asymptotic variance of the individual estimators in Theorem 1.

Remark 3. As pointed out by one of the referees, if $\boldsymbol{\mathcal{B}} = 0$ in (3), then the IVs defined above are invalid and the full rank condition in Assumption 6 (ii) fails. To account for this situation, we follow Lee and Yu (2014) and Yang (2021), and provide some discussions on how to develop the GMM estimation using both linear and quadratic moment conditions in Section S1 in the Online Supplement. But, the construction of quadratic moments for each individual requires consistent estimation of idiosyncratic errors and the parameters from all other individual regressions. This implies that all the parameters need to be determined simultaneously, which would pose a heavy computational burden, especially if N is large. Furthermore, the construction of optimal quadratic moment conditions relies on an analysis of the asymptotic properties of the GMM estimator. This analysis is challenging because of parameter heterogeneity and beyond the scope of the current study. Also, due to space constraints, we leave this important topic for future research.

3 Asymptotic Theory

3.1 The Individual Estimator

Stacking equations (1) and (2) over t, we have:

$$\boldsymbol{y}_{i.} = \rho_i \boldsymbol{y}_{i.}^* + \boldsymbol{X}_{i.}\boldsymbol{\beta}_i + \boldsymbol{F}_y \boldsymbol{\gamma}_{yi} + \boldsymbol{\varepsilon}_{i.} = \boldsymbol{Z}_{i.}\boldsymbol{\theta}_i + \boldsymbol{F}_y \boldsymbol{\gamma}_{yi} + \boldsymbol{\varepsilon}_{i.}, \tag{17}$$

$$\boldsymbol{X}_{i.} = \boldsymbol{F}_x \boldsymbol{\Gamma}_{xi} + \boldsymbol{V}_{i.}, \tag{18}$$

where $\boldsymbol{y}_{i.} = (y_{i1}, \ldots, y_{iT})'$, $\boldsymbol{y}_{i.}^* = (\boldsymbol{I}_T \otimes \boldsymbol{w}_i)\boldsymbol{y} = (y_{i1}^*, \ldots, y_{iT}^*)'$, $\boldsymbol{w}_i = \boldsymbol{b}_i'\boldsymbol{W}$ is the i-th row of $\boldsymbol{W}$, $\boldsymbol{Z}_{i.} = (\boldsymbol{y}_{i.}^*, \boldsymbol{X}_{i.})$ and $\boldsymbol{V}_{i.} = (\boldsymbol{v}_{i1}, \ldots, \boldsymbol{v}_{iT})'$. We pre-multiply (17) by $\boldsymbol{M}_{\bar{x}}$ defined in (15). Under this transformation, $\boldsymbol{f}_{2t}$ are not removed from (17). Nevertheless, by augmenting (17) with $\bar{\boldsymbol{X}}$, we can consistently estimate $\boldsymbol{\theta}_i$ using the IVs, $\tilde{\boldsymbol{Q}}_i = \boldsymbol{M}_{\bar{\boldsymbol{X}}}(\boldsymbol{I}_T \otimes \boldsymbol{b}_i')\boldsymbol{Q}$.[5] The individual CCEX-IV estimator is given by

$$\hat{\boldsymbol{\theta}}_i = \left(\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{Z}_{i.}\right)^{-1}\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{y}_{i.}, \tag{19}$$

where $\boldsymbol{\Pi}_i = \tilde{\boldsymbol{Q}}_i(\tilde{\boldsymbol{Q}}_i'\tilde{\boldsymbol{Q}}_i)^{-1}\tilde{\boldsymbol{Q}}_i'$. Since $\boldsymbol{M}_{\bar{\boldsymbol{X}}}$ is an idempotent matrix, we have $\boldsymbol{Z}_{i.}'\boldsymbol{M}_{\bar{\boldsymbol{X}}}\tilde{\boldsymbol{Q}}_i = \boldsymbol{Z}_{i.}'\tilde{\boldsymbol{Q}}_i$. This suggests that we can estimate $\boldsymbol{\theta}_i$ equally in two different ways. The first uses $\tilde{\boldsymbol{Q}}_i$ as IVs in the de-factored regression of $\boldsymbol{M}_{\bar{\boldsymbol{X}}}\boldsymbol{y}_{i.}$ against $\boldsymbol{M}_{\bar{\boldsymbol{X}}}\boldsymbol{Z}_{i.}$. The second applies the IVs, $\tilde{\boldsymbol{Q}}_i$ directly to estimate (17). Substituting (17) into (19), we obtain:

$$\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i = \left(\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{Z}_{i.}}{T}\right)^{-1}\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i(\boldsymbol{F}_1\boldsymbol{\gamma}_{1i} + \boldsymbol{F}_2\boldsymbol{\gamma}_{2i} + \boldsymbol{\varepsilon}_{i.})}{T}. \tag{20}$$

As shown in Section S2.1 in the Online Supplement, the right hand side (RHS) of (20) converges in probability to $\mathbf{0}$ and follows a limiting normal distribution for each i.

Theorem 1. *Consider the heterogeneous spatial panel data model with common factors given by (1) and (2). Under Assumptions 1–4 and 6(i)–(iii), as $(N,T)\to\infty$, the individual estimator, $\hat{\boldsymbol{\theta}}_i$ in (19), is consistent for $\boldsymbol{\theta}_i$. In addition, if Assumption 6(iv) holds and $T/N^2 \to 0$, then as $(N,T) \to \infty$,*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{\Omega}_i),$$

where $\boldsymbol{\Omega}_i = \left(\boldsymbol{\Psi}_i'\boldsymbol{\Phi}_i^{-1}\boldsymbol{\Psi}_i\right)^{-1}\left(\boldsymbol{\Psi}_i'\boldsymbol{\Phi}_i^{-1}\boldsymbol{\Sigma}_i\boldsymbol{\Phi}_i^{-1}\boldsymbol{\Psi}_i\right)\left(\boldsymbol{\Psi}_i'\boldsymbol{\Phi}_i^{-1}\boldsymbol{\Psi}_i\right)^{-1}$, and $\boldsymbol{\Psi}_i$, $\boldsymbol{\Phi}_i$, and $\boldsymbol{\Sigma}_i$ are defined in Assumption 6(iv).

[5] The instruments for the i-th individual regression can be expressed interchangeably as

$$(\boldsymbol{I}_T \otimes \boldsymbol{b}_i')(\boldsymbol{M}_{\bar{\boldsymbol{X}}} \otimes \boldsymbol{I}_N)\boldsymbol{Q} = (\boldsymbol{M}_{\bar{\boldsymbol{X}}} \otimes 1)(\boldsymbol{I}_T \otimes \boldsymbol{b}_i')\boldsymbol{Q} = \boldsymbol{M}_{\bar{\boldsymbol{X}}}(\boldsymbol{I}_T \otimes \boldsymbol{b}_i')\boldsymbol{Q}.$$

A consistent estimator for $\boldsymbol{\Omega}_i$, is given by

$$\hat{\boldsymbol{\Omega}}_i = \left(\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{Z}_{i.}}{T}\right)^{-1}\left(\frac{\boldsymbol{Z}_{i.}'\tilde{\boldsymbol{Q}}_i}{T}\right)\left(\frac{\tilde{\boldsymbol{Q}}_i'\tilde{\boldsymbol{Q}}_i}{T}\right)^{-1}\hat{\boldsymbol{\Sigma}}_i\left(\frac{\tilde{\boldsymbol{Q}}_i'\tilde{\boldsymbol{Q}}_i}{T}\right)^{-1}\left(\frac{\tilde{\boldsymbol{Q}}_i'\boldsymbol{Z}_{i.}}{T}\right)\left(\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{Z}_{i.}}{T}\right)^{-1}, \tag{21}$$

and $\hat{\boldsymbol{\Sigma}}_i$ can be constructed by the Newey and West (1987) robust estimator:

$$\hat{\boldsymbol{\Sigma}}_i = \hat{\boldsymbol{\Sigma}}_{i,0} + \sum_{h=1}^{p_T}\left(1 - \frac{h}{p_T+1}\right)\left(\hat{\boldsymbol{\Sigma}}_{i,h} + \hat{\boldsymbol{\Sigma}}_{i,h}'\right),$$

where $\hat{\boldsymbol{\Sigma}}_{i,h} = \sum_{t=h+1}^{T}\hat{e}_{it}\hat{e}_{i,t-h}\tilde{\boldsymbol{q}}_{it}\tilde{\boldsymbol{q}}_{i,t-h}'/T$, p_T is the bandwidth of the Bartlett kernel, $\hat{\boldsymbol{e}}_{i.} = \boldsymbol{M}_{\bar{\boldsymbol{X}}}(\boldsymbol{y}_{i.} - \boldsymbol{Z}_{i.}\hat{\boldsymbol{\theta}}_i) = (\hat{e}_{i1}, \ldots, \hat{e}_{iT})'$, and the $\iota \times 1$ vector $\tilde{\boldsymbol{q}}_{it}$ is the t-th column of $\tilde{\boldsymbol{Q}}_i'$.

3.2 The Mean Group Estimator

Consider the CCEX-IV Mean Group estimator for $\boldsymbol{\theta}$ defined as

$$\hat{\boldsymbol{\theta}}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_i. \tag{22}$$

Under Assumption 5 and using (20), we expand (22) as follows:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta} &= \frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_i = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\xi}_i + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{Z}_{i.}}{T}\right)^{-1}\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{F}_1\gamma_{1i}}{T}\\ &+ \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{Z}_{i.}}{T}\right)^{-1}\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{F}_2\gamma_{2i}}{T} + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{Z}_{i.}}{T}\right)^{-1}\frac{\boldsymbol{Z}_{i.}'\boldsymbol{\Pi}_i\boldsymbol{\varepsilon}_{i.}}{T}.\end{aligned} \tag{23}$$

As shown in Section S2.2 in Online Supplement, the first term, $\sum_{i=1}^{N}\boldsymbol{\xi}_i/N$ dominates and determines the asymptotic property of the Mean Group estimator.

Theorem 2. *Consider the heterogeneous spatial panel data model with common factors given by (1) and (2). Under Assumptions 1 – 5 and 6(i)–(iii), as $N \to \infty$, the Mean Group estimator $\hat{\boldsymbol{\theta}}_{MG}$ in (22), is consistent for $\boldsymbol{\theta}$ for either fixed T or $T \to \infty$. Furthermore, as $(N,T)\to\infty$, we have*

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{MG} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Omega}_{MG}\right),$$

where $\boldsymbol{\Omega}_{MG} = \boldsymbol{\Omega}_{\xi} = \mathrm{var}(\boldsymbol{\xi}_i)$.

We can estimate $\boldsymbol{\Omega}_{\xi}$ consistently by the nonparametric estimator (Pesaran (2006)):

$$\hat{\boldsymbol{\Omega}}_{MG} = \hat{\boldsymbol{\Omega}}_{\xi} = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{MG})(\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{MG})'. \tag{24}$$

3.3 The Pooled Estimator

To derive the CCEX-IV Pooled estimator for $\boldsymbol{\theta}$, we de-factor the data by pre-multiplying by $\boldsymbol{M}_{\bar{\boldsymbol{X}}}$, and obtain $\tilde{\boldsymbol{y}} = (\tilde{\boldsymbol{y}}_1', \cdots, \tilde{\boldsymbol{y}}_N')'$ with $\tilde{\boldsymbol{y}}_i = \boldsymbol{M}_{\bar{\boldsymbol{X}}}\boldsymbol{y}_{i.}$, and $\tilde{\boldsymbol{Z}} = (\tilde{\boldsymbol{Z}}_1', \cdots, \tilde{\boldsymbol{Z}}_N')'$, with $\tilde{\boldsymbol{Z}}_i = \boldsymbol{M}_{\bar{\boldsymbol{X}}}\boldsymbol{Z}_{i.}$ where $\boldsymbol{y}_{i.}$ and $\boldsymbol{Z}_{i.}$ are defined in (17). We then run the IV regression of $\tilde{\boldsymbol{y}}$ on $\tilde{\boldsymbol{Z}}$ using the $NT \times \iota$ matrix of IVs, $\tilde{\boldsymbol{Q}} = (\tilde{\boldsymbol{Q}}_1', \cdots, \tilde{\boldsymbol{Q}}_N')'$ where $\tilde{\boldsymbol{Q}}_i = \boldsymbol{M}_{\bar{\boldsymbol{X}}}(\boldsymbol{I}_T \otimes \boldsymbol{b}_i')\boldsymbol{Q}$. The Pooled estimator for $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}_P = (\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Q}}(\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{Q}})^{-1}\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{Z}})^{-1}\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Q}}(\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{Q}})^{-1}\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{y}}. \tag{25}$$

Under Assumption 5 and using (17) and (25), we have:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_P - \boldsymbol{\theta} = &\left[\left(\frac{1}{N}\sum_{i=1}^N \frac{\boldsymbol{Z}_{i.}'\tilde{\boldsymbol{Q}}_i}{T}\right)\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_i'\tilde{\boldsymbol{Q}}_i}{T}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_i'\boldsymbol{Z}_{i.}}{T}\right)\right]^{-1} \\ &\left(\frac{1}{N}\sum_{i=1}^N \frac{\boldsymbol{Z}_{i.}'\tilde{\boldsymbol{Q}}_i}{T}\right)\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_i'\tilde{\boldsymbol{Q}}_i}{T}\right)^{-1} \frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_i'(\boldsymbol{Z}_{i.}\boldsymbol{\xi}_i + \boldsymbol{F}_1\gamma_{1i} + \boldsymbol{F}_2\gamma_{2i} + \boldsymbol{\varepsilon}_{i.})}{T}.\end{aligned} \tag{26}$$

The Pooled estimator is inconsistent, due to the non-vanishing correlation between $\boldsymbol{\xi}_i$ and $\boldsymbol{y}_{i.}^*$ in $\boldsymbol{Z}_{i.}$, see Remark 4 and Section S2.3 in the Online Supplement.

In the special case where the parameters are homogeneous, i.e., $\boldsymbol{\theta}_i \equiv \boldsymbol{\theta}$ for all i, the Pooled estimator, denoted as $\hat{\boldsymbol{\theta}}_P^*$, has the following representation:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_P^* - \boldsymbol{\theta} = &\left[\left(\frac{1}{N}\sum_{i=1}^N \frac{\boldsymbol{Z}_{i.}'\tilde{\boldsymbol{Q}}_i}{T}\right)\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_i'\tilde{\boldsymbol{Q}}_i}{T}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_i'\boldsymbol{Z}_{i.}}{T}\right)\right]^{-1} \\ &\left(\frac{1}{N}\sum_{i=1}^N \frac{\boldsymbol{Z}_{i.}'\tilde{\boldsymbol{Q}}_i}{T}\right)\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_i'\tilde{\boldsymbol{Q}}_i}{T}\right)^{-1} \frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_i'(\boldsymbol{F}_1\gamma_{1i} + \boldsymbol{F}_2\gamma_{2i} + \boldsymbol{\varepsilon}_{i.})}{T}.\end{aligned} \tag{27}$$

To establish the asymptotic properties of $\hat{\boldsymbol{\theta}}_P^*$, we assume:

Assumption 7. *(i) For all $T \geq 1$, as $N \to \infty$, the $\iota \times \iota$ matrix, $\sum_{i=1}^N \tilde{\boldsymbol{Q}}_{i0}'\tilde{\boldsymbol{Q}}_{i0}/NT$ is non-singular, and the $\iota \times (k+1)$ matrix, $\sum_{i=1}^N \tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{Z}_{i0.}/NT$ has a full column rank.*

(ii) As $(N,T) \to \infty$, there exist an $\iota \times \iota$ non-singular matrix, $\boldsymbol{\Phi}$, an $\iota \times (k+1)$ matrix, $\boldsymbol{\Psi}$ with the full column rank, and a positive definite matrix. $\boldsymbol{\Sigma}_P^*$ such that

$$\boldsymbol{\Phi} = \operatorname*{plim}_{(N,T)\to\infty} \frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_{i0}'\tilde{\boldsymbol{Q}}_{i0}}{T}, \qquad \boldsymbol{\Psi} = \operatorname*{plim}_{(N,T)\to\infty} \frac{1}{N}\sum_{i=1}^N \frac{\tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{Z}_{i0.}}{T},$$

$$\boldsymbol{\Sigma}_P^* = \lim_{(N,T)\to\infty} \frac{1}{N}\sum_{i=1}^N \mathrm{E}\left(\frac{\tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{F}_2\boldsymbol{\Omega}_{\gamma_2}\boldsymbol{F}_2'\tilde{\boldsymbol{Q}}_{i0}}{T} + \frac{\tilde{\boldsymbol{Q}}_{i0}'\boldsymbol{\Omega}_{\varepsilon,i}\tilde{\boldsymbol{Q}}_{i0}}{T}\right).$$

Theorem 3. *Consider the spatial panel data model with common factors given by (1) and (2). Suppose that the parameters are heterogeneous and follow the random coefficient model (14). Then,* *the Pooled estimator $\hat{\boldsymbol{\theta}}_P$ in (25) is inconsistent for $\boldsymbol{\theta}$. In the case where the parameters are homogeneous, i.e., $\boldsymbol{\theta}_i \equiv \boldsymbol{\theta}$ for all i, the Pooled estimator $\hat{\boldsymbol{\theta}}_P^*$ is consistent as $N \to \infty$ for fixed T or $T \to \infty$ under Assumptions 1–4, 6(i)–(iii) and 7(i). In addition, if Assumption 7(ii) holds and $T/N^2 \to 0$, then as $(N,T)\to\infty$,*

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}_P^* - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{\Omega}_P^*),$$

where $\boldsymbol{\Omega}_P^* = \left(\boldsymbol{\Psi}'\boldsymbol{\Phi}^{-1}\boldsymbol{\Psi}\right)^{-1}\left(\boldsymbol{\Psi}'\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}_P^*\boldsymbol{\Phi}^{-1}\boldsymbol{\Psi}\right)\left(\boldsymbol{\Psi}'\boldsymbol{\Phi}^{-1}\boldsymbol{\Psi}\right)^{-1}$.

Following Cui et al. (2019) we estimate $\boldsymbol{\Omega}_P^*$ by

$$\hat{\boldsymbol{\Omega}}_P^* = \left(\frac{\tilde{\boldsymbol{Z}}'\boldsymbol{\Pi}\tilde{\boldsymbol{Z}}}{NT}\right)^{-1}\left(\frac{\tilde{\boldsymbol{Z}}'\tilde{\boldsymbol{Q}}}{NT}\right)\left(\frac{\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{Q}}}{NT}\right)^{-1}\hat{\boldsymbol{\Sigma}}_P^*\left(\frac{\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{Q}}}{NT}\right)^{-1}\left(\frac{\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{Z}}}{NT}\right)\left(\frac{\tilde{\boldsymbol{Z}}'\boldsymbol{\Pi}\tilde{\boldsymbol{Z}}}{NT}\right)^{-1}, \tag{28}$$

where $\boldsymbol{\Pi} = \tilde{\boldsymbol{Q}}(\tilde{\boldsymbol{Q}}'\tilde{\boldsymbol{Q}})^{-1}\tilde{\boldsymbol{Q}}'$ and

$$\hat{\boldsymbol{\Sigma}}_P^* = \frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\boldsymbol{Q}}_i'\hat{\boldsymbol{e}}_{i.}^*\hat{\boldsymbol{e}}_{i.}^{*\prime}\tilde{\boldsymbol{Q}}_i}{T}, \quad \hat{\boldsymbol{e}}_{i.}^* = \boldsymbol{y}_{i.} - \boldsymbol{Z}_{i.}. \tag{29}$$

Remark 4. It is the interplay between parameter heterogeneity and the spatial autoregressive structure that renders the Pooled estimator of $\boldsymbol{\theta}$ in (25) inconsistent. This is a new finding in the literature, which extends the seminal work by Pesaran and Smith (1995) who establish that the Pooled estimator is inconsistent in the dynamic panels with the heterogeneous parameters. To see this, we write the i-th individual de-factored regression at time t from (6) as follows:

$$\tilde{y}_{it} = \rho_i\tilde{y}_{it}^* + \tilde{\boldsymbol{x}}_{it}'\boldsymbol{\beta}_i + \tilde{\varepsilon}_{it} = \rho\tilde{y}_{it}^* + \tilde{\boldsymbol{x}}_{it}'\boldsymbol{\beta} + \tilde{e}_{it}, \tag{30}$$

where $\tilde{e}_{it} = \xi_{\rho i}\tilde{y}_{it}^* + \tilde{\boldsymbol{x}}_{it}'\boldsymbol{\xi}_{\boldsymbol{\beta}_i} + \tilde{\varepsilon}_{it}$ and we have decomposed the error term, $\boldsymbol{\xi}_i$ of the random coefficients in (14) as $\boldsymbol{\xi}_i = (\xi_{\rho_i}, \boldsymbol{\xi}_{\boldsymbol{\beta}_i}')'$. Then, the Pooled estimator for $\boldsymbol{\theta}$ defined in (25) is equivalent to applying the 2SLS estimation to (30) with the new composite error term, $\tilde{e}_{it}$. Notice that the mean of $\xi_{\rho_i}\tilde{y}_{it}^*$ in $\tilde{e}_{it}$ is no longer zero because $\tilde{y}_{it}^*$ contains ξ_{ρ_i}. More importantly, the IVs in $\tilde{\boldsymbol{Q}}$ become correlated with $\tilde{e}_{it}$ as $\tilde{\boldsymbol{Q}}$ is correlated with $\tilde{y}_{it}^*$. Thus, the Pooled estimator in (25) is inconsistent (see Section S2.3 in the Online Supplement). Notice that the term, $\tilde{\boldsymbol{x}}_{it}'\boldsymbol{\xi}_{\boldsymbol{\beta}_i}$ is uncorrelated with $\tilde{\boldsymbol{Q}}$ under Assumption 5. Hence, in the mixed case where the spatial coefficient is homogeneous (i.e., $\xi_{\rho_i} = 0$ for all i) but the slope coefficients $\boldsymbol{\beta}_i$ are heterogeneous, the Pooled estimator in (25) is still consistent. In the special case where all the parameters are homogeneous, the Pooled estimator in (27) becomes $\sqrt{NT}$-consistent.

Remark 5. The second result in Theorem 3 is slightly different from Theorem 1 in Yang (2021). This is mainly because Yang (2021) assumes that y_{it} and $\boldsymbol{x}_{it}$ share the same common factors. Importantly, the CCEX approach only requires the condition, $T/N^2 \to 0$, as the order for the third term in (S.22) is $O_p(1/\sqrt{N})$ due to the independence between $\bar{\boldsymbol{V}} = (\bar{\boldsymbol{v}}_{.1}, \ldots, \bar{\boldsymbol{v}}_{.T})'$ and $\boldsymbol{\varepsilon}_{i.}$. But, if $\bar{y}_t$ is also included as a factor proxy, the order of this term becomes $O_p\left(\sqrt{T/N}\right)$ because the factor approximation error is correlated with $\boldsymbol{\varepsilon}_{i\cdot}$. In this case, one requires the stronger condition, $T/N \to 0$.

4 Monte Carlo Simulations

We investigate the finite sample performance of the CCEX-IV and CCE-IV individual, Mean Group and Pooled estimators. To make the simulation design realistic, we construct the DGP in a data-oriented manner by calibrating the values of the parameters following our empirical application:

$$y_{it} = \rho_i y_{it}^* + \beta_{1i} x_{it1} + \beta_{2i} x_{it2} + \gamma_{1i} f_{1t} + \gamma_{2i} f_{2t} + \kappa_1 \varepsilon_{it}, \tag{31}$$

$$x_{it,p} = \Gamma_{1i,p} f_{1t} + \Gamma_{3i,p} f_{3t} + \kappa_2 v_{it,p}, \; p = 1, 2, \tag{32}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where we set the number of regressors at 2 and unobserved factors in both equations for y_{it} and $\boldsymbol{x}_{it}$ at 2.

We conduct four Monte Carlo experiments (see Table 1). In Experiments 1 and 2 we set $f_{2t} = f_{3t}$ while we generate f_{2t} and f_{3t} independently in Experiments 3 and 4. Each factor is generated as an AR(1) process:

$$f_{r,t} = \phi_{f_r} f_{r,t-1} + \xi_{f_{rt}}, \; t = -49, \dots, T; \; r = 1, 2, 3,$$

with $\phi_{f_r} = 0.5$ and $\xi_{f_{rt}} \sim IIDN(0, 1 - \phi_{f_r}^2)$. We discard the first 50 observations as the burn-in sample. We generate the factor loadings from $IIDN(0, 0.5)$.[6] In Experiments 1 and 3, we set the homogeneous parameters as

$$\rho_i = \rho = 0.5, \quad \beta_{i1} = \beta_1 = 1, \; \beta_{i2} = \beta_2 = 0.5, \; i = 1, \dots, N,$$

whereas in Experiments 2 and 4, we generate the heterogeneous parameters by

$$\rho_i = 0.5 + \xi_{\rho_i}, \quad \beta_{i1} = 1 + \xi_{\beta_{i1}}, \quad \beta_{i2} = 0.5 + \xi_{\beta_{i2}}, \; i = 1, \dots, N,$$

where $\rho_i \sim \rho + IIDU(-0.2, 0.2)$, $\xi_{\beta_{i1}} \sim IIDN(0, 0.5)$ and $\xi_{\beta_{i2}} \sim IIDN(0, 0.3)$. We also consider a higher spatial dependence with $E(\rho_i) = 0.8$.

Table 1: Experiment Settings

		Factors for y_{it} and $\boldsymbol{x}_{it}$	
		Same	Different
Coefficients	Homogeneous	Experiment 1	Experiment 3
	Heterogeneous	Experiment 2	Experiment 4

[6]We have also provided the simulation results for non-zero mean factor loadings with $\boldsymbol{\gamma}_{1i}(\boldsymbol{\gamma}_{2i}) \sim IIDN(0.5, 0.5)$ in Section S4.5 in the Online Supplement. Overall results are qualitatively similar to those reported here though RMSEs of CCE-IV estimator are more improved.

We allow idiosyncratic errors in (31) and (32) to be heteroskedastic and serially correlated. For ε_{it}, the first half cross-section units are generated from an AR(1) process:

$$\varepsilon_{it} = \phi_{\varepsilon_i}\varepsilon_{i,t-1} + \sigma_i(1-\phi_{\varepsilon_i}^2)^{0.5}\xi_{\varepsilon_{it}}, \quad i = 1,\dots \lfloor N/2 \rfloor;\ t = -49,\dots,T,$$

whilst the second half cross-section units are from a MA(1) process:

$$\varepsilon_{it} = \sigma_i(1+\psi_{\varepsilon_i}^2)^{0.5}(\xi_{\varepsilon_{it}} + \psi_{\varepsilon_i}\xi_{\varepsilon_{it-1}}), \quad i = \lfloor N/2 \rfloor + 1,\dots,N; \quad t = -49,\dots,T,$$

where $\phi_{\varepsilon_i} = \psi_{\varepsilon_i} = 0.5$ for all i, $\sigma_i^2 \sim IIDU(0.5, 1.5)$, $\xi_{\varepsilon_{it}} \sim IIDN(0,1)$, and $\lfloor . \rfloor$ denotes the integer part. We generate each component of $\boldsymbol{v}_{it}$ as an AR(1) process:

$$v_{it,p} = \phi_{v_{i,p}} v_{i,t-1,p} + \xi_{v_{i,p}},\ i = 1,\dots,N;\ t = -49,\dots,T;\ p = 1,2,$$

where $\phi_{v_{i,p}} = 0.5$ and $\xi_{v_{i,p}} \sim IIDN(0, 1-\phi_{v_{i,p}}^2)$ for all i and p. We also discard the first 50 observations of ε_{it} and $\boldsymbol{v}_{it}$.

The κ_1 in (31) and κ_2 in (32) are used to control the proportion of the variance of the unobserved components due to the idiosyncratic errors, and we set $\kappa_1 = 2$ and $\kappa_2 = 3$.[7] For the spatial weighting matrix, $\boldsymbol{W}$, we use the standard h-ahead-and-h-behind neighbour specification, i.e., its elements are zero apart from those off-diagonal elements that are within h rows on either side of the principal diagonal, which are set to be $1/2h$. Then, we apply the row-sum normalisation. In order to investigate the (robust) performance of the CCEX-IV and CCE-IV estimators against the different degrees of sparsity of the spatial weighting matrix, we consider three different values of $h = 2, 6, 0.3N$. In particular, the design with $h = 0.3N$ implies that the overall level of spatial dependence does not decay even as N increases. Each experiment is replicated 1,000 times for each (N, T) pair with $N = \{20, 50, 100\}$ and $T = \{20, 50, 100\}$.

We employ the following estimation algorithm:

- **Step 1: Construction of Factor Proxies, $\hat{\boldsymbol{F}}$**. We consider three cases:[8] (i) the CCEX estimator using $\hat{\boldsymbol{F}}_t = \bar{\boldsymbol{x}}_t$; (ii) the CCE estimator using $\hat{\boldsymbol{F}}_t = (\bar{y}_t, \bar{\boldsymbol{x}}_t')'$; and (iii) the infeasible estimator using the true factors.

- **Step 2: De-factorisation of the Data**. To deal with endogeneity caused by the correlation between the regressors and the unobserved factors, we run regressions of $\boldsymbol{y}_{i.}$, $\boldsymbol{y}_{i.}^*$, and $\boldsymbol{X}_{i.}$ on factor proxies, $\hat{\boldsymbol{F}}$, and save the respective residuals.

- **Step 3: Construction of the IVs**. To deal with endogeneity caused by the spatial lagged term, we use the instruments $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^*)$, where $\tilde{\boldsymbol{X}} = (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{I}_N)\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}^* = (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{I}_N)(\boldsymbol{I}_T \otimes$

[7] We have provided the simulation results for $\kappa_1 = 1$ and $\kappa_2 = 2$ in Section S4.6 in the Online Supplement, finding that the results are qualitatively similar.

[8] We have also used $(\bar{y}_t, \bar{y}_t^*, \bar{\boldsymbol{x}}_t')'$ as factor proxies, where $\bar{y}_t^* = \sum_{i=1}^N y_{it}^*/N$. The simulation results are generally worse than those reported here but available upon request.

$\boldsymbol{W})\boldsymbol{X}$ with $\boldsymbol{M}_{\hat{\boldsymbol{F}}} = \boldsymbol{I}_T - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'$ and $\boldsymbol{X} = (\boldsymbol{X}'_{.1}, \ldots, \boldsymbol{X}'_{.T})'$.[9]

- **Step 4: Computation of the individual, Mean Group and Pooled Estimates**. For Experiments 2 and 4 with the heterogeneous parameters, we obtain the individual estimates, $\hat{\boldsymbol{\theta}}_i = (\hat{\rho}_i, \hat{\beta}_{1i}, \hat{\beta}_{2i})'$, $i = 1, \ldots, N$, by (19) and the Mean Group estimate by (22). For Experiments 1 and 3 with the homogeneous parameters, we obtain the Pooled estimate by (25).

To evaluate the finite sample performance of the estimators, we calculate:

- **Bias** and **RMSE**. The bias and RMSE for the q-th element of $\hat{\boldsymbol{\theta}}_i$, denoted as $\hat{\theta}_{qi}, q = 1, 2, 3$, are defined as $\sum_{m=1}^{M}(\hat{\theta}_{qi}^m - \theta_{qi})/M$ and $\sqrt{\sum_{m=1}^{M}(\hat{\theta}_{qi}^m - \theta_{qi})^2/M}$, where M is the number of replications, θ_{qi} is the corresponding true value and $\hat{\theta}_{qi}^m$ is the m-th estimation result. For individual estimators we only report the average Bias and RMSE. Bias and RMSE of the Mean Group and Pooled estimators are also provided. For convenience the values of Bias and RMSE are multiplied by 100.

- **Size** and **Power**. The size of the t-test for an individual coefficient at the 5% significance level is evaluated as

$$\text{Size}_{\hat{\theta}_{qi}} = \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}\left(\frac{|\hat{\theta}_{qi}^m - \theta_{qi}|}{\hat{\sigma}_{\hat{\theta}_{qi}^m}} > 1.96\right),\ q = 1, 2, 3,$$

  where $\mathbb{1}(.)$ is the indicator function, and $\hat{\sigma}^2_{\hat{\theta}_{qi}^m}$ is the estimated variance using (21). Following the standard practice in the literature (e.g., Zhang and Boos (1994)), we report the size-adjusted power. For individual estimators, the average size and power are reported. The size and power of the Mean Group and Pooled estimators are evaluated using the variances estimated respectively by (24) and (28).

4.1 Simulation Results

We only report the simulation results for the individual and Mean Group estimators under Experiment 4.[10] Biases and RMSEs for the individual estimators are presented in Table 2. The biases of the CCEX-IV individual estimator of $(\rho_i, \beta_{1i}, \beta_{2i})$ are all relatively small in almost all cases, even in small samples with $(N, T) = (20, 20)$. These bias values are also close to those of the infeasible estimator. If the spatial weights matrix is sparse with $h = 2$, then the performance of the CCE-IV individual estimator is relatively satisfactory except for $N = 20$. As h and/or ρ rise (e.g. $h = 6$ and $\rho = 0.8$), its biases become substantially larger for $N = 20$, especially for the spatial coefficient, though the biases decline sharply with N. However, if we set h varying with N (i.e. $h = 0.3N$),

[9] We have also tried using $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^{2*})$, $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^{3*})$ and $(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{X}}^{*}, \tilde{\boldsymbol{X}}^{2*})$ as IVs, where $\tilde{\boldsymbol{X}}^{r*} = (\boldsymbol{M}_{\hat{\boldsymbol{F}}} \otimes \boldsymbol{I}_N)(\boldsymbol{I}_T \otimes \boldsymbol{W}^r)\boldsymbol{X}$. The results are quite similar to those presented here and available upon request.

[10] We provide the complete simulation results for Experiments 1 to 3 in Section S4 in the Online Supplement. Overall, the results are qualitatively similar to those reported here.

then the biases for the spatial coefficient remain substantial at all the sample sizes. This is in line with our conjecture in Section 2 that the use of $\bar{y}_t$ as factor proxies will suffer from the remaining endogeneity due to the correlation between $\bar{y}_t$ and ε_{it}, so far as the spatial weighing matrix remains dense even for large N.

RMSEs of the CCEX-IV individual estimator of $(\rho_i, \beta_{1i}, \beta_{2i})$ decrease with T, verifying Theorem 1 that the convergence rate of the individual estimator is $\sqrt{T}$. RMSEs for (β_{1i}, β_{2i}) do not depend on N, though RMSEs for ρ_i decline with N if h is fixed and rise with N if h is increasing with N. Similar patterns are observed for the CCE-IV estimator. Interestingly, we observe that RMSEs for ρ_i increase with h but fall with ρ, whereas RMSEs for (β_{1i}, β_{2i}) show little variations. Overall, we find that RMSEs for the CCE-IV estimator is no smaller than the CCEX-IV counterparts, suggesting that the potential efficiency gain from using $\bar{y}_t$ as a factor proxy does not offset the higher biases. In particular, if the spatial weighing matrix remains dense with $h = 0.3N$, then RMSEs for the spatial coefficient are significantly higher for the CCE-IV estimator than the CCEX-IV counterparts for all the sample sizes.

[Table 2 about here]

In Table 3 we report the biases and RMSEs for the Mean Group estimator. Overall, we obtain qualitatively similar results to those reported for the individual estimators. The biases of the CCEX-IV Mean Group estimator for (ρ, β_1, β_2) are more or less negligible and close to those of the infeasible estimator in almost all cases. On the other hand, the CCE-IV estimator tends to display non-negligible biases as h or ρ starts to rise, especially for $N = 20$. In particular, if the spatial weighing matrix remains dense with $h = 0.3N$, then the bias of the Mean Group estimator for ρ remains relatively large even when $N = 100$.

RMSEs of both CCEX-IV and CCE-IV Mean Group estimators of (ρ, β_1, β_2) decrease sharply with N, verifying Theorem 2 that the convergence rate of the Mean Group estimator is $\sqrt{N}$, though its speed is rather slower for the Mean Group estimation of ρ in the case with $h = 0.3N$. Overall, we observe that RMSEs of CCEX-IV and CCE-IV Mean Group estimators display similar patterns and magnitudes in most cases. This again suggests that the potential efficiency gain from using $\bar{y}_t$ as a factor proxy does not necessarily offset the higher bias.

[Table 3 about here]

The size and power of the t-tests for the individual and Mean Group estimators are summarised in Tables 4 and 5. As sample size increases (particularly T), the size of the t-tests for the CCEX-IV individual estimator (and the infeasible estimator) tends to the 5% nominal level in most cases. But, we observe that the t-tests for the CCE-IV individual spatial coefficient tend to be slightly over-sized for small h but under-sized for large h if N is small, even for large T. The powers of the t-tests for CCEX-IV and CCE-IV estimators are more or less similar, and both tend to 1 as the sample size (in particular, T) increases if h is fixed. In the case with $h = 0.3N$, the power falls with N due to RMSEs increasing with N, but rises sharply with T. Moreover, the power of the t-tests for the spatial coefficient decreases with h.

Turning to the Mean Group estimator, we find that the size and power of the t-tests display qualitatively similar patterns to those reported for the individual estimator. Now, the size distortion for the CCE-IV estimator of ρ is more noticeable, especially for small N. The power of the t-tests for all the parameters approaches 1 quickly as sample size (in particular, N) rises.

[Tables 4 and 5 about here]

In sum, we establish that the finite sample performance of the CCEX-IV estimator is quite satisfactory in almost all cases considered. On the other hand, the performance of the CCE-IV estimator depends crucially upon the degrees of sparsity of the spatial weighting matrix. If the network is relatively sparse, then its performance is satisfactory. However, if the network is relatively dense, the biases of the CCE-IV estimated spatial coefficient remain substantial at all the sample sizes. In this case the RMSEs are also significantly higher than the CCEX-IV counterparts even for large N. This is in line with our conjecture that the use of $\bar{y}_t$ as factor proxies may suffer from the remaining endogeneity due to the correlation between $\bar{y}_t$ and idiosyncratic errors, which leads to bias that becomes larger when the network becomes denser.

We also report the complete simulation results in the Online Supplement, which provide the support for the robust performance of the CCEX-IV estimator under the different experiments. Furthermore, we establish that the Pooled estimators exhibit substantial biases under parameter heterogeneity (see Sections S4.2 and S4.4 in the Online Supplement).

5 Empirical Application

The study of comovements in house prices in the UK has a long history (Giussani and Hadjimatheou (1991)). Recently, spatial techniques have become more popular in studying comovement of regional house prices and the ripple effects in the UK (Holly et al. (2011); Gray (2012); Chelva and Shin (2017)). Furthermore, Holly et al. (2011) argue that the comovement of house prices could also be originated from (unobserved) economy-wide common shocks (e.g. the introduction of *Help to Buy* or *Buy to Let* policies in UK) that affect all the regions. Another important issue is the parameter heterogeneity. This will be of great relevance when the analysis is based on the regional level instead of the national level, because of different economic conditions and house market structures across regions (Meen (1999)). Many empirical studies have also confirmed the parameter heterogeneity for regional studies in the UK, (Holly et al. (2011); Meen (2001); Chelva and Shin (2017)) and for other countries (Aquaro et al. (2021)).

To analyse the spatial patterns of the UK house price growths, we apply our proposed model, that can address the spatial dependence, unobserved common shocks and parameter heterogeneity simultaneously, to the quarterly changes of real house price for 339 Local Authority Districts (LAD) in the UK over the period 1997Q1-2016Q4 ($T = 80$).[11] Following the literature, we focus on the demand side[12] by analysing the two most popular determinants of house price growths: personal income and population growths.[13] See Section S5.1 in the Online Supplement for the data details.

As a preliminary examination of the data, we apply the cross-section dependence (CD) test developed by Pesaran (2015) to house price growth. We find the CD statistic is 781.9, well above the critical value of 1.96 at the 5% level, and strongly reject the null hypothesis of weak CSD. We also report the estimate of CSD exponent (denoted α hereafter) proposed by Bailey et al. (2016b), which is quite close to 1 (0.99). Both results indicate that there exists strong CSD in real house price growths. Hence, we employ the heterogeneous spatial data model with common factors:

$$hp_{it} = \rho_i hp^*_{it} + \beta_i^{pop} pop_{it} + \beta_i^{inc} inc_{it} + e_{it}, \qquad (33)$$
$$e_{it} = \boldsymbol{\gamma}'_{yi} \boldsymbol{f}_{yt} + \varepsilon_{it},\ i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T,$$

where hp_{it} is the real house price growth, pop_{it} the population growth and inc_{it} the per capita personal income growth for LAD i at time t. We include the spatial term, $hp^*_{it} = \sum_{j=1}^{N} w_{ij} hp_{it}$ with w_{ij} being the (i, j)-th element of the spatial weighting matrix. As pointed out by Meen (2001), the spatial weighting matrix should specify the relationship between house prices in different locations, using the spatial contiguity or distance decay (see Section S5.2 in the Online Supplement for the construction of the spatial weighting matrices). Here we focus on the distance-based spatial weighting matrix.[14]

To control for spatial endogeneity and unobserved factors, we employ the CCEX-IV estimator using $\bar{\boldsymbol{x}}_t$ only and the CCE-IV estimator using $(\bar{y}_t, \bar{\boldsymbol{x}}'_t)'$ as factor proxies. In both cases we use the combinations of $\tilde{\boldsymbol{X}}^*$ and $\tilde{\boldsymbol{X}}^{2*}$ as IVs for the spatial lagged term.

5.1 Estimation Results for Individual LADs

First, we analyse the CCEX-IV estimation results for each LAD. To provide an overall visualisation of the heterogeneous effects of the spatial term and explanatory variables on the real house price growths, we display spatial maps of the individual coefficients in Figure 1. We match the individual spatial coefficient, $\hat{\rho}_i$, to the corresponding LAD on the choropleth map in Figure 1a. LADs colored in green depict positive coefficients: darker shades are associated with estimates closer to unity while

[11] LADs are local governments that are responsible for providing full or partial services to the public. In 2019, there are 382 LADs in the UK (317 in England, 22 in Wales, 11 in Northern Ireland and 32 in Scotland). Due to the data availability, we only analyse 339 LADs in England and Wales.

[12] Housing economics theory suggests that any demand shock will lead to a temporary change in real house prices while housing supply is inelastic in the short-run (Meen (2001)). Extensions to including dynamics and characterising house price growths from the supply side deserve a separate future research.

[13] An extensions of our study to include more determinants, such as interest rate and unemployment rate, is straightforward. But, their effects on house prices are negligibly small and tend to appear with the wrong sign, see Case and Shiller (2003); McQuinn and O'Reilly (2008); Chelva and Shin (2017).

[14] When employing the contiguity-based spatial weighting matrix, the estimation results are less satisfactory. In particular, the spatial coefficients fall outside (-1,1) for 92 LADS (as compared to 51 in the model with the distance-based one). These results at the regional and national level are presented in Table S35 in the Online Supplement.

the lighter shades refer to $\hat{\rho}_i$ closer to zero. Similarly, blue areas are associated with negative spatial coefficients, with the lighter shade indicating $\hat{\rho}_i$ closer to zero while darker blue areas representing more sizable coefficients. We find that 51 of 339 individual spatial coefficients fall outside the interval (-1,1), which are regarded as outliers, plotted in the category 'Outlier'.

[Figure 1 about here]

The individual spatial coefficients are quite heterogeneous, but overall positive (269, 93.4%) and significant (174, 60.4%). This finding is in line with *a priori* expectation that the house price increase in nearby area causes the local house demand to rise as they share amenities, but it also causes the local house supply to fall as the suppliers will switch to an area with high house price growth. Only 19 LADs exhibit a negative spatial coefficient but all insignificant. Negative spatial coefficients can be observed under very specific cases in which the less developed districts, where the house is of very poor quality in terms of economic and social values, are surrounded by more developed areas.[15]

Turning to the CCE-IV results, we find that their individual spatial coefficients are much smaller than CCEX-IV counterparts. There are much more negative spatial coefficients (91, 31.6%) (see Figure 1b),[16] and many of them are insignificant. This rather unsatisfactory outcome may reflect the combination of two perspectives. First, we observe that the correlation between hp_t^* and $\bar{hp}_t$ tends to be higher than 0.5 for 275 out of 339 LADs. Such a high proportion of positive correlations may cause multicollinearity in the CCE-IV estimation, leading to unstable and insignificant estimation results. Next, we observe that the distance-based weighting matrix is relatively dense. Then, the smaller CCE-IV spatial coefficients may indicate downward biases of the CCE-IV estimator, which could occur if the network is relatively dense, as shown in Monte Carlo experiments in Section 4.

Next, we analyse the impacts of population growth, obtained by CCEX-IV estimation and plotted in Figure 1c, which are quite heterogeneous. The population growth exerts a positive effect on the real house price changes for 172 LADs (50 significant) whilst its impact is negative for 116 LADs (34 significant). We expect a positive relationship between population growth and house price growth since a natural population growth will increase the demand for house (Day (2018)). However, if population growth is mainly due to migration flows, its effects on house price would be ambiguous. As discussed by Sá (2015), immigration may be associated with negative house price changes through the following channels: (i) immigration accompanied by native out-migration from

[15] This is the case for Ribble Valley ($\hat{\rho}_i = -0.25$), a mainly rural district surrounded by the Greater Manchester area, for South Bucks ($\hat{\rho}_i = -0.21$) and Slough ($\hat{\rho}_i = -0.07$), largely rural districts surrounded by the Great London area, and for Monmouthshire ($\hat{\rho}_i = -0.11$), a mainly rural district in the south coast of Wales, close to the two cities, Cardiff and Newport. See the Rural-Urban Classification defined in Section S5.3 in the Online Supplement.

[16] Applying the CCE-based de-factoring, Bailey et al. (2016a) and Aquaro et al. (2021) have reported a substantially high incidence of negative spatial connections in the study of the US house price. In the context of the housing market, this outcome is puzzling given the lack of convincing explanations. Chelva (2017) has also obtained an unexpectedly high proportion of negative spatial coefficients when the similar method is applied to English house price inflation. As described earlier, some areas may be inversely related to neighbourhood house price changes due to substantial dissimilar values of the house. Following Chelva and Shin (2017), we hypothesise these cases would be in the minority.

the richer areas, has adverse effects on the average local income; (ii) immigration may generate more crime or affect the quality of local public goods due to overcrowding. In Figure 1c we find that negative estimates of β_i^{pop} appear mostly in London, the Great Manchester and their neighbourhood areas, which have higher immigration rates.[17]

Finally, we analyse the CCEX-IV coefficients, $\hat{\beta}_i^{inc}$, plotted in Figure 1e, which are also quite heterogeneous. Although many (95) are negative, they are small and only 13 of them are significant. We expect that the effects of income growth on house price are generally positive, since the higher income makes the house purchase more affordable. However, there are two situations under which income growth could be associated with negative house price growths. The first is when income growth involves redistribution of income from the poor to the rich such that the positive impact by the rich may be overtaken by the negative impact from decreased income in poor people. The overall house price may go down, see Määttänen and Terviö (2014). The second is the rural-urban migrants decision. Zang et al. (2015) find that the income growth in rural areas encourages local people to migrate to urban areas for better job and education, etc. This will in turn lower the demand for house in rural areas because of out-migration.[18] Since the measurement of income inequality data (e.g. the Gini coefficient) is unavailable at LAD level, we use the income deprivation index, which measures the proportion of the population experiencing deprivation due to low income. For the 13 LADs with significantly negative $\hat{\beta}_i^{inc}$, 6 are located in the high income deprived LADs: Thanet (-5.16), Enfield (-3.38), Nuneaton and Bedworth (-2.57), Wolverhampton (-2.29), Bolton (-1.91) and Kingston upon Hull (-1.64), while 7 are located in rural or suburban areas: Surrey Heath (-2.77), East Hertfordshire (-2.49), Fareham (-2.01), Basingstoke and Deane (-1.38), South Oxfordshire (-1.27), Hambleton (-0.90) and Richmondshire (-0.86).

Turning to the CCE-IV results, we find that the spatial patterns of $\hat{\beta}_i^{pop}$ and $\hat{\beta}_i^{inc}$, are not as clear as CCEX-IV counterparts. There are several cases which are difficult to interpret, for example, negative $\hat{\beta}_i^{pop}$ in several rural LADs in South West and negative $\hat{\beta}_i^{inc}$ in mainly urban LADs, Somerset West & Taunton (-1.44) and Trafford (-1.32).

5.2 Estimation Results at the Regional and National Level

England is divided into 9 regions: North East, The Yorkshire and Humber, North West, East Midlands, West Midlands, East of England, London, South East and South West. To ensure that each region contain a reasonable number of LADs, we merge North East and The Yorkshire and Humber (denoted North East & York). Finally, together with Wales, we have 9 regions.

In Table 6 we report the Mean Group estimation results at the regional and national level.[19]

---

[17]Inspecting further, negative estimates are observed for LADs with high immigration rate, such as Brent (-1.25), Ealing (-1.06) and Westerminister (-0.35), Reading (-0.67), Oxford (-0.48) and Brighton and Hove (-5.3), South Cambridgeshire (-4.08) and Norwich (-3.14).

[18]These two reasons also turn out to explain the negative estimates for β_i^{inc} in our analysis.

[19]In computing the Mean Group estimates, we exclude 51 and 49 LADs with their spatial coefficients outside (-1,1) for CCEX-IV and CCE-IV. In addition, we exclude 6 and 9 more LADs that exhibit extremely large, $\hat{\beta}_i^{pop}$ and $\hat{\beta}_i^{inc}$, greater than 15. Hence, we exclude 57 and 58 LADs respectively for CCEX-IV and CCE-IV estimation.

Overall the regional Mean Group estimates are heterogeneous, but we are now able to provide more vivid spatial patterns across regions. We first analyse the Mean Group estimation results for CCEX-IV. The spatial coefficients are all positive and significant, mitigating the small number of negative and insignificant individual coefficients. They are tightly clustered around the national mean of 0.57, suggesting the existence of positive spatial linkage of house prices growths across regions. Notice that the spatial coefficient for London is the lowest at 0.40. The lower spatial coefficient implies that the house price growth is more likely to be affected by fundamental determinants rather than the house price growth in neighbouring regions. For the regions sharing a border with London (East of England and South East), they are not only influenced by London, but also affecting neighbour regions, making their spatial coefficients relatively lower than those in other regions. This provides the support for the central role played by London.

On the other hand, the spatial coefficients by the CCE-IV estimation are negligibly small and insignificant for some regions with the national mean of 0.23, which is not in line with the widespread evidence of the spatial dependence and the ripple effects in the UK housing market documented in the literature (e.g Meen (1999) and Holly et al. (2011)). Further, some estimation results, for example, that London has the fourth highest spatial coefficient among all the regions, are hard to explain. These unsatisfactory results could be due to the small sample bias or the downward bias of the CCE-IV estimator under the distance-based weighting matrix that is relatively dense, as shown in Monte Carlo experiments in Section 4.

Next, the coefficients on population and income growths display interesting spatial patterns. $\hat{\beta}^{pop}_{MG,r}$ is larger and significant for North East & York, North West and South West, whilst $\hat{\beta}^{inc}_{MG,r}$ is larger and significant in the rest of regions. Population growth will increase house price when it contributes the new formation of house (Day (2018)). From Table 7, we observe that the household growths for East Midlands, London, East of England and South East are lower than in other regions. This implies that population growth does not necessarily induce the demand for new house. Indeed, population growths in rich regions are largely caused by immigrants. Then, they are likely to live in the multi-family and two or more unrelated adults household types.[20] Thus, house price growths in these regions are relatively less affected by population growths. By contrast, population growths in less rich regions transform into the demand for new house as observed in their relatively high household growth (Table 7).

Our finding that house prices are more responsive to income changes in South-East regions is in line with the regional "ripple effects" in the UK house market, see Meen (1999). The significant impacts of income growth on house price growth provide support for the leading position of South-East regions (in particular, London). Overall, the CCEX-IV results suggest that personal income growth is the dominant factor behind the house price growth in rich regions while the population growth would be the more dominant factor for less rich regions.

There are two regions, South West and Wales, which may require further explanations. South West has the fourth highest average income with the lowest population density, making it largely

[20] These two types of households account for almost 10% of the total households in London.

a rural region due to its geographical and climate features. Income growth in rural area could be associated with the lower local demand because of out-migration. We find that 10 out of 19 LADs display negative impacts of income growths, which are all rural or suburban LADs. Thus, the house price growth in South West, which has the third highest household growth, is mainly driven by population growth. Next, the impact of population growth in Wales is higher than the income effect. A high household growth in Wales suggests that the population growth has induced more demand for house. The personal incomes in Wales and North East are very similar, but the average house price in Wales is 8.7% higher, rendering the house in Wales less affordable. This also makes the income growth a significant determinant.

By contrast, it is rather difficult to interpret the spatial patterns of the results provided by CCE-IV. For example, we observe that the population growth is a more dominant driver than the income growth behind the house price growth in London, which deems unreliable, given its highest population density and lowest household growth.

Turning to the CCEX-IV results at the national level, we find that all the coefficients are statistically significant at 1% level while the impacts of population growths are substantially larger than income growths, in line with the US studies reported by Aquaro et al. (2021) and Yang (2021).

[Tables 6 and 7 about here]

Finally, we notice that the CD test rejects the null of weak CSD for some regions and at the national level while the CSD exponent, $\hat{\alpha}$ becomes moderate. This suggests that house price growth may be affected by other factors that are not associated with population and income growth. However, our proposed CCEX-IV method can still provide consistent estimates if strong factors from the regressors can be removed effectively.[21]

5.3 The GCM Analysis of Direct, Spill-in and Spill-out Effects

In a model with the spatial dependence, it is now standard to provide a summary measure of the direct and indirect effects of the regressors on the dependent variable using the following system representation of (33):

$$hp_{.t} = \boldsymbol{\rho} \boldsymbol{W} hp_{.t} + \boldsymbol{\beta}^{pop} pop_{.t} + \boldsymbol{\beta}^{inc} inc_{.t} + e_{.t}, \tag{34}$$

where $\boldsymbol{\rho}$ is a $N \times N$ diagonal matrix with i th diagonal element, ρ_i, and $\boldsymbol{\beta}^{pop}$ and $\boldsymbol{\beta}^{inc}$ are defined similarly with ith diagonal elements, β_i^{pop} and β_i^{inc}. We construct heterogeneous direct, spill-in and spill-out effects for the ith LAD (see LeSage and Chih (2016)):

[21] We find that the CD test results are 182.5 and 1613.8 for the raw data, pop_{it} and inc_{it} with $\hat{\alpha}$ being 0.86 and 1, respectively. When applying to the defactored data obtained using $(\overline{pop}_t, \overline{inc}_t)$ as factor proxies, we find that the CD test results reduce to 5.18 and -2.91 while $\hat{\alpha}$ become 0.73 and 0.74. This shows that unobserved factors have been effectively removed by CCEX.

- Heterogeneous Direct Effect (HDE): the direct effect of the population or income growth on the real house price changes, given by the ith diagonal element of $(\boldsymbol{I}_N - \boldsymbol{\rho}\boldsymbol{W})^{-1}\boldsymbol{\beta}^{pop}$ or $(\boldsymbol{I}_N - \boldsymbol{\rho}\boldsymbol{W})^{-1}\boldsymbol{\beta}^{inc}$.
- Heterogeneous Spill-in Effect (HSI): the sum of the effects of population or income growths from all the other LADs on house price changes in the ith LAD, given by ith row-sum minus ith diagonal element of $(\boldsymbol{I}_N - \boldsymbol{\rho}\boldsymbol{W})^{-1}\boldsymbol{\beta}^{pop}$ or $(\boldsymbol{I}_N - \boldsymbol{\rho}\boldsymbol{W})^{-1}\boldsymbol{\beta}^{inc}$.
- Heterogeneous Spill-out Effect (HSO): the sum of the effects of population or income growth from the ith LAD on house price changes in all the other LADs, given by the ith column-sum minus ith diagonal element of $(\boldsymbol{I}_N - \boldsymbol{\rho}\boldsymbol{W})^{-1}\boldsymbol{\beta}^{pop}$ or $(\boldsymbol{I}_N - \boldsymbol{\rho}\boldsymbol{W})^{-1}\boldsymbol{\beta}^{inc}$.

Our approach may operate at two extremes: (i) complete aggregation, where the $N(N-1)$ bilateral linkages among N individual LADs are aggregated into the single indices at the national level, and (ii) no aggregation, where the $N(N-1)$ bilateral linkages are studied at the individual LAD level. On the one hand, a single index to summarise the connectedness of the system will obscure the amount of details of regions completely. On the other hand, for large N, it would become almost infeasible to study the network topology on a disaggregated basis. In this regard, we follow the GCM approach by Greenwood-Nimmo et al. (2021) and introduce intermediate levels of aggregation by analysing the $R(R-1)$ bilateral linkages among R regions rather than among N individual LADs.[22] We focus on the CCEX-IV estimation results with $N = 282$ but $R = 9$. The number of LADs in the ℓth region is denote by N_ℓ for $\ell = 1, \ldots, R$.

We re-order the individual LADs so that they are gathered together into desired regions. We write the $N \times N$ matrix, $(\boldsymbol{I}_N - \boldsymbol{\rho}\boldsymbol{W})^{-1}\boldsymbol{\beta}^{pop}$ or $(\boldsymbol{I}_N - \boldsymbol{\rho}\boldsymbol{W})^{-1}\boldsymbol{\beta}^{inc}$ as

$$\underset{(N\times N)}{\mathbf{C}} = \begin{bmatrix} \phi_{1\leftarrow 1} & \cdots & \phi_{1\leftarrow N_1} & \phi_{1\leftarrow N_1+1} & \cdots & \phi_{1\leftarrow N_1+N_2} & \cdots & \phi_{1\leftarrow N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{N_1\leftarrow 1} & \cdots & \phi_{N_1\leftarrow N_1} & \phi_{N_1\leftarrow N_1+1} & \cdots & \phi_{N_1\leftarrow N_1+N_2} & \cdots & \phi_{N_1\leftarrow N} \\ \phi_{N_1+1\leftarrow 1} & \cdots & \phi_{N_1+1\leftarrow N_1} & \phi_{N_1+1\leftarrow N_1+1} & \cdots & \phi_{N_1+1\leftarrow N_1+N_2} & \cdots & \phi_{N_1+1\leftarrow N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{N_1+N_2\leftarrow 1} & \cdots & \phi_{N_1+N_2\leftarrow N_1} & \phi_{N_1+N_2\leftarrow N_1+1} & \cdots & \phi_{N_1+N_2\leftarrow N_1+N_2} & \cdots & \phi_{N_1+N_2\leftarrow N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{N\leftarrow 1} & \cdots & \phi_{N\leftarrow N_1} & \phi_{N\leftarrow N_1+1} & \cdots & \phi_{N\leftarrow N_1+N_2} & \cdots & \phi_{N\leftarrow N} \end{bmatrix}. \quad (35)$$

The (k, ℓ)th block in (35), denoted as $\mathbf{B}_{k\leftarrow\ell}$ for $k, \ell = 1, ..., R$, is given by:

$$\underset{(N_k\times N_\ell)}{\mathbf{B}_{k\leftarrow\ell}} = \begin{bmatrix} \phi_{\tilde{N}_k+1\leftarrow\tilde{N}_\ell+1} & \cdots & \phi_{\tilde{N}_k+1\leftarrow\tilde{N}_\ell+N_\ell} \\ \vdots & \ddots & \vdots \\ \phi_{\tilde{N}_k+N_k\leftarrow\tilde{N}_\ell+1} & \cdots & \phi_{\tilde{N}_k+N_k\leftarrow\tilde{N}_\ell+N_\ell} \end{bmatrix}, \quad (36)$$

[22] We provide the results for individual HDE, HSI and HSO in Section S5.4 in the Online Supplement. Their spatial patterns are closely related to those of individual coefficients reported in Figure 1c and 1e. For all LADs, the signs of HDE are the same as the corresponding individual coefficients. Moreover, if population (income) growth in the ith LAD increases its own house price, then this will induce positive spill-outs from the ith LAD as well as positive spill-ins for its neighbourhood LADs, because the house price growths are mostly positively spatially correlated.

where $\widetilde{N}_k = \sum_{j=1}^{k-1} N_j$ for $k = 2, \ldots, R$, and $\widetilde{N}_1 = 0$. We evaluate the sum of the elements of $\mathbf{B}_{k\leftarrow\ell}$ and normalise it by the average number of LADs in the pair as:

$$\psi_{k\leftarrow\ell} = \frac{1}{0.5(N_k + N_\ell)} \boldsymbol{\iota}'_{N_k} \mathbf{B}_{k\leftarrow\ell} \boldsymbol{\iota}_{N_\ell}, \tag{37}$$

where $\boldsymbol{\iota}_{N_k}$ is an $N_k \times 1$ column vector of ones. Then, we can construct the following $R \times R$ connectedness matrix at the regional level:

$$\underset{(R\times R)}{\mathbf{C}_R} = \begin{bmatrix} \psi_{1\leftarrow1} & \psi_{1\leftarrow2} & \cdots & \psi_{1\leftarrow R} \\ \psi_{2\leftarrow1} & \psi_{2\leftarrow2} & \cdots & \psi_{2\leftarrow R} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{R\leftarrow1} & \psi_{R\leftarrow2} & \cdots & \psi_{R\leftarrow R} \end{bmatrix}. \tag{38}$$

Now, it is straightforward to derive the direct, spill-in and spill-out effects of population and income growths on the house price growths at the regional level using (38), denoted RDE, RSI and RSO, respectively, defined by

$$RDE_i = \psi_{i\leftarrow i}; \quad RSI_i = \sum_{j=1, j\neq i}^{R} \psi_{i\leftarrow j}; \quad RSO_i = \sum_{j=1, j\neq i}^{R} \psi_{j\leftarrow i}.$$

We construct the regional net effect (RNE) by the difference between RSO and RSI, which may allow us to distinguish between net-transmitting and net-receiving regions with respect to population or income growth shocks (e.g. Antonakakis et al. (2018)).

The regional direct/spillover results are summarised in Table 8. We first analyse the bilateral impacts of population growth. The spatial pattern of RDE of population growth is qualitatively similar to that of regional Mean Group coefficients, $\beta^{pop}_{MG,r}$ as reported in Table 6. RDE is substantially larger for North East & York, South West and Wales. Notice that the RDE of population growth for London is the lowest, reflecting the combined effects of its highest population density and lowest household growth (see Table 7).

Turning to the interplay between RSI and RSO, North East & York exhibits the largest RSI, mainly due to the large spill-ins from North West and East Midlands, whilst having the smallest RSO. Thus, its net effect is most negative. Since the sum of net effects is zero by construction, North East & York can be regarded as the most passive receiver in terms of the population growth impacts.

By contrast, London displays the smallest RSI. This negative spill-in effect from South East and East of England mainly reflects that more people tend to escape from London to neighbouring regions where houses are more affordable but they remain close to commute, see Meen (2001). Hence, population growths in South East and East of England are likely to reflect out-migration from London,[23] thus exerting negative effects on house price growth in London. London shows the

[23] Out-migration from London to South East and Est of England is nearly twice as out-migration from London to

larger RSO. Thus, London becomes the most influential transmitter of population growth shocks.

East Midlands and North West are another influential transmitters of population growth shocks. They exert large positive spill-outs to North East & York. Though East Midlands is also influenced significantly by South East, its net effect is the largest, albeit slightly larger than London, reflecting that it is a region with a very strong manufacturing sector. By contrast, North West receives relatively small spill-ins. South East and South West belong to the class of receivers. But, the former can be regarded as the active receiver as it has the second largest RSI and the fourth largest RSO.

The remaining three regions (West Midlands, East of England and Wales) are relatively neutral with their spill-outs approximately matching their spill-ins. East of England and West Midlands are mildly active in terms of its sizable and contradictory RSI and RSO associated with neighbors. On the other hand, Wales appears to be rather independent of the regions in England.

Next, we analyse the bilateral impacts of personal income growth. The spatial pattern of RDE of income growth is somewhat similar to the regional Mean Group coefficients, $\beta^{inc}_{MG,r}$ as reported in Table 6. RDE is substantially larger for East Midland, South East, London and Wales. On the other hand, RDE of income growth for South West is the lowest followed by North West. As discussed above, South West is largely a rural region, and its lowest RDE may be in line with the rural-urban migrants decision.

East Midlands and London become the two most influential transmitters of income growth shocks. London exerts the large positive spill-outs particularly on South East while receiving rather small spill-ins from South East and East of England. East Midlands produces the larger positive spill-outs to North East & York, while receiving relatively small spill-ins from South East and West Midlands.

South East exhibits the largest RSI of income growth mainly due to the substantial spill-ins from London, but it also has the fourth largest RSO. Though it has the largest negative net effects, it can be regarded as the active receiver. North East & York has the second largest RSI (receiving the largest spill-ins from East Midlands) while exerting negligible spill-outs to the other regions. Thus, it is still regarded as the most passive receiver. South West belongs to the class of receiver, mainly due to the spill-ins from South East and Wales.

The remaining four regions are neutral with their net effects being rather small, though East of England is mildly active in terms of its contradictory RSI and RSO. Wales is again independent of the regions in England. Interestingly, North West exerts a slightly negative spill-outs to North East & York, suggesting that its income growth induces the residents in North East & York to make the rural-urban migration decision.

[Table 8 about here]

Finally, we define a pair of indices to address two issues of particular interest (i) 'how dependent is the ith region on external conditions from other regions'; and, (ii) 'to what extent does the ith

all other regions during 2002-2012, see a report by Mayor of London available at https://www.london.gov.uk/sites/default/files/wp62-migration-commuting-final.pdf.

region influence/is the ith region influenced by the system as a whole?' We follow Shin and Thornton (2020) and construct the External Motivation (EM) and Systemic Influence (SI) indices given by

$$EM_i = \frac{RSI_i}{ATOT_{i\leftarrow\bullet}}; \quad SI_i = \frac{RNE_i}{TNP_i}, \tag{39}$$

where $ATOT_{i\leftarrow\bullet} = \sum_{j=1}^{R} |\psi_{i\leftarrow j}|$ is the absolute row-sum for region i, and $TNP_i = 0.5\sum_{i=1}^{N} |NE_i|$ is the total absolute net effects. EM_i measures the relative importance and direction of RSI in determining the conditions in the ith region while SI_i captures the systemic influence of the ith region.[24]

The coordinate pair (EM_i, SI_i) displayed in Figure 2, will provide a vivid representation of region's relative position in the UK house market. There is the tendency for regions to cluster along a line north-west to south-east, since positive (negative) spill-ins contribute negatively (positively) to a region's net effect. A region in the north-east quadrant would be one for which spill-ins were outweighed by spill-outs, leading to a positive net connectedness. In our empirical application, this corresponds to East Midlands and North West in terms of population growth shocks. A region in the south-west quadrant would be one for which spill-ins outweigh spill-outs, leading to a negative net effect. This corresponds to South East in terms of income growth shocks.

[Figure 2 about here]

From Figure 2a, we find that only London has negative EM with respect to population growth shocks while London has positive spill-outs to East of England and South East. As a result, London is located in the far north-west as the most influential net transmitter of population growth shocks in the UK. On the other hand, North East & York is positioned at the far south-east, clearly showing that it is the most passive receiver. The position of Wales is close to the coordinate pair (0,0), showing that its housing market condition may be independent of those in England.

From Figure 2b, we find that the external motivation is universally positive for all the regions. London and East Midlands are located in the far north-west as the most influential net transmitters of income growth shocks whereas North East & York is positioned at the far south-east as the most passive receiver. Wales still maintains a rather independent position with its coordinate pair close to (0,0). Interestingly, North West is now located at the south-west quadrant with slightly negative net effects.

In sum, we have identified London and East Midlands as the most influential transmitters of population and income growth shocks affecting the house price changes in the UK. However, our findings do not provide full support for the London-centric view of ripple effects where house price appreciation begins in South East and London before spreading to the rest of the country. The

[24] Both EM_i and SI_i stay within $[-1, 1]$. If $EM_i \to 1(-1)$, then the house market condition in region i is dominated by positive (negative) RSIs, as opposed to direct effects. If region i receives contradictory spill-ins and/or if the magnitude of RSI is small in comparison to direct effects, then $EM_i \to 0$. If $0 \leq SI_i \leq 1$ $(-1 \leq SI_i \leq 0)$, then region i is a net shock transmitter (receiver). If SI_i is close to zero, then region i is neutral with its RSOs matching its RSIs.

ripple effects originated from London may have eventually spread to North East & York and South West via spatial interactions with East Midlands and South East. But, we fail to detect evidence that this effect does spread to North West and Wales. Both regions seem to have their own house market drivers, because North West has the second largest urban area in the UK while Wales is a country with its own economic development strategy, see also Chelva and Shin (2017).

6 Concluding Remarks

In this paper, we have developed a unifying econometric framework for the analysis of heterogeneous panel data models that can account for spatial dependence and common factors, simultaneously. In particular, we develop a CCEX-IV estimation procedure, that approximates unobserved factors by CSA of the regressors, and then deals with the spatial endogeneity through employing the (internal) instrumental variables.

We show that the individual parameters can be estimated consistently by applying the de-factored IVs directly to the original regressions and the resulting estimators are asymptotically normal-distributed. We also establish $\sqrt{N}$ -consistency and asymptotic normality for the Mean Group estimator. By contrast, we find that the Pooled estimator is inconsistent in the presence of parameter heterogeneity. This is a new finding, extending the work by Pesaran and Smith (1995) who establish that the Pooled estimator is inconsistent in heterogeneous dynamic panels.

Monte Carlo simulations confirm that the finite sample performance of the proposed estimators is quite satisfactory. We also demonstrate the usefulness of our approach with an application to the house price growth for Local Authority Districts in the UK over 1997Q1-2016Q4.

We conclude by noting a few avenues for future research. A natural extension is to develop a dynamic heterogeneous spatial panel data model with interactive effects, which can shed further lights on improving our understanding of the dynamic network, e.g., Shin and Thornton (2020). Another extension is the development of nonlinear/quantile heterogeneous panel data models with spatial dependence and common factors. In particular, the CCEX estimator can be easily extended to developing the binary choice model with unobserved factors, e.g., Boneva and Linton (2017). Finally, as discussed in Remark 3, we aim to develop the GMM estimation procedure using both linear and quadratic moment conditions.

References

Antonakakis, N., Chatziantoniou, I., Floros, C., and Gabauer, D. (2018). The dynamic connectedness of UK regional property returns. *Urban Studies*, 55:3110–3134.

Aquaro, M., Bailey, N., and Pesaran, M. H. (2021). Estimation and inference for spatial models with heterogeneous coefficients: an application to US house prices. *Journal of Applied Econometrics*, 36:18–44.

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

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

Bai, J. and Li, K. (2014). Spatial panel data models with common shocks. *mimeo.,* University of Columbia.

Bai, J. and Li, K. (2021). Dynamic spatial panel data models with common shocks. *Journal of Econometrics,* Forthcoming.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146:304–317.

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

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

Boneva, L. and Linton, O. (2017). A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance. *Journal of Applied Econometrics*, 32:1226–1243.

Case, K. E. and Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity*, 2003:299–362.

Chelva, B. (2017). Estimating spillover effects in the English housing market: A heterogeneous spatial autoregressive panel approach. *mimeo.,* University of Leeds.

Chelva, B. and Shin, Y. (2017). Spatial dependence and common factors in the English housing market: A spatio-temporal autoregressive model with factors. *mimeo.,* University of Leeds.

Chen, M. and Yan, J. (2019). Unbiased CCE estimator for interactive fixed effects panels. *Economics Letters*, 175:1–4.

Cui, G., Hayakawa, K., Nagata, S., and Yamagata, T. (2019). A robust approach to heteroskedasticity, error serial correlation and slope heterogeneity for large linear panel data models with interactive effects. *mimeo.,* Osaka University.

Day, C. (2018). Population and house prices in the United Kingdom. *Scottish Journal of Political Economy*, 65:127–141.

Giussani, B. and Hadjimatheou, G. (1991). Modeling regional house prices in the United Kingdom. *Papers in Regional Science*, 70:201–219.

Gray, D. (2012). District house price movements in England and Wales 1997–2007: An exploratory spatial data analysis approach. *Urban Studies*, 49:1411–1434.

Greene, W. H. (2010). *Econometric Analysis.* Pearson Education, Upper Saddle River, NJ.

Greenwood-Nimmo, M., Nguyen, V. H., and Shin, Y. (2021). Measuring the connectedness of the global economy. *International Journal of Forecasting*, 37:899–919.

Holly, S., Pesaran, M. H., and Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics*, 69:2–23.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17:99–121.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533.

Kuersteiner, G. M. and Prucha, I. R. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, 88:2109–2146.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72:1899–1925.

Lee, L.-F. and Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, 180:174–197.

LeSage, J. P. and Chih, Y.-Y. (2016). Interpreting heterogeneous coefficient spatial autoregressive panel models. *Economics Letters*, 142:1–5.

LeSage, J. P. and Chih, Y.-Y. (2018). A bayesian spatial panel model with heterogeneous coefficients. *Regional Science and Urban Economics*, 72:58–73.

Lu, L. (2017). Simultaneous spatial panel data models with common shocks. *mimeo.*, Federal Reserve Bank of Boston.

Määttänen, N. and Terviö, M. (2014). Income distribution and housing prices: An assignment model approach. *Journal of Economic Theory*, 151:381–410.

Mastromarco, C., Serlenga, L., and Shin, Y. (2016). Modelling technical efficiency in cross sectionally dependent stochastic frontier panels. *Journal of Applied Econometrics*, 31:281–297.

McQuinn, K. and O'Reilly, G. (2008). Assessing the role of income and interest rates in determining house prices. *Economic Modelling*, 25:377–390.

Meen, G. (1999). Regional house prices and the ripple effect: A new interpretation. *Housing Studies*, 14:733–753.

Meen, G. (2001). *Modelling Spatial Housing Markets: Theory, Analysis, and Policy.* Springer Science & Business Media, LLC.

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

Norkuté, M., Saradis, V., Yamagata, T., and Cui, G. (2021). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. *Journal of Econometrics*, 220:416–446.

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

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

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

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

Sá, F. (2015). Immigration and house prices in the UK. *The Economic Journal*, 125:1393–1424.

Shi, W. and Lee, L.-F. (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 197:323–347.

Shin, Y. and Thornton, M. (2020). Dynamic network analysis via diffusion multipliers. *mimeo.,* University of York.

Yang, C. F. (2021). Common factors and spatial dependence: An application to US house prices. *Econometric Reviews*, 40:14–50.

Zang, B., Lv, P., and Warren, C. M. (2015). Housing prices, rural–urban migrants' settlement decisions and their regional differences in China. *Habitat International*, 50:149–159.

Zhang, J. and Boos, D. D. (1994). Adjusted power estimates in Monte Carlo experiments. *Communications in Statistics-Simulation and Computation*, 23:165–173.

Table 2: The Finite Sample Performance of Individual Estimators under Experiment 4

		Bias (× 100)									RMSE (× 100)								
		20			50			100			20			50			100		
	(N,T)	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}
		$\rho = 0.5,\ h = 2$																	
	20	0.01	0.05	-0.10	-0.02	0.02	0.00	0.01	-0.08	-0.04	9.99	14.67	14.81	7.60	10.99	10.76	4.97	7.44	7.68
	50	-0.03	-0.04	-0.02	-0.04	0.02	0.05	-0.02	0.04	0.04	9.00	14.66	14.47	7.06	10.59	10.71	4.53	7.23	7.48
	100	-0.02	0.03	0.03	0.02	-0.05	-0.01	0.00	0.05	-0.07	8.86	14.47	14.47	6.59	10.75	10.77	4.54	7.19	7.32
		$\rho = 0.5,\ h = 6$																	
	20	-0.05	0.02	-0.06	0.06	0.07	-0.02	0.04	-0.01	0.03	20.34	14.56	14.62	16.57	10.77	10.95	11.55	7.52	7.44
	50	0.05	-0.02	0.04	-0.01	-0.08	0.03	-0.03	0.01	0.00	15.74	14.37	14.37	11.74	10.58	10.61	7.95	7.33	7.30
	100	-0.02	0.01	-0.02	-0.08	-0.02	0.04	-0.02	0.00	-0.02	14.44	14.25	14.24	10.66	10.48	10.53	7.46	7.29	7.15
CCEX-IV		$\rho = 0.8,\ h = 6$																	
	20	0.04	0.04	0.09	0.09	0.00	-0.01	0.07	-0.02	0.06	16.76	14.85	14.87	11.40	11.04	11.07	6.46	7.74	7.64
	50	0.03	0.01	-0.11	0.00	-0.08	0.04	0.00	-0.01	0.04	8.50	14.53	14.66	6.31	10.75	10.83	4.53	7.28	7.29
	100	-0.04	0.02	-0.05	-0.01	-0.04	-0.05	-0.02	0.04	0.00	8.21	14.24	14.60	6.03	10.55	10.56	3.99	7.10	7.25
		$\rho = 0.8,\ h = 0.3N$																	
	20	0.06	0.04	0.02	-0.06	0.08	-0.05	0.01	-0.03	0.02	17.10	14.90	14.80	11.74	10.95	11.22	7.81	7.76	7.64
	50	0.02	0.05	-0.09	-0.01	-0.03	-0.04	-0.05	0.01	-0.01	21.38	14.43	14.34	14.92	10.55	10.53	10.42	7.19	7.31
	100	0.07	0.01	0.06	-0.02	-0.02	-0.02	0.02	-0.06	-0.05	26.51	14.17	14.32	21.07	10.34	10.39	14.21	7.17	7.14
		$\rho = 0.5,\ h = 2$																	
	20	-0.25	0.01	-0.21	-0.23	-0.08	-0.02	-0.19	-0.13	-0.07	10.05	14.97	14.84	7.75	11.13	10.84	5.00	7.71	7.48
	50	-0.06	-0.06	-0.06	-0.08	0.03	0.03	-0.01	0.05	0.03	8.95	14.46	14.79	7.19	10.72	10.71	4.72	7.41	7.37
	100	-0.03	0.01	0.02	0.02	-0.07	-0.01	-0.01	0.02	-0.03	8.89	14.23	14.69	6.71	10.65	10.68	4.60	7.20	7.32
		$\rho = 0.5,\ h = 6$																	
	20	-1.94	-0.16	-0.21	-2.06	0.03	-0.23	-1.87	-0.05	-0.11	23.04	14.46	14.67	17.85	10.84	10.80	12.44	7.72	7.42
	50	-0.13	-0.01	0.03	-0.12	-0.05	0.00	-0.14	0.03	-0.01	15.95	14.50	14.53	11.87	10.55	10.70	8.00	7.30	7.19
	100	-0.03	0.01	-0.01	-0.13	-0.01	0.03	-0.04	-0.01	-0.02	14.68	14.21	14.24	10.73	10.46	10.54	7.55	7.17	7.18
CCE-IV		$\rho = 0.8,\ h = 6$																	
	20	-2.89	0.06	-0.11	-3.34	-0.19	-0.27	-3.56	-0.23	-0.28	18.26	14.83	14.88	15.05	11.08	10.93	10.02	7.66	7.43
	50	-0.18	-0.01	-0.10	-0.19	-0.08	0.00	-0.33	-0.03	0.05	9.13	14.77	14.63	6.44	10.67	10.87	4.81	7.27	7.23
	100	-0.05	-0.01	-0.07	-0.04	-0.04	-0.07	-0.08	0.03	0.01	8.22	14.26	14.86	6.11	10.65	10.55	3.87	7.15	7.26
		$\rho = 0.8,\ h = 0.3N$																	
	20	-2.97	-0.08	-0.28	-3.16	-0.04	-0.20	-3.69	-0.23	-0.30	20.24	14.95	14.78	14.22	10.98	10.93	11.30	7.74	7.78
	50	-2.58	0.02	-0.15	-2.52	-0.08	-0.11	-2.64	-0.02	-0.12	24.28	14.42	14.37	18.50	10.55	10.51	13.19	7.29	7.27
	100	-2.34	-0.03	-0.01	-2.38	-0.01	-0.07	-2.39	-0.07	-0.09	31.34	14.19	14.33	24.55	10.45	10.43	17.28	7.22	7.12
		$\rho = 0.5,\ h = 2$																	
	20	0.00	0.02	0.01	-0.03	-0.06	0.04	-0.02	-0.08	-0.06	7.52	13.58	13.84	5.55	10.09	10.23	3.74	6.89	6.96
	50	0.07	0.00	-0.03	-0.03	0.01	0.03	0.01	0.00	0.04	7.66	13.74	13.71	5.94	9.95	10.19	3.79	6.79	6.99
	100	-0.04	0.02	0.01	0.05	-0.05	0.02	-0.02	0.03	-0.01	7.87	13.58	13.75	5.78	9.93	10.11	3.92	6.78	6.95
		$\rho = 0.5,\ h = 6$																	
	20	-0.01	0.06	-0.02	-0.01	-0.03	0.02	0.01	0.01	0.05	10.14	13.47	13.61	7.50	10.00	9.97	5.23	6.88	6.75
	50	0.06	0.01	0.06	-0.03	-0.03	0.07	0.00	0.00	-0.01	10.70	13.56	13.56	7.91	9.92	10.04	5.45	6.83	6.88
	100	0.01	0.05	-0.03	-0.03	0.01	0.03	-0.03	0.00	-0.03	10.87	13.58	13.59	8.02	10.02	10.12	5.72	6.82	6.80
Infeasible		$\rho = 0.8,\ h = 6$																	
	20	0.03	-0.04	0.03	0.03	0.00	-0.03	0.05	0.03	0.01	6.28	13.79	14.07	4.47	10.02	10.31	2.78	6.89	7.04
	50	0.03	-0.03	-0.08	0.00	-0.05	0.01	-0.01	0.03	0.04	4.68	13.71	13.90	3.56	10.07	10.24	2.52	6.80	6.91
	100	-0.03	0.04	-0.01	-0.01	-0.06	-0.05	0.01	0.02	-0.02	5.34	13.64	13.74	4.04	9.99	10.10	2.71	6.79	6.95
		$\rho = 0.8,\ h = 0.3N$																	
	20	0.01	0.00	-0.03	-0.04	0.05	0.00	-0.01	-0.03	0.01	5.81	13.87	14.17	4.49	10.03	10.35	2.90	6.99	6.94
	50	0.02	0.05	-0.06	0.02	-0.05	-0.08	-0.02	0.05	0.02	8.39	13.84	13.82	5.97	10.19	10.16	4.29	6.88	7.02
	100	-0.01	-0.04	0.04	-0.02	0.00	0.01	0.05	-0.03	-0.02	11.92	13.74	13.86	8.36	10.18	10.09	3.67	6.85	6.94

Notes: The simulation results are based on the DGP specified in Section 4.

Table 3: The Finite Sample Performance of Mean Group Estimators under Experiment 4

		Bias (× 100)									RMSE (× 100)								
		20			50			100			20			50			100		
	(N,T)	ρ	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2
										$\rho = 0.5,\ h = 2$									
	20	-0.04	0.00	0.02	-0.05	-0.04	-0.06	0.01	0.06	0.00	4.05	12.27	8.27	3.11	11.60	7.81	2.71	11.54	7.27
	50	0.02	0.07	-0.07	-0.03	0.06	-0.02	-0.03	-0.01	0.01	2.33	7.62	5.52	2.02	7.43	4.87	1.53	7.36	4.36
	100	-0.04	0.01	0.01	0.02	0.08	0.03	0.02	0.01	0.00	1.41	5.30	3.86	1.04	5.22	3.45	0.58	5.11	2.91
										$\rho = 0.5,\ h = 6$									
	20	0.07	0.06	0.02	-0.01	-0.01	-0.02	-0.04	0.05	0.02	8.54	12.01	8.64	6.12	11.81	7.98	4.56	11.79	7.19
	50	0.06	0.01	-0.06	-0.07	0.09	0.07	0.01	0.04	-0.02	3.64	7.72	5.31	2.71	7.23	4.79	1.99	7.14	4.55
	100	-0.08	-0.09	0.03	-0.03	0.08	-0.02	-0.06	-0.08	-0.06	2.40	5.43	3.78	1.82	5.20	3.47	1.26	5.15	2.84
CCEX-IV										$\rho = 0.8,\ h = 6$									
	20	-0.03	-0.01	-0.06	-0.02	0.02	0.08	0.07	0.04	0.04	7.41	12.36	8.41	4.24	11.73	7.95	3.26	11.25	7.08
	50	-0.07	0.03	-0.01	-0.02	0.04	-0.02	0.07	0.00	-0.07	2.44	7.44	5.44	1.81	7.38	4.79	1.53	7.20	4.51
	100	-0.03	0.00	0.02	-0.07	0.05	0.04	-0.01	0.07	0.00	1.33	5.51	3.90	1.07	5.31	3.32	0.97	5.11	2.79
										$\rho = 0.8,\ h = 0.3N$									
	20	0.07	0.00	0.09	0.01	0.08	-0.07	0.07	0.04	0.03	8.68	12.65	8.24	4.86	11.56	7.44	3.27	11.45	7.45
	50	0.03	-0.02	-0.06	0.06	0.02	0.01	-0.05	-0.05	-0.05	5.95	7.80	5.57	4.08	7.59	4.90	2.69	7.14	4.29
	100	-0.06	-0.08	0.07	0.07	-0.07	-0.09	0.02	0.04	0.03	5.22	5.17	3.91	3.62	5.06	3.41	2.42	4.89	3.20
										$\rho = 0.5,\ h = 2$									
	20	-0.30	0.02	0.02	-0.29	-0.04	-0.11	-0.26	0.04	-0.05	4.13	12.25	8.43	3.13	11.69	7.87	2.59	11.47	7.29
	50	-0.08	0.07	-0.06	-0.05	0.07	0.00	-0.10	0.00	0.04	2.34	7.59	5.43	1.91	7.46	4.87	1.52	7.55	4.48
	100	-0.09	0.04	0.01	-0.03	0.11	0.01	0.02	0.00	0.01	1.51	5.29	3.91	1.13	5.27	3.30	0.490	5.14	2.86
										$\rho = 0.5,\ h = 6$									
	20	-2.35	-0.21	-0.15	-2.13	-0.07	-0.19	-2.22	-0.10	-0.22	8.86	12.03	8.63	6.62	11.87	7.96	5.05	11.81	7.33
	50	-0.15	0.04	-0.06	-0.19	0.09	0.07	-0.17	0.13	-0.03	3.64	7.86	5.40	2.88	7.21	4.87	2.07	7.08	4.59
	100	-0.04	-0.07	0.01	0.07	0.06	-0.03	-0.12	-0.10	-0.07	2.58	5.39	3.81	1.84	5.09	3.35	1.27	5.02	2.82
CCE-IV										$\rho = 0.8,\ h = 6$									
	20	-4.23	-0.18	-0.26	-4.08	-0.20	-0.36	-4.23	-0.11	-0.24	7.87	12.30	8.33	6.41	11.76	7.89	5.72	11.18	7.20
	50	-0.31	-0.03	-0.07	-0.26	0.05	0.01	-0.14	0.03	-0.04	2.43	7.52	5.32	1.90	7.37	4.85	1.55	7.25	4.40
	100	-0.07	-0.02	0.03	-0.11	0.04	0.01	-0.05	0.09	0.00	1.48	5.52	3.93	1.52	5.32	3.29	0.99	5.09	2.83
										$\rho = 0.8,\ h = 0.3N$									
	20	-4.16	-0.17	-0.23	-4.05	-0.05	-0.12	-4.27	-0.25	-0.06	7.69	12.47	8.28	6.29	11.49	7.34	5.07	11.40	7.48
	50	-3.59	0.04	-0.55	-3.36	0.05	-0.17	-3.64	-0.18	-0.22	6.17	7.80	5.38	5.05	7.60	4.87	3.97	7.20	4.51
	100	-2.73	-0.14	0.03	-2.88	-0.16	0.11	-2.76	0.13	-0.15	6.13	5.20	3.96	4.34	5.04	3.51	3.51	4.99	3.16
										$\rho = 0.5,\ h = 2$									
	20	-0.09	0.04	0.07	-0.07	-0.03	0.04	0.00	0.05	-0.02	3.40	12.17	8.23	2.70	11.57	7.70	2.45	11.30	6.96
	50	-0.01	0.05	0.08	-0.01	0.08	-0.04	-0.02	-0.01	0.04	2.15	7.49	5.29	1.68	7.42	4.71	1.48	7.50	4.30
	100	-0.05	-0.01	-0.07	0.02	0.10	0.00	-0.02	0.01	0.00	1.39	5.32	3.45	1.08	5.21	3.19	0.39	5.16	2.65
										$\rho = 0.5,\ h = 6$									
	20	0.02	0.03	-0.02	-0.04	-0.03	-0.09	0.09	-0.02	-0.01	4.18	11.94	7.97	3.32	11.87	7.63	2.70	11.68	7.15
	50	0.09	-0.09	0.00	-0.01	0.05	0.09	0.05	0.05	-0.06	2.56	7.76	5.13	2.14	7.19	4.77	1.52	7.16	4.30
	100	-0.05	-0.07	0.06	0.05	0.09	0.01	-0.06	-0.05	-0.07	1.83	5.43	3.47	1.65	5.10	3.24	1.17	5.03	2.70
Infeasible										$\rho = 0.8,\ h = 6$									
	20	-0.03	-0.06	-0.03	-0.03	-0.02	0.03	0.06	0.08	0.04	2.73	11.95	8.15	2.37	11.63	7.62	2.03	11.12	6.98
	50	-0.05	0.10	-0.03	0.00	0.02	-0.04	0.06	0.02	-0.02	1.56	7.36	5.27	1.38	7.36	4.76	1.41	7.08	4.19
	100	-0.08	-0.06	0.01	-0.03	0.03	0.02	-0.02	0.08	0.02	1.16	5.31	3.53	0.88	5.24	3.30	0.84	5.08	2.67
										$\rho = 0.8,\ h = 0.3N$									
	20	0.02	0.04	0.07	0.03	0.09	-0.01	0.09	-0.02	-0.05	2.84	12.35	8.17	2.45	11.53	7.31	2.23	11.43	7.34
	50	0.00	0.04	0.03	0.03	0.04	-0.01	0.00	-0.01	0.00	1.90	7.55	5.25	1.70	7.41	4.73	1.30	7.13	4.36
	100	0.02	-0.02	0.08	0.03	0.05	0.09	-0.03	0.02	-0.04	1.24	5.14	3.90	1.15	5.03	3.55	0.86	4.93	3.00

Notes: The simulation results are based on the DGP specified in Section 4.

Table 4: The Size and Power for Individual Estimators under Experiment 4

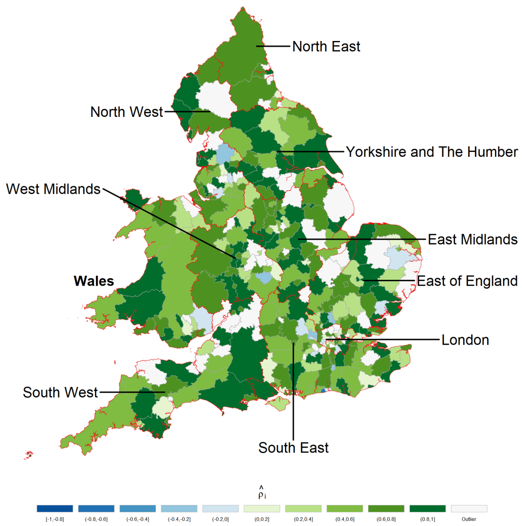
		Size									Power								
		20			50			100			20			50			100		
	(N,T)	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}	ρ_i	β_{1i}	β_{2i}
										$\rho = 0.5,\ h = 2$									
CCEX-IV	20	0.076	0.075	0.078	0.068	0.074	0.070	0.054	0.048	0.050	0.547	0.249	0.270	0.717	0.421	0.436	0.956	0.747	0.723
	50	0.073	0.076	0.070	0.063	0.071	0.079	0.055	0.062	0.059	0.571	0.282	0.262	0.779	0.450	0.450	0.986	0.768	0.772
	100	0.077	0.073	0.077	0.065	0.074	0.062	0.059	0.046	0.050	0.586	0.289	0.260	0.792	0.466	0.467	0.993	0.754	0.761
										$\rho = 0.5,\ h = 6$									
	20	0.073	0.073	0.064	0.044	0.069	0.057	0.052	0.060	0.065	0.258	0.258	0.267	0.348	0.433	0.439	0.517	0.737	0.740
	50	0.056	0.073	0.074	0.062	0.075	0.061	0.059	0.047	0.054	0.279	0.275	0.269	0.444	0.458	0.458	0.726	0.787	0.769
	100	0.076	0.072	0.071	0.069	0.067	0.061	0.053	0.051	0.054	0.322	0.295	0.279	0.492	0.488	0.453	0.753	0.776	0.779
										$\rho = 0.8,\ h = 6$									
	20	0.078	0.067	0.059	0.059	0.051	0.061	0.048	0.056	0.055	0.428	0.244	0.242	0.579	0.432	0.390	0.882	0.730	0.708
	50	0.059	0.079	0.072	0.052	0.073	0.063	0.049	0.066	0.051	0.695	0.268	0.289	0.871	0.460	0.451	0.975	0.757	0.762
	100	0.067	0.076	0.078	0.058	0.070	0.059	0.052	0.050	0.052	0.710	0.295	0.276	0.866	0.445	0.443	0.996	0.785	0.765
										$\rho = 0.8,\ h = 0.3N$									
	20	0.058	0.058	0.052	0.055	0.054	0.055	0.054	0.055	0.052	0.484	0.268	0.265	0.607	0.420	0.434	0.757	0.740	0.724
	50	0.062	0.055	0.049	0.051	0.053	0.044	0.058	0.061	0.046	0.354	0.279	0.258	0.483	0.442	0.452	0.620	0.763	0.750
	100	0.058	0.073	0.041	0.057	0.035	0.038	0.047	0.048	0.047	0.249	0.250	0.279	0.332	0.467	0.441	0.492	0.760	0.764
										$\rho = 0.5,\ h = 2$									
CCE-IV	20	0.072	0.072	0.088	0.075	0.079	0.066	0.071	0.060	0.058	0.543	0.246	0.270	0.727	0.418	0.442	0.954	0.744	0.752
	50	0.069	0.079	0.089	0.076	0.062	0.063	0.058	0.059	0.060	0.569	0.268	0.258	0.779	0.441	0.449	0.981	0.762	0.780
	100	0.073	0.081	0.082	0.055	0.064	0.060	0.055	0.055	0.062	0.601	0.261	0.290	0.809	0.445	0.446	0.967	0.772	0.758
										$\rho = 0.5,\ h = 6$									
	20	0.042	0.063	0.069	0.046	0.055	0.051	0.044	0.073	0.065	0.238	0.251	0.258	0.303	0.434	0.450	0.498	0.728	0.739
	50	0.071	0.083	0.071	0.050	0.077	0.064	0.063	0.054	0.054	0.272	0.281	0.273	0.408	0.439	0.452	0.72	0.770	0.753
	100	0.066	0.072	0.075	0.056	0.076	0.065	0.052	0.053	0.054	0.319	0.281	0.273	0.496	0.462	0.476	0.749	0.762	0.762
										$\rho = 0.8,\ h = 6$									
	20	0.035	0.050	0.050	0.036	0.048	0.044	0.034	0.051	0.037	0.420	0.244	0.246	0.529	0.428	0.417	0.738	0.698	0.688
	50	0.040	0.078	0.075	0.049	0.070	0.060	0.046	0.054	0.064	0.691	0.254	0.277	0.837	0.433	0.437	0.948	0.753	0.774
	100	0.052	0.069	0.074	0.061	0.063	0.074	0.049	0.054	0.050	0.700	0.261	0.255	0.889	0.480	0.452	0.976	0.791	0.762
										$\rho = 0.8,\ h = 0.3N$									
	20	0.038	0.041	0.051	0.038	0.056	0.060	0.037	0.043	0.044	0.412	0.256	0.240	0.559	0.406	0.433	0.667	0.706	0.703
	50	0.037	0.042	0.053	0.031	0.045	0.040	0.043	0.038	0.048	0.310	0.275	0.264	0.453	0.430	0.434	0.550	0.752	0.748
	100	0.025	0.034	0.046	0.038	0.046	0.040	0.028	0.047	0.049	0.229	0.263	0.251	0.306	0.440	0.437	0.425	0.769	0.753
										$\rho = 0.5,\ h = 2$									
Infeasible	20	0.074	0.077	0.076	0.054	0.064	0.063	0.052	0.051	0.042	0.681	0.310	0.292	0.898	0.470	0.500	1.000	0.829	0.791
	50	0.068	0.073	0.074	0.062	0.051	0.059	0.055	0.051	0.049	0.706	0.322	0.310	0.847	0.519	0.476	1.000	0.795	0.788
	100	0.075	0.075	0.078	0.052	0.060	0.054	0.045	0.057	0.054	0.659	0.312	0.292	0.876	0.483	0.496	0.984	0.801	0.799
										$\rho = 0.5,\ h = 6$									
	20	0.074	0.082	0.067	0.059	0.064	0.056	0.052	0.049	0.046	0.468	0.281	0.310	0.723	0.498	0.493	0.939	0.798	0.805
	50	0.069	0.069	0.075	0.058	0.052	0.062	0.050	0.059	0.060	0.456	0.318	0.286	0.668	0.510	0.489	0.929	0.822	0.799
	100	0.077	0.069	0.076	0.059	0.055	0.065	0.043	0.061	0.051	0.415	0.280	0.312	0.671	0.522	0.496	0.906	0.795	0.788
										$\rho = 0.8,\ h = 6$									
	20	0.069	0.075	0.078	0.067	0.062	0.063	0.046	0.042	0.045	0.803	0.305	0.265	0.972	0.515	0.464	0.992	0.810	0.790
	50	0.076	0.083	0.073	0.057	0.056	0.060	0.050	0.048	0.058	0.955	0.305	0.287	0.976	0.501	0.484	0.992	0.801	0.784
	100	0.067	0.066	0.081	0.056	0.057	0.064	0.055	0.043	0.057	0.906	0.297	0.290	1.000	0.514	0.504	1.000	0.822	0.789
										$\rho = 0.8,\ h = 0.3N$									
	20	0.059	0.069	0.066	0.051	0.051	0.069	0.050	0.057	0.057	0.862	0.304	0.270	0.979	0.501	0.479	0.988	0.801	0.784
	50	0.061	0.064	0.069	0.041	0.056	0.058	0.047	0.058	0.043	0.655	0.289	0.313	0.863	0.494	0.478	1.000	0.811	0.798
	100	0.044	0.070	0.065	0.044	0.049	0.054	0.046	0.045	0.048	0.450	0.305	0.300	0.689	0.488	0.492	1.000	0.809	0.795

Notes: The simulation results are based on the DGP specified in Section 4. The alternative for the power test is $\boldsymbol{\theta}_i^a = \boldsymbol{\theta}_i + 0.2$.

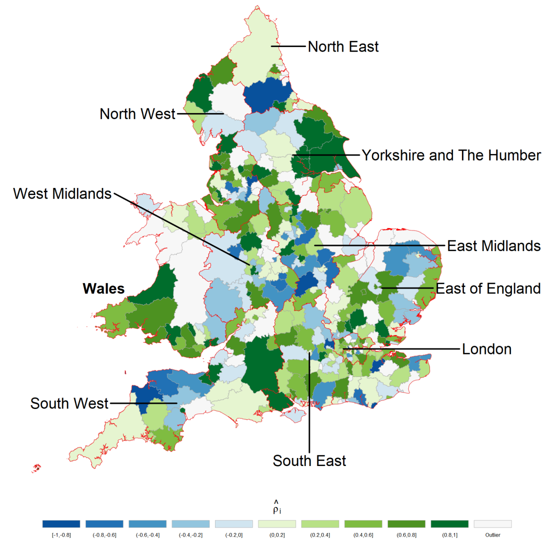
Table 5: The Size and Power for Mean Group Estimators under Experiment 4

		Size									Power								
		20			50			100			20			50			100		
	(N,T)	ρ	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2	ρ	β_1	β_2
	$\rho = 0.5,\ h = 2$																		
	20	0.072	0.067	0.075	0.059	0.055	0.055	0.052	0.048	0.064	0.605	0.307	0.470	0.776	0.407	0.684	0.920	0.419	0.927
	50	0.065	0.044	0.075	0.057	0.053	0.064	0.049	0.048	0.057	0.944	0.648	0.812	0.973	0.753	0.981	0.993	0.852	1.000
	100	0.048	0.040	0.051	0.054	0.053	0.054	0.055	0.046	0.057	0.982	0.951	1.000	1.000	0.983	0.982	1.000	1.000	0.991
	$\rho = 0.5,\ h = 6$																		
	20	0.065	0.061	0.062	0.066	0.059	0.060	0.063	0.066	0.057	0.240	0.334	0.496	0.334	0.386	0.630	0.492	0.458	0.944
	50	0.046	0.061	0.048	0.056	0.057	0.060	0.062	0.046	0.046	0.795	0.700	0.855	0.881	0.755	0.999	0.963	0.823	1.000
	100	0.048	0.051	0.056	0.050	0.046	0.051	0.054	0.049	0.047	0.938	0.960	1.000	1.000	0.981	0.989	0.982	0.997	1.000
CCEX-IV	$\rho = 0.8,\ h = 6$																		
	20	0.054	0.051	0.061	0.051	0.065	0.058	0.048	0.055	0.045	0.446	0.298	0.493	0.549	0.345	0.679	0.782	0.505	0.894
	50	0.050	0.059	0.058	0.055	0.045	0.060	0.048	0.055	0.051	0.956	0.695	0.804	0.980	0.805	0.991	0.992	0.884	0.981
	100	0.055	0.059	0.046	0.052	0.051	0.065	0.049	0.055	0.056	0.979	1.000	0.988	0.985	0.998	1.000	1.000	1.000	1.000
	$\rho = 0.8,\ h = 0.3N$																		
	20	0.060	0.063	0.044	0.064	0.055	0.054	0.047	0.047	0.062	0.428	0.273	0.503	0.576	0.355	0.626	0.758	0.514	0.830
	50	0.055	0.047	0.058	0.052	0.057	0.056	0.048	0.057	0.055	0.521	0.630	0.797	0.689	0.696	0.886	0.893	0.756	0.951
	100	0.058	0.062	0.080	0.048	0.056	0.064	0.048	0.046	0.047	0.523	0.919	0.970	0.761	0.933	0.994	0.933	0.998	1.000
	$\rho = 0.5,\ h = 2$																		
	20	0.078	0.076	0.076	0.071	0.083	0.047	0.076	0.063	0.075	0.607	0.296	0.432	0.756	0.422	0.678	0.903	0.418	0.897
	50	0.073	0.045	0.060	0.062	0.062	0.076	0.049	0.066	0.043	0.957	0.650	0.816	0.971	0.746	0.960	0.997	0.840	1.000
	100	0.060	0.054	0.050	0.059	0.058	0.039	0.048	0.057	0.058	0.983	0.980	1.000	1.000	1.000	1.000	0.988	0.993	1.000
	$\rho = 0.5,\ h = 6$																		
	20	0.079	0.072	0.059	0.088	0.082	0.067	0.093	0.078	0.057	0.209	0.343	0.483	0.343	0.348	0.639	0.525	0.433	0.941
	50	0.052	0.069	0.050	0.057	0.062	0.055	0.059	0.048	0.057	0.788	0.706	0.847	0.903	0.742	0.975	0.981	0.825	1.000
	100	0.054	0.054	0.064	0.050	0.044	0.056	0.053	0.074	0.055	0.937	0.969	1.000	1.000	0.999	0.986	0.993	1.000	1.000
CCE-IV	$\rho = 0.8,\ h = 6$																		
	20	0.089	0.070	0.068	0.088	0.073	0.065	0.085	0.084	0.063	0.317	0.281	0.514	0.488	0.396	0.678	0.618	0.500	0.907
	50	0.062	0.070	0.068	0.053	0.052	0.054	0.063	0.058	0.045	0.928	0.713	0.826	0.972	0.798	0.968	0.994	0.881	1.000
	100	0.049	0.063	0.068	0.048	0.049	0.052	0.047	0.042	0.064	0.982	0.984	0.983	1.000	0.984	0.993	1.000	0.991	1.000
	$\rho = 0.8,\ h = 0.3N$																		
	20	0.096	0.072	0.062	0.098	0.066	0.060	0.095	0.072	0.080	0.301	0.268	0.509	0.421	0.353	0.629	0.554	0.500	0.829
	50	0.060	0.057	0.069	0.078	0.052	0.066	0.086	0.064	0.055	0.450	0.633	0.843	0.557	0.701	0.918	0.777	0.739	0.942
	100	0.070	0.054	0.081	0.066	0.045	0.075	0.097	0.043	0.047	0.404	0.898	0.989	0.734	0.936	0.989	0.889	0.977	1.000
	$\rho = 0.5,\ h = 2$																		
	20	0.076	0.051	0.070	0.047	0.064	0.051	0.054	0.060	0.063	0.755	0.322	0.505	0.878	0.381	0.717	0.919	0.433	0.924
	50	0.060	0.060	0.072	0.070	0.063	0.058	0.060	0.048	0.050	0.982	0.658	0.865	0.980	0.749	0.972	1.000	0.863	1.000
	100	0.053	0.044	0.055	0.048	0.060	0.057	0.056	0.060	0.049	1.000	0.973	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$\rho = 0.5,\ h = 6$																		
	20	0.077	0.061	0.061	0.055	0.068	0.062	0.041	0.056	0.058	0.593	0.336	0.546	0.791	0.354	0.706	0.864	0.456	0.927
	50	0.047	0.069	0.072	0.048	0.051	0.058	0.051	0.044	0.057	0.905	0.742	0.872	0.989	0.781	0.991	0.986	0.849	0.987
	100	0.048	0.053	0.049	0.051	0.041	0.042	0.049	0.063	0.046	0.985	0.975	0.982	1.000	0.978	1.000	1.000	1.000	0.987
Infeasible	$\rho = 0.8,\ h = 6$																		
	20	0.049	0.055	0.064	0.052	0.064	0.068	0.059	0.045	0.054	0.927	0.387	0.520	0.955	0.389	0.743	0.982	0.563	0.916
	50	0.043	0.056	0.060	0.053	0.055	0.047	0.048	0.051	0.041	1.000	0.707	0.861	0.983	0.807	0.981	1.000	0.909	1.000
	100	0.055	0.054	0.060	0.050	0.043	0.051	0.051	0.055	0.060	0.987	0.989	0.994	1.000	1.000	1.000	1.000	0.988	0.986
	$\rho = 0.8,\ h = 0.3N$																		
	20	0.060	0.070	0.067	0.063	0.054	0.061	0.053	0.053	0.053	0.884	0.384	0.525	0.950	0.388	0.749	0.967	0.542	0.925
	50	0.055	0.055	0.059	0.057	0.066	0.063	0.052	0.058	0.049	0.977	0.660	0.839	0.994	0.772	0.885	0.980	0.861	0.957
	100	0.047	0.055	0.057	0.051	0.050	0.059	0.052	0.045	0.051	0.988	0.904	0.987	1.000	0.956	0.991	1.000	1.000	1.000

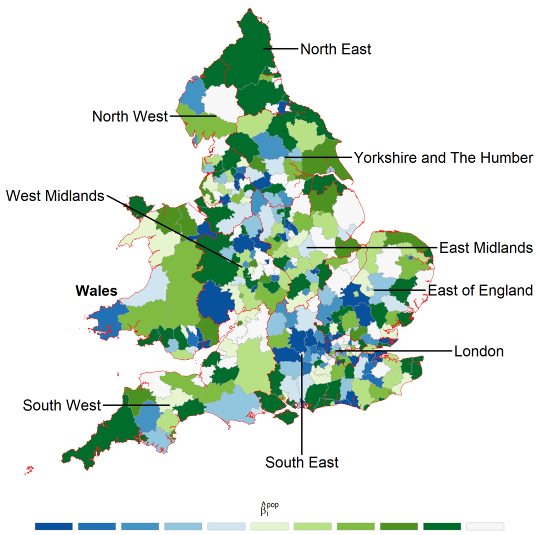
Notes: The simulation results are based on the DGP specified in Section 4. The alternative for the power test is $\boldsymbol{\theta}^a = \boldsymbol{\theta} + 0.2$.



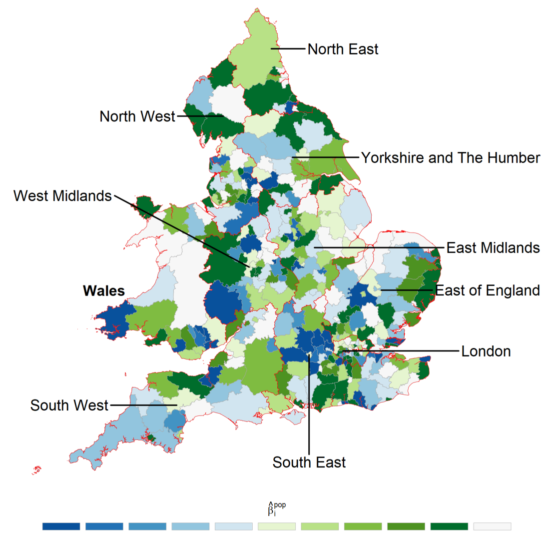
(a) CCEX-IV estimates of ρ_i



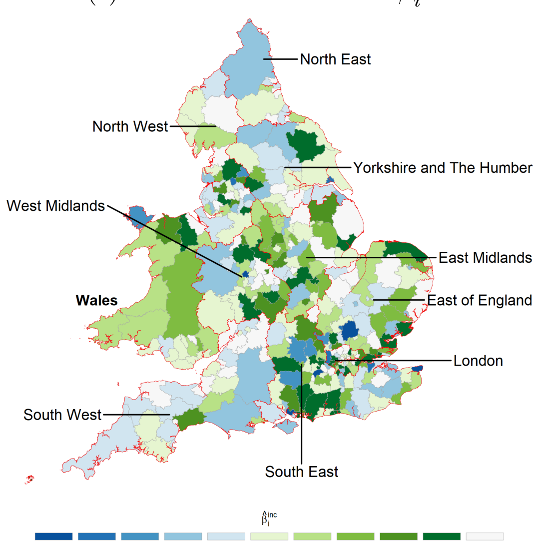
(b) CCE-IV estimates of ρ_i



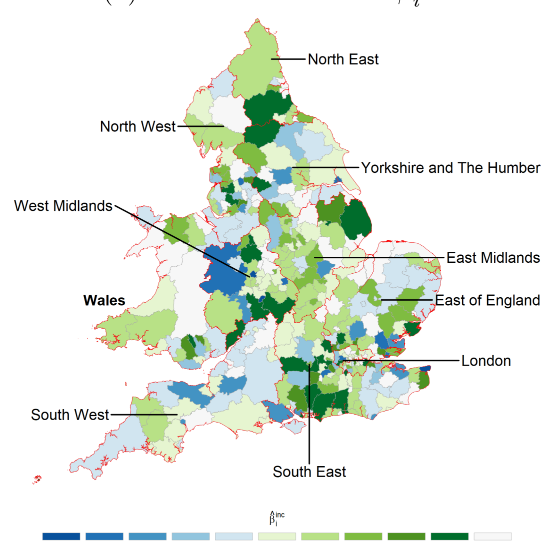
(c) CCEX-IV estimates of β_i^{pop}



(d) CCE-IV estimates of β_i^{pop}

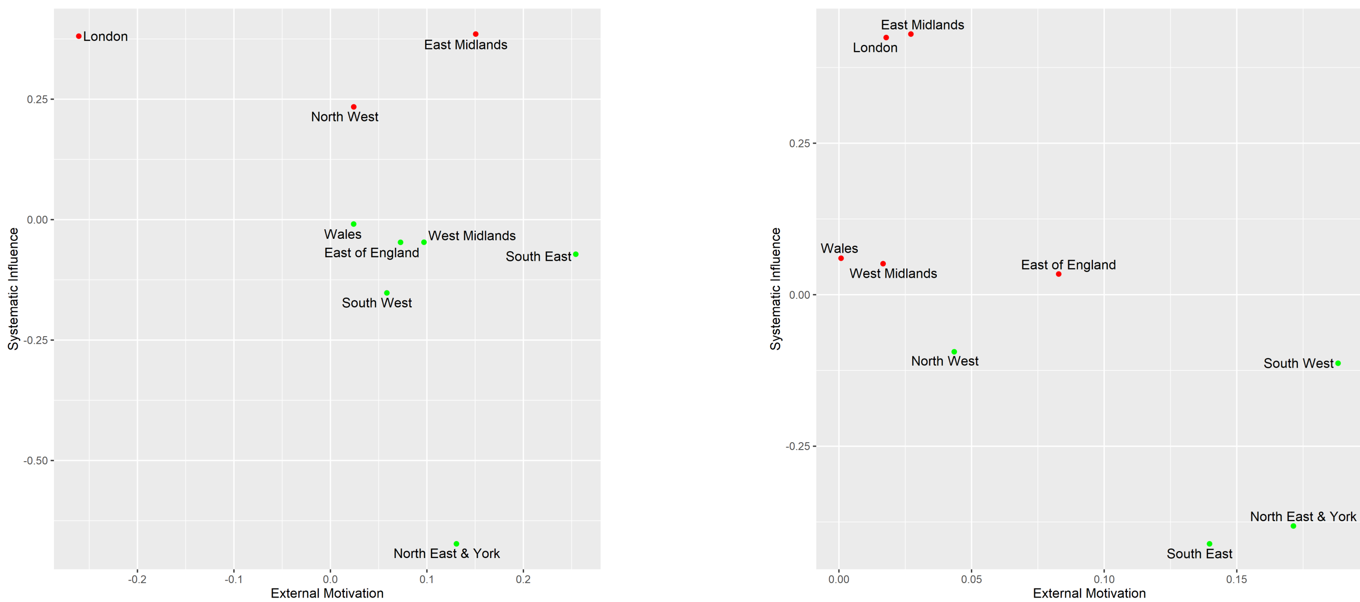


(e) CCEX-IV estimates of β_i^{inc}



(f) CCE-IV estimates of β_i^{inc}

Figure 1: The Spatial Patterns of Individual Coefficients for LADs



(a) The population growth (b) The income growth

Figure 2: Regional GCM Analysis

Table 6: Mean Group Estimation Results at the Regional and National Level

Region	N_r	$\hat{\rho}_{MG,r}$	$\hat{\beta}^{pop}_{MG,r}$	$\hat{\beta}^{inc}_{MG,r}$	CD test	$\hat{\alpha}$	N_r	$\hat{\rho}_{MG,r}$	$\hat{\beta}^{pop}_{MG,r}$	$\hat{\beta}^{inc}_{MG,r}$	CD test	$\hat{\alpha}$
		CCEX-IV Estimation						CCE-IV Estimation				
North East&York	26	$0.699^{\ddagger}$	$2.504^{\ddagger}$	0.161	1.44	0.65	24	$0.415^{\ddagger}$	$2.077^{\dagger}$	0.421^{*}	0.40	0.60
		(0.041)	(0.917)	(0.185)	[0.15]	(0.04)		(0.086)	(1.004)	(0.246)	[0.69]	(0.04)
North West	33	$0.596^{\ddagger}$	1.340^{*}	0.283	4.97	0.76	32	$0.346^{\ddagger}$	1.164	0.284	3.03	0.63
		(0.053)	(0.813)	(0.199)	[0.00]	(0.04)		(0.091)	(0.852)	(0.188)	[0.00]	(0.04)
East Midlands	35	$0.633^{\ddagger}$	0.492	$0.863^{\ddagger}$	6.38	0.76	36	0.064	0.680	$0.724^{\ddagger}$	3.14	0.65
		(0.038)	(0.726)	(0.138)	[0.00]	(0.05)		(0.091)	(0.512)	(0.155)	[0.00]	(0.02)
West Midlands	24	$0.539^{\ddagger}$	0.352	$0.862^{\ddagger}$	6.86	0.79	26	0.078	0.258	0.366	0.79	0.54
		(0.062)	(1.022)	(0.338)	[0.00]	(0.04)		(0.092)	(0.738)	(0.420)	[0.43]	(0.03)
East of England	40	$0.550^{\ddagger}$	0.963	$0.747^{\ddagger}$	13.28	0.82	36	$0.240^{\ddagger}$	$1.605^{\dagger}$	$0.455^{\ddagger}$	2.44	0.68
		(0.051)	(0.770)	(0.261)	[0.00]	(0.04)		(0.069)	(0.765)	(0.194)	[0.02]	(0.04)
London	25	$0.400^{\ddagger}$	0.161	$1.059^{\ddagger}$	2.83	0.70	25	$0.308^{\ddagger}$	$1.693^{\ddagger}$	$0.629^{\ddagger}$	0.83	0.60
		(0.081)	(0.660)	(0.332)	[0.00]	(0.03)		(0.075)	(0.589)	(0.262)	[0.41]	(0.04)
South East	59	$0.512^{\ddagger}$	0.194	0.632^{*}	13.86	0.82	60	$0.145^{\ddagger}$	-0.360	$0.707^{\ddagger}$	10.88	0.76
		(0.044)	(0.653)	(0.323)	[0.00]	(0.04)		(0.056)	(0.624)	(0.166)	[0.00]	(0.03)
South West	21	$0.668^{\ddagger}$	$2.453^{\ddagger}$	0.187	7.68	0.81	26	$0.200^{\dagger}$	0.754	0.231	2.59	0.74
		(0.057)	(0.810)	(0.156)	[0.00]	(0.04)		(0.096)	(0.691)	(0.297)	[0.01]	(0.03)
Wales	19	$0.605^{\ddagger}$	0.972	$0.573^{\ddagger}$	1.79	0.70	16	$0.470^{\ddagger}$	0.889	0.409^{*}	-1.51	0.61
		(0.069)	(0.855)	(0.162)	[0.00]	(0.04)		(0.089)	(0.902)	(0.244)	[0.13]	(0.03)
England&Wales	282	$0.569^{\ddagger}$	$0.922^{\ddagger}$	$0.557^{\ddagger}$	9.04	0.68	281	$0.225^{\ddagger}$	$0.821^{\ddagger}$	$0.505^{\ddagger}$	0.73	0.56
		(0.019)	(0.272)	(0.079)	[0.00]	(0.03)		(0.028)	(0.251)	(0.078)	[0.46]	(0.02)

Notes: All the Mean Group estimates are calculated as simple averages from district level parameter estimates and the standard errors in () are calculated using the formula in (24). The superscripts ‡,† and * denote the significance at 1, 5 and 10% level. CD test is the statistic for the null of weak cross section dependence proposed by Pesaran (2015) with the figure in [] indicating the p-value. $\hat{\alpha}$ is the estimate of exponent of cross section dependence proposed by Bailey et al. (2016b) with the figure in () indicating the standard error.

Table 7: Three-Year Average (2014-2016) of Demographic Statistics for Each Region

Region	North East	North West	Yorkshire & Humber	East Midlands	West Midlands	East of England	London	South East	South West	Wales
Population Density (people/km^2)	308	515	354	305	451	323	5620	476	233	151
Per Capita Income (£)	15725	16833	16224	16921	16667	19995	26862	22218	19079	15754
House Price (£)	150753	170169	169996	182726	192571	274760	538803	324207	247792	163723
Household Growth (%)	2.48	0.73	1.38	-0.57	0.84	0.72	0.54	0.53	1.18	2.24

Table 8: Regional Direct, Spill-in, Spill-out and Net Effects

Region	North East & York	North West	East Midlands	West Midlands	East of England	London	South East	South West	Wales
	Population Growth								
	Regional Connectedness Matrix (%)								
North East & York	563.55	24.20	60.41	0.04	0.02	0.00	0.00	0.00	0.00
North West	2.10	273.57	-0.01	5.10	0.00	0.00	0.00	0.00	-0.39
East Midlands	1.38	4.94	83.18	-1.19	0.14	-0.01	9.90	0.01	0.00
West Midlands	0.04	0.57	1.11	110.02	0.00	0.00	-0.02	7.59	2.53
East of England	0.04	0.00	0.02	0.00	158.56	19.97	-6.57	0.00	0.00
London	0.00	0.00	0.00	0.00	-4.17	31.59	-6.98	0.00	0.00
South East	0.00	0.00	-0.05	0.30	11.79	14.72	78.48	0.04	0.00
South West	0.07	0.22	0.04	0.86	0.00	0.01	21.83	494.13	7.66
Wales	0.04	5.02	0.02	1.05	0.00	0.00	0.00	4.76	442.10
	Regional Effects (%)								
RSI	84.68	6.80	15.17	11.82	13.46	-11.15	26.8	30.71	10.90
RSO	3.68	34.97	61.54	6.17	7.79	34.69	18.16	12.39	9.80
RNE	-81.00	28.17	46.36	-5.65	-5.67	45.84	-8.64	-18.31	-1.10
	Personal Income Growth								
	Regional Connectedness Matrix (%)								
North East & York	84.27	-2.46	20.79	0.10	0.00	0.00	0.00	0.00	0.00
North West	0.36	44.66	0.02	1.57	0.00	0.00	0.00	0.00	0.08
East Midlands	0.84	0.03	182.64	1.44	0.09	0.02	2.67	0.00	0.00
West Midlands	0.04	0.20	0.50	93.50	0.00	0.00	0.01	-0.02	0.86
East of England	0.00	0.00	2.59	0.00	99.40	5.90	0.49	0.00	0.00
London	0.00	0.00	0.00	0.00	1.01	150.43	1.72	0.00	0.00
South East	0.00	0.00	0.450	0.17	9.40	15.85	160.55	0.21	0.00
South West	0.02	0.02	0.06	0.62	0.01	0.04	2.69	22.93	1.86
Wales	0.02	0.01	0.02	-0.01	0.00	0.00	0.01	0.05	121.97
	Regional Effects (%)								
RSI	18.44	2.03	5.09	1.58	8.98	2.73	26.08	5.31	0.10
RSO	1.28	-2.21	24.44	3.88	10.52	21.81	7.59	0.23	2.80
RNE	-17.16	-4.23	19.34	2.30	1.54	19.08	-18.49	-5.08	2.71