CrossMark

A multidimensional spatial lag panel data model with spatial moving average nested random effects errors

Bernard Fingleton[1] · **Julie Le Gallo[2]** ·
Alain Pirotte[3]

Abstract This paper focuses on a three-dimensional model that combines two different types of spatial interaction effects, i.e. endogenous interaction effects *via* a spatial lag on the dependent variable and interaction effects among the disturbances *via* a spatial moving average (SMA) nested random effects errors. A three-stage procedure is proposed to estimate the parameters. In a first stage, the spatial lag panel data model is estimated using an instrumental variable (IV) estimator. In a second stage, a generalized moments (GM) approach is developed to estimate the SMA parameter and the variance components of the disturbance process using IV residuals from the first stage. In a third stage, to purge the equation of the specific structure of the disturbances a Cochrane–Orcutt-type transformation is applied combined with the IV principle. This leads to the GM spatial IV estimator and the regression parameter estimates. Monte Carlo simulations show that our estimators are not very different in terms of root mean square error from those produced by maximum likelihood. The approach is applied to European Union regional employment data for regions nested within countries.

✉ Bernard Fingleton
bf100@cam.ac.uk

Julie Le Gallo
julie.le-gallo@agrosupdijon.fr

Alain Pirotte
alain.pirotte@u-paris2.fr

[1] Department of Land Economy, University of Cambridge, Cambridge CB3 9EP, UK

[2] CESAER UMR1041, AgroSup Dijon, INRA, University Bourgogne Franche-Comté,
26, Boulevard Petitjean, 21000 Dijon, France

[3] CRED, University Paris II Panthéon-Assas, 12 Place du Panthéon, 75 231 Paris Cedex 05, France

🙲 Springer

1 Introduction

Recently, Fingleton et al. (2016) introduced a generalization of the Kapoor et al. (2007) (hereafter KKP) generalized moments (GM) procedure to multidimensional panel data models assuming that the disturbances follow a first-order spatial autoregressive (SAR) process, which includes a nested random effects structure, namely SAR-NRE. They refer to this specification as a panel data model with spatially nested random effects disturbances. They derive a spatial feasible generalized least squares (S-FGLS) estimator for the model's regression parameters which uses the GM parameter estimates of the SAR parameter and the variance components of the disturbance process, namely GM-S-FGLS. This estimator is based on a spatial counterpart to the Cochrane–Orcutt transformation, as well as on transformations which are used in the estimation of classical error component models.

In this paper, we consider a more general multidimensional panel data model which includes a spatial lag and where the disturbances are assumed to follow a spatial moving average (SMA) process (local spatial spillover effects) in the spirit of Fingleton (2008). This structure constitutes an alternative to incorporating spatial lags on the explanatory variables. In the cross-sectional case, when the model contains a spatial lag dependent variable, Kelejian and Prucha (1998, 1999) suggest a 2SLS procedure. They propose that the instrument set should be kept to a low order to avoid linear dependence and retain full column rank for the matrix of instruments, and thus recommend that (X, WX) should be used, if the number of regressors is large. Inclusion of spatial lags of the explanatory variables could have a major impact on the performance of the estimation procedures if one were to keep to this recommendation. Pace et al. (2012) show that instrumental variable estimation suffers greatly in situations where spatial lags of the explanatory variables (WX) are included in the model specification. The reason is that this requires the use of $(W^2X, W^3X, \ldots)$ as instruments, in place of the conventional instruments that rely on WX, and this appears to result in a weak instrument problem. Our motivation for the adoption of a SMA specification of the error process, which has been largely neglected in spatial econometrics, is that it mitigates against the problem for instrumental variable estimation identified by Pace et al. (2012). Naturally the choice of this specification should be predicated on the applied researcher, at a preliminary stage, examining the nature of local spillovers in order to establish its appropriateness for the empirical application at hand.

We propose a three-stage procedure to estimate the parameters. In a first stage, the spatial lag panel data model is estimated using an instrumental variable (IV) estimator. In a second stage, a GM approach is developed to estimate the SMA parameter and the variance components of the disturbance process using IV residuals from the first stage. In a third stage, to purge the equation of the specific structure of the disturbances, a Cochrane–Orcutt-type transformation combined with the IV principle is applied.

This leads to the GM spatial IV estimator and the regression parameter estimates of the spatial lag model. Monte Carlo simulations show that our estimates are not very different in terms of root mean square error compared to those produced by maximum likelihood (ML).

The outline of the paper is as follows: Sect. 2 presents the spatial lag panel data model with spatial moving average nested random effects errors, and Sect. 3 focuses on estimation methods. This section introduces a spatial GM instrumental variable approach to estimate the parameters of the model. Section 4 presents the Monte Carlo design and describes the Monte Carlo results. Section 5 illustrates our approach using an application to EU regional employment data for regions nested within countries. The last section concludes.

2 The spatial model

Our point of departure is a three-dimensional model that combines two different types of spatial interaction effects, i.e. endogenous interaction effects *via* a spatial lag on the dependent variable and interaction effects among the disturbances *via* a spatial moving average (SMA) process on the error term. The notation is as follows: the dependent variable y_{ijt} is observed along three indices, with $i = 1, \ldots, N$, $j = 1, \ldots, M_i$ and $t = 1, \ldots, T$. N denotes the number of groups. M_i denotes the number of individuals in group i, so in total there are $S = \sum_{i=1}^{N} M_i$ individuals. Since this model allows for an unequal number of individuals across the N groups, it is therefore unbalanced in the spatial dimension, although it is balanced in the time dimension. Hence, the model describes a hierarchical structure with the index j pertaining to individuals that are nested within the N groups. Assuming that spatial autocorrelation only takes place at the individual level and that the slope coefficients are homogenous, the model can be written as:

$$ y_{ijt} = \rho \sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{ij,gh} y_{ght} + x_{ijt}\beta + \varepsilon_{ijt}, \tag{1} $$

where y_{ijt} is the dependent variable; x_{ijt} is a $(1 \times K)$ vector of explanatory exogenous variables; β represents a $(K \times 1)$ vector of parameters to be estimated; and ε_{ijt} is the disturbance, the properties of which will be discussed below.

The weight $w_{ij,gh} = w_{k,l}$ is the $(k = ij; l = gh)$ element of the spatial matrix W_S with ij denoting individual j within group i, and similarly for gh. Thus, $k, l = 1, \ldots, S$ and W_S is a $(S \times S)$ matrix of known spatial weights which has zero on the leading diagonal and is usually row-normalized so that for row k, $\sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{k,gh} = 1$, although as we will illustrate in the empirical example other normalizations are permissible. We maintain the standard assumption concerning the weight matrix, i.e. W_S is assumed non-stochastic, and its row and column sums are required to be uniformly bounded in absolute value. ρ is the spatial lag parameter to be estimated. This coefficient is bounded numerically to ensure spatial stationarity, i.e. $e_{\min}^{-1} < \rho < 1$ where $e_{\min}$ is the minimum real characteristic root of W_S.

In this paper, we consider the case of disturbances ε_{ijt} that are contemporaneously correlated through a moving average process at the individual level:

$$\varepsilon_{ijt} = u_{ijt} - \lambda \sum_{g=1}^{N} \sum_{h=1}^{M_g} m_{ij,gh} u_{ght}. \tag{2}$$

The weight $m_{ij,gh}$ is an element of the spatial matrix M_S which satisfies the same assumptions as for W_S. For simplicity in the following, we assume that $M_S = W_S$. λ is the spatial moving average parameter to be estimated. u_{ijt} is assumed to be i.i.d.$(0, \sigma_u^2)$. Spatial heterogeneity is captured through a random effects structure for the errors u_{ijt}: it contains an unobserved permanent unit-specific error component α_i, a nested permanent unit-specific error component μ_{ij} together with a remainder error component v_{ijt}. Hence, we envisage a time-invariant group effect applying equally to all individuals nested within a group, time-invariant individual group-specific effects and transient effects that vary at random across groups, individuals and time. More formally:

$$u_{ijt} = \alpha_i + \mu_{ij} + v_{ijt}, \tag{3}$$

in which α_i is the unobservable group-specific time-invariant effect which is assumed to be i.i.d.$N\left(0, \sigma_\alpha^2\right)$; μ_{ij} is the nested effect of individual j within the ith group which is assumed to be i.i.d.$N\left(0, \sigma_\mu^2\right)$ and v_{ijt} is the remainder term which is also assumed to be i.i.d.$N\left(0, \sigma_v^2\right)$. The α_i's, μ_{ij}'s and v_{ijt}'s are independent of each other and among themselves.

In contrast to the classical literature on panel data, grouping the data by periods rather than units is more convenient when we consider spatial autocorrelation. For a cross section t, Eqs. (1), (2) and (3) can be written as:

$$y_t = \rho W_S y_t + x_t \beta + \varepsilon_t, \tag{4}$$

where y_t is of dimension $(S \times 1)$ and x_t is an $(S \times K)$ matrix of explanatory variables that are assumed to be exogenous and non-stochastic and have elements uniformly bounded in absolute value. The first-order moving average error process ϵ_t is given by

$$\varepsilon_t = u_t - \lambda W_S u_t, \tag{5}$$

with

$$u_t = \text{diag}\left(\iota_{M_i}\right)\alpha + \mu + v_t, \tag{6}$$

where u_t is $(S \times 1)$, α is the vector of group effects of dimension $(N \times 1)$, $\mu^\top = \left(\mu_1^\top, \ldots, \mu_N^\top\right)$, a vector of dimension $(1 \times S)$, $\mu_i^\top = \left(\mu_{i1}, \ldots, \mu_{iM_i}\right)$, a vector of dimension $(1 \times M_i)$, ι_{M_i} is a vector of ones of dimension $(M_i \times 1)$. By $\text{diag}\left(\iota_{M_i}\right)$, we mean $\text{diag}\left(\iota_{M_1}, \ldots, \iota_{M_N}\right)$. Finally v_t is of dimension $(S \times 1)$.

Stacking the T cross sections gives

$$y = Z\delta + \varepsilon, \tag{7}$$

and

$$\varepsilon = u - \lambda W u, \tag{8}$$

with $Z = [Wy, X]$ and $\delta = [\rho, \beta]^\top$. y and X are the vector and matrix of the dependent and explanatory variables (covariates), respectively, of size $(TS \times 1)$ and $(TS \times K)$; β is the vector of the slope parameters of size $(K \times 1)$; and finally ε is the vector of the disturbance terms of dimension $(TS \times 1)$. Given that I_T is an identity matrix of dimension $(T \times T)$, then $W = (I_T \otimes W_S)$ is of size $(TS \times TS)$. Finally, for the full $(TS \times 1)$ vector u, we have:

$$u = \left(\iota_T \otimes \mathrm{diag}\left(\iota_{M_i} \right) \right) \alpha + \left(\iota_T \otimes I_S \right) \mu + v. \tag{9}$$

In order to compute the GM-S-IV estimator of δ, which is described in Sect. 3.2, we need to obtain the inverse of the covariance matrix of u, which is Ω_u^{-1}. This is achieved by means of the spectral decomposition. Following Baltagi et al. (2014), the covariance matrix of u_t is:

$$E\left[u_t u_t^\top \right] = \sigma_\alpha^2 \mathrm{diag}\left(J_{M_i} \right) + \left(\sigma_\mu^2 + \sigma_v^2 \right) I_S, \tag{10}$$

where $I_S \left(= \mathrm{diag}\left(I_{M_i} \right) \right)$ is an identity matrix of dimension S. $J_{M_i} = \left(\iota_{M_i} \iota_{M_i}^\top \right)$ is a matrix of ones of dimension $(M_i \times M_i)$. The covariance matrix of u corresponds to:

$$\begin{aligned}
\Omega_u &= \sigma_\alpha^2 \left(Z_\alpha Z_\alpha^\top \right) + \sigma_\mu^2 \left(Z_\mu Z_\mu^\top \right) + \sigma_v^2 \left(I_T \otimes I_S \right) \\
&= \sigma_\alpha^2 \left(J_T \otimes \mathrm{diag}\left(J_{M_i} \right) \right) + \left(\sigma_\mu^2 J_T + \sigma_v^2 I_T \right) \otimes I_S,
\end{aligned} \tag{11}$$

where $Z_\alpha = \iota_T \otimes \mathrm{diag}\left(\iota_{M_i} \right)$, $Z_\mu = \iota_T \otimes I_S$ and $J_T = \left(\iota_T \iota_T^\top \right)$ is a matrix of ones of dimension $(T \times T)$. Replace J_T by its idempotent counterpart $T \bar{J}_T$, J_{M_i} by $M_i \bar{J}_{M_i}$ with $\bar{J}_T = J_T / T$ and $\bar{J}_{M_i} = J_{M_i} / M_i$. Also, define $E_T = I_T - \bar{J}_T$, and $E_{M_i} = I_{M_i} - \bar{J}_{M_i}$, and replace I_T by $\left(E_T + \bar{J}_T \right)$, I_{M_i} by $\left(E_{M_i} + \bar{J}_{M_i} \right)$. Collecting terms with the same matrices, one gets the spectral decomposition of Ω_u:

$$\Omega_u = \theta_1 Q_1 + \theta_2 Q_2 + \left(I_T \otimes \mathrm{diag}\left(\theta_{3i} I_{M_i} \right) \right) Q_3, \tag{12}$$

with

$$\theta_1 = \sigma_v^2, \theta_2 = T\sigma_\mu^2 + \sigma_v^2, \theta_{3i} = M_i T \sigma_\alpha^2 + T \sigma_\mu^2 + \sigma_v^2. \tag{13}$$

These equalities occur because of the definitions[1] of Q_1, Q_2 and Q_3. It turns out that Q_1 relates to within transformation. Q_2 and Q_3 relate, respectively, to between and mean transformation matrices. More formally,

$$Q_1 = E_T \otimes I_S, Q_2 = \bar{J}_T \otimes \mathrm{diag}\left(E_{M_i} \right), \tag{14}$$

$$Q_3 = \bar{J}_T \otimes \mathrm{diag}\left(\bar{J}_{M_i} \right). \tag{15}$$

---

[1] To save space, details are available upon request from the authors.

The operators Q_1, Q_2 and Q_3 are symmetric and idempotent, with their rank equal to their trace. Moreover, they are pairwise orthogonal and sum to the identity matrix. From (12), we can easily obtain Ω_u^{-1} as:

$$\Omega_u^{-1} = \theta_1^{-1} Q_1 + \theta_2^{-1} Q_2 + \left(I_T \otimes \text{diag} \left(\theta_{3i}^{-1} I_{M_i} \right) \right) Q_3. \tag{16}$$

For the full $(T S \times 1)$ vector ε, we then have:

$$\varepsilon = u - \lambda(I_T \otimes W_S)u, \tag{17}$$

or

$$\varepsilon = [I_{TS} - \lambda(I_T \otimes W_S)]u = (I_T \otimes G_S)u, \tag{18}$$

where $G_S = I_S - \lambda W_S$. The corresponding $(T S \times T S)$ covariance matrix is given by:

$$\Omega_\varepsilon = A \Omega_u A^\top, \tag{19}$$

where A is a block diagonal matrix equal to $(I_T \otimes G_S)$. Following the properties of the matrices Ω_u and A, we obtain the inverse covariance matrix of ε defined as:

$$\Omega_\varepsilon^{-1} = \left(A' \right)^{-1} \Omega_u^{-1} A^{-1}. \tag{20}$$

3 Estimation methods

The estimation methods of multidimensional spatial panel models are direct extensions of the ones that have been created for the standard spatial panel data econometrics. This means that two main approaches are used to estimate these models: one based on ML principle and the other one linked to method of moments procedures.

3.1 Maximum likelihood estimation

Upton and Fingleton (1985), Anselin (1988), LeSage and Pace (2009) and Elhorst (2014) provide the general framework for ML estimation of spatial models. Under normality of the disturbances, the log-likelihood function is

$$\ln L = - \frac{T S}{2} \ln (2\pi) - \frac{1}{2} \ln |\Omega_\varepsilon| + T \ln |D_S|$$
$$- \frac{1}{2} (Dy - X\beta)' \Omega_\varepsilon^{-1} (Dy - X\beta), \tag{21}$$

where $D_S = (I_S - \rho W_S)$ and $D = (I_T \otimes D_S)$. For a SMA process for the disturbances ε, and after some mathematical manipulations, we obtain

$$\ln L = -\frac{TS}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega_u| - T \ln |G_S| + T \ln |D_S|$$
$$- \frac{1}{2} (Dy - X\beta)' \Omega_\varepsilon^{-1} (Dy - X\beta). \tag{22}$$

Let $\gamma_1 = \sigma_\alpha^2/\sigma_v^2$, $\gamma_2 = \sigma_\mu^2/\sigma_v^2$ and $\Omega_\varepsilon = \sigma_v^2 \Sigma$, then the log-likelihood function (22) can be written as [2]

$$\ln L = -\frac{TS}{2} \ln(2\pi) - \frac{TS}{2} \ln \sigma_v^2 - \frac{1}{2} \sum_{i=1}^{N} \ln(T(M_i \gamma_1 + \gamma_2) + 1)$$
$$- \frac{1}{2} \sum_{i=1}^{N} (M_i - 1) \ln(T\gamma_2 + 1)$$
$$- T \sum_{i=1}^{N} \sum_{j=1}^{M_i} \ln(1 - \omega_{ij}\lambda) + T \sum_{i=1}^{N} \sum_{j=1}^{M_i} \ln(1 - \eta_{ij}\rho)$$
$$- \frac{1}{2\sigma_v^2} (Dy - X\beta)' \Sigma^{-1} (Dy - X\beta). \tag{23}$$

The first-order conditions for the parameters in (22) and (23) are intertwined which means that they are nonlinear, i.e. the equations cannot be solved analytically. Therefore, a numerical solution by means of an iterative procedure is needed in the spirit of Anselin (1988).

3.2 GM and instrumental variables

There are several issues with ML procedures. First, they call for explicit distributional assumptions, which may be difficult to satisfy, although quasi-ML (QML) approaches may to some extent allay this problem. Second, specifying and maximizing likelihood functions appropriate to extensions to more complex models may be problematic, especially if there are endogenous variables other than the spatial lag, as ML estimation is not possible when endogeneity is in implicit form. Finally, there are very computationally intensive. In view of the desirability of estimation approaches that avoid some of these challenges posed by ML, Kelejian and Prucha (1998, 1999) suggested an alternative instrumental variable estimation procedure for the cross-sectional spatial lag model also including a SAR process for the disturbances. This approach is based on a GM estimator of the parameter in the SAR process. The procedures suggested in Kelejian and Prucha (1998, 1999) are computationally feasible even for very large sample sizes. In a panel data context with a spatial error autoregressive process, KKP (2007) derive a GM estimator, which is computationally feasible even for large sample size, while Fingleton et al. (2016) extend this procedure to capture spatial autoregres-

---

[2] See Baltagi et al. (2001) who give the expression of the log-likelihood function for the multidimensional panel data model with nested random effects errors but without considering any spatial effects.

sive nested random effects errors. We follow this and adapt the moments conditions in order to consider SMA nested random effects errors.

3.2.1 Moments conditions

We follow Fingleton et al. (2016) to develop a GM approach leading to estimators of λ, σ_α^2, σ_μ^2, σ_v^2, or equivalently of λ, σ_α^2, θ_2 $\left(= T\sigma_\mu^2 + \sigma_v^2\right)$ and σ_v^2, relies on moments conditions related to $E[u^\top Q_i u]$, $E[\bar{u}^\top Q_i u]$, $E[\bar{\bar{u}}^\top Q_i u]$, $E[u^\top Q_i \bar{u}]$, $E[\bar{\bar{u}}^\top Q_i \bar{u}]$, $E[\bar{\bar{u}}^\top Q_i \bar{\bar{u}}]$, $i = 1, 2, 3$. For notational convenience, we have

$$\bar{\varepsilon} = (I_T \otimes W_S)\,\varepsilon, \tag{24}$$

$$\bar{\bar{\varepsilon}} = (I_T \otimes W_S)\,\bar{\varepsilon}, \tag{25}$$

$$\bar{u} = (I_T \otimes W_S)\,u, \tag{26}$$

$$\bar{\bar{u}} = (I_T \otimes W_S)\,\bar{u}. \tag{27}$$

Following (17), we have

$$\varepsilon = u - \lambda\bar{u}, \tag{28}$$

$$\bar{\varepsilon} = \bar{u} - \lambda\bar{\bar{u}}. \tag{29}$$

First, we compute the quadratic moments with respect to Q_1:

$$\begin{aligned}\varepsilon^\top Q_1 \varepsilon &= (u - \lambda\bar{u})^\top Q_1 (u - \lambda\bar{u}) \\ &= u^\top Q_1 u + \lambda^2 \bar{u}^\top Q_1 \bar{u} - 2\lambda\bar{u}^\top Q_1 u,\end{aligned} \tag{30}$$

$$\begin{aligned}\bar{\varepsilon}^\top Q_1 \varepsilon &= (\bar{u} - \lambda\bar{\bar{u}})^\top Q_1 (u - \lambda\bar{u}) \\ &= \bar{u}^\top Q_1 u + \lambda^2 \bar{\bar{u}}^\top Q_1 \bar{u} - \lambda\left[\bar{u}^\top Q_1 \bar{u} + \bar{\bar{u}}^\top Q_1 u\right],\end{aligned} \tag{31}$$

$$\begin{aligned}\bar{\varepsilon}^\top Q_1 \bar{\varepsilon} &= (\bar{u} - \lambda\bar{\bar{u}})^\top Q_1 (\bar{u} - \lambda\bar{\bar{u}}) \\ &= \bar{u}^\top Q_1 \bar{u} + \lambda^2 \bar{\bar{u}}^\top Q_1 \bar{\bar{u}} - 2\lambda\bar{\bar{u}}^\top Q_1 \bar{u}.\end{aligned} \tag{32}$$

Then, the expectations of the quadratic moments (30), (31), (32) depend on the moments $E[u^\top Q_1 u]$, $E[\bar{u}^\top Q_1 u]$, $E[\bar{u}^\top Q_1 \bar{u}]$, $E[\bar{\bar{u}}^\top Q_1 \bar{u}]$, $E[\bar{\bar{u}}^\top Q_i \bar{\bar{u}}]$, $E[\bar{\bar{u}}^\top Q_1 u]$. After some computations,[3] these latter expectations are given by:

$$E[u^\top Q_1 u] = \sigma_v^2 S\,(T - 1), \tag{33}$$

$$E[\bar{u}^\top Q_1 u] = 0, \tag{34}$$

$$E[\bar{u}^\top Q_1 \bar{u}] = \sigma_v^2\,(T - 1)\,\mathrm{tr}\left(W_S^\top W_S\right), \tag{35}$$

$$E[\bar{\bar{u}}^\top Q_1 \bar{u}] = \sigma_v^2\,(T - 1)\,\mathrm{tr}\left(W_S^\top W_S^\top W_S\right), \tag{36}$$

$$E[\bar{\bar{u}}^\top Q_1 \bar{\bar{u}}] = \sigma_v^2\,(T - 1)\,\mathrm{tr}\left(W_S^\top W_S^\top W_S W_S\right), \tag{37}$$

---

[3] To save space, all the proofs in the following are available upon request from the authors.

$$E[\overline{\overline{u}}^{\top} Q_1 u] = \sigma_v^2 (T - 1) \operatorname{tr} \left(W_S^{\top} W_S^{\top} \right). \tag{38}$$

Substituting (33) to (38) into (30), (31) and (32) gives:

$$E[\varepsilon^{\top} Q_1 \varepsilon] = \sigma_v^2 S (T - 1) \\ + \lambda^2 \sigma_v^2 (T - 1) \operatorname{tr} \left(W_S^{\top} W_S \right), \tag{39}$$

$$E[\overline{\varepsilon}^{\top} Q_1 \varepsilon] = \lambda^2 \sigma_v^2 (T - 1) \operatorname{tr} \left(W_S^{\top} W_S^{\top} W_S \right) \\ - \lambda \sigma_v^2 (T - 1) \operatorname{tr} \left(W_S^{\top} W_S + W_S^{\top} W_S^{\top} \right), \tag{40}$$

$$E[\overline{\varepsilon}^{\top} Q_1 \overline{\varepsilon}] = \sigma_v^2 (T - 1) \operatorname{tr} \left(W_S^{\top} W_S \right) \\ + \lambda^2 \sigma_v^2 (T - 1) \operatorname{tr} \left(W_S^{\top} W_S^{\top} W_S W_S \right) \\ - \lambda \sigma_v^2 2 (T - 1) \operatorname{tr} \left(W_S^{\top} W_S^{\top} W_S \right). \tag{41}$$

We proceed in a similar fashion as a result of replacing Q_1 in (30), (31) and (32) by Q_2 and by Q_3. The moments of $E[u^{\top} Q_i u]$, $E[\overline{u}^{\top} Q_i u]$, $E[\overline{u}^{\top} Q_i \overline{u}]$, $E[\overline{\overline{u}}^{\top} Q_i \overline{u}]$, $E[\overline{\overline{u}}^{\top} Q_i \overline{\overline{u}}]$, $E[\overline{\overline{u}}^{\top} Q_i u]$, $i = 2, 3$, are:

$$E[u^{\top} Q_2 u] = \theta_2 (S - N), \tag{42}$$

$$E[\overline{u}^{\top} Q_2 u] = \theta_2 \operatorname{tr} \left(W_S^{\bullet \top} \right), \tag{43}$$

$$E[\overline{u}^{\top} Q_2 \overline{u}] = \theta_2 \operatorname{tr} \left(W_S^{\bullet \top} W_S^{\bullet} \right) + T \sigma_\alpha^2 \operatorname{tr} \left(\Gamma W_S^{\bullet \top} W_S^{\bullet} \right), \tag{44}$$

$$E[\overline{\overline{u}}^{\top} Q_2 \overline{u}] = \theta_2 \operatorname{tr} \left(W_S^{\bullet \bullet \top} W_S^{\bullet} \right) + T \sigma_\alpha^2 \operatorname{tr} \left(\Gamma W_S^{\bullet \bullet \top} W_S^{\bullet} \right), \tag{45}$$

$$E[\overline{\overline{u}}^{\top} Q_2 \overline{\overline{u}}] = \theta_2 \operatorname{tr} \left(W_S^{\bullet \bullet \top} W_S^{\bullet \bullet} \right) + T \sigma_\alpha^2 \operatorname{tr} \left(\Gamma W_S^{\bullet \bullet \top} W_S^{\bullet \bullet} \right), \tag{46}$$

$$E[\overline{\overline{u}}^{\top} Q_2 u] = \theta_2 \operatorname{tr} \left(W_S^{\bullet \bullet \top} \right), \tag{47}$$

and

$$E[u^{\top} Q_3 u] = N \theta_2 + S T \sigma_\alpha^2, \tag{48}$$

$$E[\overline{u}^{\top} Q_3 u] = \theta_2 \operatorname{tr} \left(W_S^{* \top} \right) + T \sigma_\alpha^2 \operatorname{tr} \left(\Gamma W_S^{\top} \right), \tag{49}$$

$$E[\overline{u}^{\top} Q_3 \overline{u}] = \theta_2 \operatorname{tr} \left(W_S^{* \top} W_S^{*} \right) + T \sigma_\alpha^2 \operatorname{tr} \left(\Gamma W_S^{* \top} W_S^{*} \right), \tag{50}$$

$$E[\overline{\overline{u}}^{\top} Q_3 \overline{u}] = \theta_2 \operatorname{tr} \left(W_S^{** \top} W_S^{*} \right) + T \sigma_\alpha^2 \operatorname{tr} \left(\Gamma W_S^{** \top} W_S^{*} \right), \tag{51}$$

$$E[\overline{\overline{u}}^{\top} Q_3 \overline{\overline{u}}] = \theta_2 \operatorname{tr} \left(W_S^{** \top} W_S^{**} \right) + T \sigma_\alpha^2 \operatorname{tr} \left(\Gamma W_S^{** \top} W_S^{**} \right), \tag{52}$$

$$E[\overline{\overline{u}}^{\top} Q_3 u] = \theta_2 \operatorname{tr} \left(W_S^{** \top} \right). \tag{53}$$

Then, the use of these moments leads to:

$$E[\varepsilon^\top Q_2 \varepsilon] = \theta_2 \left(S - N\right)$$
$$+ \lambda^2 \left[\theta_2 \text{tr}\left(W_S^{\bullet\top} W_S^\bullet\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{\bullet\top} W_S^\bullet\right)\right]$$
$$- \lambda\theta_2 2\text{tr}\left(W_S^{\bullet\top}\right), \tag{54}$$

$$E[\bar{\varepsilon}^\top Q_2 \varepsilon] = \theta_2 \text{tr}\left(W_S^{\bullet\top}\right)$$
$$+ \lambda^2 \left[\theta_2 \text{tr}\left(W_S^{\bullet\bullet\top} W_S^\bullet\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{\bullet\bullet\top} W_S^\bullet\right)\right]$$
$$- \lambda \left[\theta_2 \text{tr}\left(W_S^{\bullet\top} W_S^\bullet + W_S^{\bullet\bullet\top}\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{\bullet\top} W_S^\bullet\right)\right], \tag{55}$$

$$E[\bar{\varepsilon}^\top Q_2 \bar{\varepsilon}] = \theta_2 \text{tr}\left(W_S^{\bullet\top} W_S^\bullet\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{\bullet\top} W_S^\bullet\right)$$
$$+ \lambda^2 \left[\theta_2 \text{tr}\left(W_S^{\bullet\bullet\top} W_S^{\bullet\bullet}\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{\bullet\bullet\top} W_S^{\bullet\bullet}\right)\right]$$
$$- \lambda \left[\theta_2 2\text{tr}\left(W_S^{\bullet\bullet\top} W_S^\bullet\right) + T\sigma_\alpha^2 2\text{tr}\left(\Gamma W_S^{\bullet\bullet\top} W_S^\bullet\right)\right], \tag{56}$$

where $\Gamma = \text{diag}(J_{M_i})$, $W_S^\bullet = \text{diag}(E_{M_i})W_S$, $W_S^{\bullet\bullet} = \text{diag}(E_{M_i})W_S W_S$ and

$$E\left[\varepsilon^\top Q_3 \varepsilon\right] = \theta_2 N + \sigma_\alpha^2 ST$$
$$+ \lambda^2 \left[\theta_2 \text{tr}\left(W_S^{*\top} W_S^*\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{*\top} W_S^*\right)\right]$$
$$- \lambda \left[\theta_2 2\text{tr}\left(W_S^{*\top}\right) + T\sigma_\alpha^2 2\text{tr}\left(\Gamma W_S^\top\right)\right], \tag{57}$$

$$E[\bar{\varepsilon}^\top Q_3 \varepsilon] = \theta_2 \text{tr}\left(W_S^{*\top}\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^\top\right)$$
$$+ \lambda^2 \left[\theta_2 \text{tr}\left(W_S^{**\top} W_S^*\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{**\top} W_S^*\right)\right]$$
$$- \lambda \left[\theta_2 \text{tr}\left(W_S^{*\top} W_S^* + W_S^{**\top}\right)\right. \tag{58}$$
$$\left. + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{*\top} W_S^* + \Gamma W_S^\top W_S^\top\right)\right], \tag{59}$$

$$E[\bar{\varepsilon}^\top Q_3 \bar{\varepsilon}] = \theta_2 \text{tr}\left(W_S^{*\top} W_S^*\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{*\top} W_S^*\right)$$
$$+ \lambda^2 \left[\theta_2 \text{tr}\left(W_S^{**\top} W_S^{**}\right) + T\sigma_\alpha^2 \text{tr}\left(\Gamma W_S^{**\top} W_S^{**}\right)\right]$$
$$- \lambda \left[\theta_2 2\text{tr}\left(W_S^{**\top} W_S^*\right) + T\sigma_\alpha^2 2\text{tr}\left(\Gamma W_S^{**\top} W_S^*\right)\right], \tag{60}$$

where $W_S^* = \text{diag}(\bar{J}_{M_i})W_S$ and $W_S^{**} = \text{diag}(\bar{J}_{M_i})W_S W_S$. Overall, we obtain a system of nine equations involving the second moments of ε, $\bar{\varepsilon}$:

$$\Lambda\Upsilon - \gamma = 0, \tag{61}$$

where

$$\Lambda = \begin{pmatrix}
\kappa_1\kappa_2 & \kappa_2 t_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \kappa_2 t_2 & \kappa_2 t_3 & 0 & 0 & 0 & 0 & 0 & 0 \\
\kappa_2 t_1 & \kappa_2 t_4 & 2\kappa_2 t_2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \kappa_3 & t_1^\bullet & 2t_6^\bullet & 0 & \kappa_4 t_5^\bullet & 0 \\
0 & 0 & 0 & t_6^\bullet & t_7 & t_8^\bullet & 0 & \kappa_4 t_9 & \kappa_4 t_5^\bullet \\
0 & 0 & 0 & t_1^\bullet & t_1^{\bullet\bullet} & 2t_7 & \kappa_4 t_5^\bullet & \kappa_4 t_5^{\bullet\bullet} & 2\kappa_4 t_9^\bullet \\
0 & 0 & 0 & \kappa_5 & t_1^* & 2t_6^* & \kappa_1\kappa_4 & \kappa_4 t_5^* & 2\kappa_4 t_{10} \\
0 & 0 & 0 & t_6^* & t_{11} & t_8^* & \kappa_4 t_{10} & \kappa_4 t_9^* & \kappa_4 t_{12} \\
0 & 0 & 0 & t_1^* & t_1^{**} & 2t_{11} & \kappa_4 t_5^* & \kappa_4 t_5^{**} & 2\kappa_4 t_9^*
\end{pmatrix}, \qquad (62)$$

and

$$\Upsilon = \begin{pmatrix}
\sigma_v^2 \\
\lambda^2\sigma_v^2 \\
-\lambda\sigma_v^2 \\
\theta_2 \\
\lambda^2\theta_2 \\
-\lambda\theta_2 \\
\sigma_\alpha^2 \\
\lambda^2\sigma_\alpha^2 \\
-\lambda\sigma_\alpha^2
\end{pmatrix}, \gamma = \begin{pmatrix}
E\left[\varepsilon^\top Q_1 \varepsilon\right] \\
E\left[\bar\varepsilon^\top Q_1 \varepsilon\right] \\
E\left[\bar\varepsilon^\top Q_1 \bar\varepsilon\right] \\
E\left[\varepsilon^\top Q_2 \varepsilon\right] \\
E\left[\bar\varepsilon^\top Q_2 \varepsilon\right] \\
E\left[\bar\varepsilon^\top Q_2 \bar\varepsilon\right] \\
E\left[\varepsilon^\top Q_3 \varepsilon\right] \\
E\left[\bar\varepsilon^\top Q_3 \varepsilon\right] \\
E\left[\bar\varepsilon^\top Q_3 \bar\varepsilon\right]
\end{pmatrix}, \qquad (63)$$

with $\kappa_1 = S$, $\kappa_2 = T - 1$, $\kappa_3 = S - N$, $\kappa_4 = T$, $\kappa_5 = N$, $t_1 = \operatorname{tr}\left(W_S^\top W_S\right)$, $t_2 = \operatorname{tr}\left(W_S^\top W_S^\top W_S\right)$, $t_3 = \operatorname{tr}\left(W_S^\top W_S + W_S^\top W_S^\top\right)$, $t_4 = \operatorname{tr}\left(W_S^\top W_S^\top W_S W_S\right)$, $t_1^\bullet = \operatorname{tr}\left(W_S^{\bullet\top} W_S^\bullet\right)$, $t_1^* = \operatorname{tr}\left(W_S^{*\top} W_S^*\right)$, $t_1^{\bullet\bullet} = \operatorname{tr}\left(W_S^{\bullet\bullet\top} W_S^{\bullet\bullet}\right)$, $t_1^{**} = \operatorname{tr}\left(W_S^{**\top} W_S^{**}\right)$, $t_5^\bullet = \operatorname{tr}\left(\Gamma W_S^{\bullet\top} W_S^\bullet\right)$, $t_5^* = \operatorname{tr}\left(\Gamma W_S^{*\top} W_S^*\right)$, $t_5^{\bullet\bullet} = \operatorname{tr}\left(\Gamma W_S^{\bullet\bullet\top} W_S^{\bullet\bullet}\right)$, $t_5^{**} = \operatorname{tr}\left(\Gamma W_S^{**\top} W_S^{**}\right)$, $t_6^\bullet = \operatorname{tr}\left(W_S^{\bullet\top}\right)$, $t_6^* = \operatorname{tr}\left(W_S^{*\top}\right)$, $t_7 = \operatorname{tr}\left(W_S^{\bullet\bullet\top} W_S^\bullet\right)$, $t_8^\bullet = \operatorname{tr}\left(W_S^{\bullet\top} W_S^\bullet + W_S^{\bullet\bullet\top}\right)$, $t_8^* = \operatorname{tr}\left(W_S^{*\top} W_S^* + W_S^{**\top}\right)$, $t_9^\bullet = \operatorname{tr}\left(\Gamma W_S^{\bullet\bullet\top} W_S^\bullet\right)$, $t_9^* = \operatorname{tr}\left(\Gamma W_S^{**\top} W_S^*\right)$, $t_{10} = \operatorname{tr}\left(\Gamma W_S^\top\right)$, $t_{11} = \operatorname{tr}\left(W_S^{**\top} W_S^*\right)$ and $t_{12} = t_5^* + \operatorname{tr}\left(\Gamma W_S^\top W_S^\top\right)$.

In practice, to obtain the GM estimators of λ, σ_α^2, θ_2 and σ_v^2, we have to use the sample counterparts of the terms in Eq. (61), i.e. $\widetilde\Lambda$ and $\widetilde\gamma$. Nevertheless, to estimate λ and σ_v^2, it is possible to use only the moments from (33) to (38). Then, the estimates of θ_2 and σ_α^2 follow from the moments (42) and (48). This estimator is called the unweighted GM estimator. In other words, the GM estimators of λ and σ_v^2 are obtained from the reduced system

$$\Lambda^\bullet \Upsilon^\bullet - \gamma^\bullet = 0, \qquad (64)$$

where

$$\Lambda^\bullet = \begin{pmatrix}
\kappa_1\kappa_2 & \kappa_2 t_1 & 0 \\
0 & \kappa_2 t_2 & \kappa_2 t_3 \\
\kappa_2 t_1 & \kappa_2 t_4 & 2\kappa_2 t_2
\end{pmatrix}, \qquad (65)$$

and

$$\Upsilon^{\bullet} = \begin{pmatrix} \sigma_v^2 \\ \lambda^2 \sigma_v^2 \\ -\lambda \sigma_v^2 \end{pmatrix}, \gamma^{\bullet} = \begin{pmatrix} E\left[\varepsilon^{\top} Q_1 \varepsilon\right] \\ E\left[\bar{\varepsilon}^{\top} Q_1 \varepsilon\right] \\ E\left[\bar{\varepsilon}^{\top} Q_1 \bar{\varepsilon}\right] \end{pmatrix}. \tag{66}$$

Then, from Eqs. (42) and (48) we obtain, respectively:

$$\hat{\theta}_2 = \frac{1}{(S-N)} \left(\hat{G}^{-1}\hat{\varepsilon}\right)^{\top} Q_2 \hat{G}^{-1}\hat{\varepsilon}, \tag{67}$$

and

$$\hat{\sigma}_{\alpha}^2 = \frac{1}{ST} \left(\hat{G}^{-1}\hat{\varepsilon}\right)^{\top} Q_3 \hat{G}^{-1}\hat{\varepsilon} - \frac{N}{ST}\hat{\theta}_2, \tag{68}$$

where $\hat{G}^{-1} = \left[I_T \otimes \hat{G}_S^{-1}\right]$ with $\hat{G}_S = I_S - \hat{\lambda} W_S$.

The above approach relates to unweighted GM. Nevertheless, the literature on generalized method of moments estimators indicates that it is optimal to use the inverse of the variance–covariance matrix of the sample moments at the true parameter values as a weighting matrix in order to obtain asymptotic efficiency. In the following, our Monte Carlo simulations show that our results are not very different from those produced by ML, especially when our three-stage procedure is iterated. We therefore leave the weighted GM method for further research.

3.2.2 The GM spatial IV estimator

To obtain the GM-S-IV estimator of $\delta \left(= [\rho, \beta]^{\top}\right)$, one first calculates the unweighted GM estimates of $\lambda, \sigma_v^2, \theta_2$ and σ_{α}^2, following a three-stage procedure:

- in the first stage, the model (7) is estimated using an IV approach based on the matrix of instruments H which is given by $\left(X, WX, W^2X, \dots\right)$. Thus, the IV estimator of δ is defined as:

$$\hat{\delta}_{IV} = \left(Z^{\top} P_H Z\right)^{-1} Z^{\top} P_H y, \tag{69}$$

where $P_H = H \left(H^{\top} H\right)^{-1} H^{\top}$;
- in the second stage, the parameters $\lambda, \sigma_v^2, \theta_2$ and σ_{α}^2 are estimated using the GM approach from Sect. 3.2 based on IV residuals, i.e. $\hat{\varepsilon} = y - Z\hat{\delta}_{IV}$. The GM estimates are obtained from the sample counterpart of the reduced system (64) which is:

$$\tilde{\Lambda}^{\bullet} \Upsilon^{\bullet} - \tilde{\gamma}^{\bullet} = \xi\left(\lambda, \sigma_v^2\right), \tag{70}$$

where $\xi\left(\lambda, \sigma_v^2\right)$ is a vector of residuals. The unweighted GM estimators of λ and σ_v^2 are the nonlinear least squares estimators based on (70):

$$\left(\widehat{\lambda}, \widehat{\sigma}_v^2\right) = \arg\min\left\{\xi\left(\lambda, \sigma_v^2\right)^\top \xi\left(\lambda, \sigma_v^2\right)\right\}. \tag{71}$$

Then, the estimated parameters of θ_2 and σ_α^2 are obtained using, respectively (67) and (68);

- in the third stage, we need the estimated variance–covariance matrix $\widehat{\Omega}_u$ obtained using the first stage estimates of $\widehat{\sigma}_v^2$, $\widehat{\sigma}_\mu^2$, $\widehat{\sigma}_\alpha^2$. In order to obtain an equation in terms of u, from which spatial autocorrelation is absent, rather than in terms of ε in which it is present, we can purge the equation of spatial dependence by pre-multiplication by $\hat{G}^{-1}$. This can be seen to be a type of Cochrane–Orcutt transformation appropriate to spatially dependent data. Hence, pre-multiplication of the model (7) by $\hat{G}^{-1}$ yields:

$$y^* = Z^*\delta + u, \tag{72}$$

where $Z^* = \hat{G}^{-1}Z$, $y^* = \hat{G}^{-1}y$. If we are guided by the classical panel data random effects literature (see Baltagi 2013), and transform the model in (72) by pre-multiplying it by $\widehat{\Omega}_u^{-1/2}$, then applying the IV principle gives the GM-S-IV estimator $\hat{\delta}_{GM-S-IV}$ which corresponds to:

$$\widehat{\delta}_{GM-S-IV} = \left(Z^{**\top} P_{H^{**}} Z^{**}\right)^{-1} Z^{**\top} P_{H^{**}} y^{**}, \tag{73}$$

where $Z^{**} = \widehat{\Omega}_u^{-1/2} Z^*$, $y^{**} = \widehat{\Omega}_u^{-1/2} y^*$, $H^{**} = \widehat{\Omega}_u^{-1/2} H^*$, $H^* = \hat{G}^{-1} H$, $P_{H^{**}} = H^{**}\left(H^{**\top} H^{**}\right)^{-1} H^{**\top}$.

This three-stage procedure can be iterated. After the first iteration, i.e. the application of the procedure describes above, the GM-S-IV residuals are computed. Then, they are used to compute new sets of unweighted GM estimates. Last, these latter are used to obtain new GM-S-IV parameter estimates of the multidimensional spatial lag model and so on.

4 A Monte Carlo study

The idea here is to demonstrate the comparative performance of the various estimators described thus far, namely ML and the GM-S-IV approach. For this purpose, we generate data using a model with known parameters and see how accurately the different estimators recover the true parameter values. Our data generating process is the spatial lag regression model:

$$y_t = D_S^{-1}\left[\beta_0 \iota_t + \beta_1 x_t + \varepsilon_t\right] \qquad t = 1, \ldots, T, \tag{74}$$

where y_t is of dimension $(S \times 1)$ as is the exogenous variable x_t. Likewise, ι_t is a vector of ones of dimension $(S \times 1)$. $D_S = I_S - \rho W_S$ where W_S is the spatial matrix of size $(S \times S)$. We retain the spatial structure proposed by Kelejian and Prucha (1999), which are referred to as "J ahead and J behind", with the nonzero elements equal to $1/2J$. Note that, as J increase, the value of nonzero elements $1/2J$ decreases and this is turn may reduce the amount of spatial correlation. Here, we consider $J = 2, 6$ and 10. The error term ε_t has a SMA structure

$$\varepsilon_t = u_t - \lambda W_S u_t, \tag{75}$$

and u_t has a nested random components structure given by

$$u_t = \text{diag}\left(\iota_{M_i}\right)\alpha + \mu + v_t, \tag{76}$$

where u_t is $(S \times 1)$, α is the vector of group effects of dimension $(N \times 1)$, $\mu^\top = \left(\mu_1^\top, \ldots, \mu_N^\top\right)$, a vector of dimension $(1 \times S)$, $\mu_i^\top = \left(\mu_{i1}, \ldots, \mu_{iM_i}\right)$, a vector of dimension $(1 \times M_i)$, ι_{M_i} is a vector of ones of dimension $(M_i \times 1)$. v_t is of dimension $(S \times 1)$.

Throughout the experiment, the parameters of (74) and (75) were set at $\beta_0 = 5$, $\beta_1 = 2$ and $\rho = 0.3, 0.6$ and $\lambda = -0.2, -0.5, -0.9$, i.e. positive dependence. The explanatory variable x_{ijt} is generated by a similar method to that of Nerlove (1971), Antweiler (2001) and Baltagi et al. (2001). More precisely, we have:

$$x_{ijt} = 0.3t + 0.8x_{ijt-1} + \omega_{ijt}, \tag{77}$$

where $i = 1, \ldots, N$, $j = 1, \ldots, M_i$, and ω_{ijt} is a random variable uniformly distributed on the interval $[-0.5, 0.5]$ and $x_{ij0} = 60 + 30\omega_{ij0}$. Observations over the first 10 periods are discarded to minimize the effect of initial values. For the data generating process for the errors, we assume $\alpha_i \sim iid.N\left(0, \sigma_\alpha^2\right)$, $\mu_{ij} \sim iid.N\left(0, \sigma_\mu^2\right)$ and $v_{ijt} \sim iid.N\left(0, \sigma_v^2\right)$. We fix $\sigma_u^2 = \sigma_\alpha^2 + \sigma_\mu^2 + \sigma_v^2 = 2$ and define $\gamma_1 = \sigma_\alpha^2/\sigma_u^2$ and $\gamma_2 = \sigma_\mu^2/\sigma_u^2$. These two ratios vary over the set $(0.2, 0.4, 0.6)$ such that $(1 - \gamma_1 - \gamma_2)$ is always positive. For all experiments, we have 20 groups observed over 5 periods, hence $(N, T) = (20, 5)$, and we have $S = 100$ individuals, so the sample size (i.e. TS) is fixed at 500. We consider the 3 unbalanced patterns proposed by Fingleton et al. (2016) denoted by P_1, P_2 and P_3, with individuals nested within the N groups with differing frequencies $(M_1, \ldots, M_{20})$. More precisely, considering $N = 20$, P_1 is characterized by $M_i = 5$, $i = 1, \ldots, 20$. P_2, $M_i = 3$, $i = 1, \ldots, 12$, $M_i = 4$, $i = 13, \ldots, 16$ and $M_i = 12$, $i = 17, \ldots, 20$. For P_3, we have $M_i = 2$, $i = 1, \ldots, 8$, $M_i = 3$, $i = 9, \ldots, 12$, $M_i = 4$, $i = 13, \ldots, 18$ and $M_i = 24$, $i = 19, 20$.

For each experiment, we focus on the estimates of the parameters ρ, β_0, β_1, λ, σ_α^2, σ_μ^2 and σ_v^2. Following KKP (2007), we adopt a measure of dispersion which is closely related to the standard measure of RMSE defined as follows:

$$\text{RMSE} = \left[\text{bias}^2 + \left(\frac{IQ}{1.35}\right)^2\right]^{1/2}, \tag{78}$$

where bias corresponds to the difference between the median and the true value of the parameter, while *IQ* is the interquantile range defined as $q_1 - q_2$ where q_1 is the 0.75 quantile and q_2 is the 0.25 quantile.

In the tables below, three sets of RMSE parameters are reported. They are the outcomes of ML, unweighted GM-S-IV and iterated unweighted GM-S-IV estimators. More precisely, $(\hat{\lambda}, \hat{\sigma}_\alpha^2, \hat{\sigma}_\mu^2, \hat{\sigma}_v^2)$ and $(\hat{\lambda}^{(1)}, \hat{\sigma}_\alpha^{2(1)}, \hat{\sigma}_\mu^{2(1)}, \hat{\sigma}_v^{2(1)})$ denote the unweighted and iterated unweighted GM estimates, respectively, whereas $(\hat{\lambda}^{ML}, \hat{\sigma}_\alpha^{2,ML}, \hat{\sigma}_\mu^{2,ML}, \hat{\sigma}_v^{2,ML})$ denote the ML estimates. These estimates are based on IV residuals. Subsequently, the GM-S-IV estimates of ρ, β_0 and β_1 are computed, i.e. $(\hat{\rho}, \hat{\beta}_0, \hat{\beta}_1)$ and $(\hat{\rho}^{(1)}, \hat{\beta}_0^{(1)}, \hat{\beta}_1^{(1)})$, respectively. The ML estimates are denoted $(\hat{\rho}^{ML}, \hat{\beta}_0^{ML}, \hat{\beta}_1^{ML})$. The results of 1, 000 replications for P_1 (balanced subgroups pattern, $M_i = 5, \forall i = 1, \ldots, N$) and $\rho = 0.3$ are given in Table 1, whereas Tables 2 and 3 give results for the unbalanced patterns P_2 and P_3.

From Table 1, it is apparent that while ML is the most efficient for all parameters, the iterated GM-S-IV is almost equally as good for almost all parameters. For example, on average, the RMSE of both GM-S-IV estimators of the spatial autoregressive parameter ρ is approximately only 2% larger than the ML estimate $\hat{\rho}^{ML}$. The differences for β_0, β_1, σ_α^2, σ_μ^2 are also very small between ML and iterated GM-S-IV and never larger than 4%. This is important because the parameters β_0, β_1 are of particular interest in applied economics. It also means that the computational benefits associated with the use of the GM approach do not seem to have much cost in terms of efficiency. For β_0, β_1, σ_α^2, σ_μ^2, the differences between ML and the simple (i.e. not iterated GM-S-IV) are a bit larger (up to 5% for β_1, σ_α^2, σ_μ^2 and 28% for β_0). While iterating the GM-S-IV estimator is likely to achieve marginally more efficient estimates, this is definitely not the case for λ, especially when λ is near the upper end of its range. Indeed, it appears that the RMSE of the iterated GM-S-IV estimator $\hat{\lambda}^{(1)}$ is 32% larger on average than the RMSE of $\hat{\lambda}^{ML}$. Looking in more details, the difference is especially high for $\lambda = -0.9$ (up to 100%), while it remains acceptable for smaller values (in absolute value) of λ (for instance: 17% for $\lambda = -0.2$). Hence, caution is in order as the absolute value of λ tends to unity.

Tables 2 and 3 concern two unbalanced patterns P_2 and P_3. More precisely, the distribution of individuals over the twenty subgroups changes but the sample size remain fixed at $TS = 5 \times 100 = 500$. In Table 2, based on the least unbalanced of the two, the results are qualitatively similar to those of Table 1. In terms of averages, the RMSE of both the simple and iterated GM-S-IV estimator for the spatial autoregressive parameter ρ is approximately 0.73% larger than that produced by ML estimator. The differences for the regression parameters β_0, β_1 are very small (1% for β_0 and even -0.58% for β_1). Conversely, the differences between ML and iterated GM-S-IV for the variances σ_α^2, σ_μ^2 are a bit larger than in the balanced case and up to 35% for σ_α^2. These values are even higher for the simple GM-S-IV estimator, highlighting the less efficient estimates of GM compared with ML estimation, and the need to consider this in relation to the advantages provided by GM. With respect to λ, we find again that the RMSE for the iterated GM-S-IV estimator is larger than under ML, with an average difference of 35% produced by an assumption that the true value of λ is -0.9. With

Table 1 RMSEs of the estimators of ρ, β_0, β_1, λ, σ_α^2, σ_μ^2 and σ_v^2 for pattern P_1 considering $(N, T) = (20, 5)$, $\rho = 0.3$, 1,000 replications

Parameter values			RMSE									
λ	γ_1	γ_2	$\hat\rho^{ML}$	$\hat\beta_0^{ML}$	$\hat\beta_1^{ML}$	$\hat\lambda^{ML}$	$\hat\sigma_\alpha^{2,ML}$	$\hat\sigma_\mu^{2,ML}$	$\hat\sigma_v^{2,ML}$	$\hat\rho$	$\hat\beta_0$	$\hat\beta_1$
$J = 2$												
-0.2	0.2	0.6	0.0122	0.5664	0.0218	0.0699	0.2005	0.1951	0.0263	0.0122	0.5565	0.0217
-0.2	0.4	0.4	0.0117	0.5851	0.0209	0.0688	0.2967	0.1373	0.0267	0.0118	0.5852	0.0207
-0.2	0.6	0.2	0.0113	0.5967	0.0195	0.0671	0.3988	0.0736	0.0271	0.0114	0.6014	0.0196
-0.5	0.2	0.6	0.0129	0.6783	0.0201	0.0681	0.1941	0.1924	0.0266	0.0130	0.6746	0.0204
-0.5	0.4	0.4	0.0129	0.7120	0.0199	0.0692	0.2930	0.1348	0.0264	0.0127	0.7038	0.0195
-0.5	0.6	0.2	0.0124	0.7229	0.0180	0.0679	0.3808	0.0741	0.0265	0.0127	0.7290	0.0183
-0.9	0.2	0.6	0.0145	0.8321	0.0186	0.0548	0.1990	0.1931	0.0281	0.0151	0.8426	0.0185
-0.9	0.4	0.4	0.0146	0.8802	0.0176	0.0563	0.2931	0.1338	0.0277	0.0151	0.8848	0.0178
-0.9	0.6	0.2	0.0140	0.8891	0.0162	0.0553	0.3777	0.0733	0.0279	0.0147	0.8912	0.0166
$J = 6$												
-0.2	0.2	0.6	0.0139	0.6160	0.0235	0.1418	0.2000	0.1990	0.0286	0.0136	0.6090	0.0229
-0.2	0.4	0.4	0.0132	0.6019	0.0220	0.1368	0.2960	0.1374	0.0267	0.0137	0.6077	0.0217
-0.2	0.6	0.2	0.0128	0.6121	0.0205	0.1289	0.3821	0.0735	0.0265	0.0129	0.6102	0.0204
-0.5	0.2	0.6	0.0160	0.7502	0.0222	0.1527	0.1956	0.2007	0.0270	0.0162	0.7560	0.0221
-0.5	0.4	0.4	0.0156	0.7670	0.0215	0.1504	0.2953	0.1364	0.0273	0.0160	0.7519	0.0212

Table 1 continued

	Parameter values		RMSE									
λ	γ_1	γ_2	$\hat{\rho}^{ML}$	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\lambda}^{ML}$	$\hat{\sigma}_\alpha^{2,ML}$	$\hat{\sigma}_\mu^{2,ML}$	$\hat{\sigma}_v^{2,ML}$	$\hat{\rho}$	$\hat{\beta}_0$	$\hat{\beta}_1$
−0.5	0.6	0.2	0.0151	0.7591	0.0197	0.1470	0.3877	0.0744	0.0274	0.0154	0.7604	0.0197
−0.9	0.2	0.6	0.0184	0.9394	0.0219	0.1437	0.1959	0.2009	0.0272	0.0190	0.9448	0.0217
−0.9	0.4	0.4	0.0181	0.9381	0.0208	0.1438	0.2941	0.1344	0.0271	0.0188	0.9394	0.0213
−0.9	0.6	0.2	0.0177	0.9525	0.0188	0.1364	0.3800	0.0743	0.0274	0.0182	0.9556	0.0186
$J = 10$												
−0.2	0.2	0.6	0.0144	0.6166	0.0241	0.1813	0.1998	0.1964	0.0271	0.0144	0.6318	0.0240
−0.2	0.4	0.4	0.0141	0.6221	0.0228	0.1802	0.3027	0.1360	0.0267	0.0142	0.6313	0.0227
−0.2	0.6	0.2	0.0136	0.6413	0.0214	0.1768	0.3852	0.0751	0.0273	0.0137	0.6210	0.0211
−0.5	0.2	0.6	0.0168	0.7603	0.0236	0.2033	0.1968	0.1955	0.0273	0.0171	0.7779	0.0239
−0.5	0.4	0.4	0.0167	0.7544	0.0226	0.2037	0.3060	0.1386	0.0276	0.0169	0.7721	0.0223
−0.5	0.6	0.2	0.0162	0.7841	0.0208	0.1953	0.3782	0.0755	0.0277	0.0163	0.7728	0.0209
−0.9	0.2	0.6	0.0205	0.9681	0.0229	0.1917	0.1912	0.1981	0.0275	0.0206	0.9733	0.0230
−0.9	0.4	0.4	0.0200	0.9668	0.0218	0.1939	0.3019	0.1368	0.0274	0.0200	0.9666	0.0222
−0.9	0.6	0.2	0.0192	0.9448	0.0207	0.1839	0.3869	0.0749	0.0272	0.0193	0.9628	0.0202
Averages			0.0151	0.7577	0.0209	0.1322	0.2929	0.1358	0.0272	0.0154	0.7598	0.0209

Table 1 continued

Parameter values			RMSE										
λ	γ_1	γ_2	$\hat{\lambda}$	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_\mu^2$	$\hat{\sigma}_v^2$	$\hat{\rho}^{(1)}$	$\hat{\beta}_0^{(1)}$	$\hat{\beta}_1^{(1)}$	$\hat{\lambda}^{(1)}$	$\hat{\sigma}_\alpha^{2(1)}$	$\hat{\sigma}_\mu^{2(1)}$	$\hat{\sigma}_v^{2(1)}$
$J = 2$													
−0.2	0.2	0.6	0.1086	0.2126	0.2037	0.0314	0.0123	0.5534	0.0219	0.0817	0.2042	0.2007	0.0268
−0.2	0.4	0.4	0.1179	0.3192	0.1463	0.0325	0.0119	0.5903	0.0208	0.0811	0.3094	0.1371	0.0266
−0.2	0.6	0.2	0.1294	0.4359	0.0830	0.0337	0.0114	0.5946	0.0194	0.0813	0.4032	0.0718	0.0268
−0.5	0.2	0.6	0.1270	0.2128	0.2126	0.0323	0.0132	0.6898	0.0203	0.0970	0.2057	0.2064	0.0273
−0.5	0.4	0.4	0.1417	0.3153	0.1556	0.0337	0.0130	0.7169	0.0196	0.0961	0.3075	0.1375	0.0274
−0.5	0.6	0.2	0.1595	0.4271	0.0953	0.0348	0.0129	0.7327	0.0181	0.0970	0.4048	0.0736	0.0275
−0.9	0.2	0.6	0.1167	0.2009	0.2087	0.0448	0.0153	0.8454	0.0185	0.1221	0.2044	0.2092	0.0332
−0.9	0.4	0.4	0.1193	0.3106	0.1440	0.0511	0.0152	0.8939	0.0176	0.1213	0.2999	0.1436	0.0332
−0.9	0.6	0.2	0.1184	0.4158	0.0774	0.0558	0.0148	0.9092	0.0165	0.1208	0.3981	0.0756	0.0332
$J = 6$													
−0.2	0.2	0.6	0.1974	0.2031	0.2040	0.0294	0.0136	0.6111	0.0233	0.1483	0.2008	0.2026	0.0266
−0.2	0.4	0.4	0.2386	0.3038	0.1474	0.0294	0.0137	0.6145	0.0219	0.1479	0.3093	0.1360	0.0266
−0.2	0.6	0.2	0.2793	0.4120	0.0818	0.0306	0.0128	0.6034	0.0205	0.1486	0.4026	0.0739	0.0266
−0.5	0.2	0.6	0.2407	0.2027	0.2040	0.0300	0.0160	0.7541	0.0223	0.1783	0.1984	0.2051	0.0268
−0.5	0.4	0.4	0.2866	0.3002	0.1486	0.0314	0.0159	0.7598	0.0211	0.1787	0.3115	0.1364	0.0268

Table 1 continued

Parameter values			RMSE										
λ	γ_1	γ_2	$\hat{\lambda}$	$\hat{\sigma}^2_\alpha$	$\hat{\sigma}^2_\mu$	$\hat{\sigma}^2_v$	$\hat{\rho}^{(1)}$	$\hat{\beta}^{(1)}_0$	$\hat{\beta}^{(1)}_1$	$\hat{\lambda}^{(1)}$	$\hat{\sigma}^{2(1)}_\alpha$	$\hat{\sigma}^{2(1)}_\mu$	$\hat{\sigma}^{2(1)}_v$
-0.5	0.6	0.2	0.3316	0.4098	0.0812	0.0342	0.0151	0.7547	0.0197	0.1798	0.4074	0.0742	0.0267
-0.9	0.2	0.6	0.1656	0.1980	0.2020	0.0397	0.0190	0.9554	0.0217	0.2144	0.2007	0.2009	0.0303
-0.9	0.4	0.4	0.1505	0.2997	0.1361	0.0461	0.0187	0.9445	0.0211	0.2142	0.3119	0.1356	0.0303
-0.9	0.6	0.2	0.1361	0.4032	0.0752	0.0543	0.0180	0.9688	0.0188	0.2152	0.4001	0.0726	0.0303
$J = 10$													
-0.2	0.2	0.6	0.2311	0.1986	0.1982	0.0286	0.0145	0.6201	0.0244	0.1977	0.2012	0.2050	0.0272
-0.2	0.4	0.4	0.2677	0.3118	0.1391	0.0292	0.0144	0.6212	0.0228	0.1979	0.3095	0.1389	0.0272
-0.2	0.6	0.2	0.2932	0.4079	0.0765	0.0292	0.0139	0.6339	0.0213	0.1984	0.4102	0.0754	0.0270
-0.5	0.2	0.6	0.2779	0.1995	0.2009	0.0290	0.0171	0.7688	0.0239	0.2483	0.1995	0.2032	0.0269
-0.5	0.4	0.4	0.3178	0.3075	0.1396	0.0292	0.0171	0.7624	0.0226	0.2474	0.3093	0.1379	0.0268
-0.5	0.6	0.2	0.3429	0.4045	0.0787	0.0306	0.0164	0.7823	0.0210	0.2470	0.4054	0.0748	0.0269
-0.9	0.2	0.6	0.2010	0.1991	0.1998	0.0340	0.0206	0.9693	0.0230	0.2893	0.2005	0.2007	0.0292
-0.9	0.4	0.4	0.1922	0.3092	0.1327	0.0373	0.0200	0.9655	0.0220	0.2889	0.3090	0.1397	0.0291
-0.9	0.6	0.2	0.1669	0.3988	0.0727	0.0405	0.0193	0.9669	0.0205	0.2873	0.4000	0.0745	0.0289
Averages			0.2021	0.3081	0.1424	0.0357	0.0154	0.7623	0.0209	0.1750	0.3046	0.1386	0.0282

Table 2 RMSEs of the estimators of ρ, β_0, β_1, λ, σ_α^2, σ_μ^2 and σ_v^2 for pattern P_2 considering $(N, T) = (20, 5)$, $\rho = 0.3$, 1,000 replications

Parameter values			RMSE									
λ	γ_1	γ_2	$\hat\rho\text{ML}$	$\hat\beta_0\text{ML}$	$\hat\beta_1\text{ML}$	$\hat\lambda\text{ML}$	$\hat\sigma_\alpha^{2,ML}$	$\hat\sigma_\mu^{2,ML}$	$\hat\sigma_v^{2,ML}$	$\hat\rho$	$\hat\beta_0$	$\hat\beta_1$
$J = 2$												
−0.2	0.2	0.6	0.0111	0.5386	0.0186	0.0702	0.2259	0.1946	0.0264	0.0113	0.5424	0.0187
−0.2	0.4	0.4	0.0111	0.5611	0.0175	0.0701	0.3384	0.1378	0.0264	0.0111	0.5620	0.0175
−0.2	0.6	0.2	0.0105	0.5646	0.0165	0.0700	0.4106	0.0755	0.0263	0.0106	0.5684	0.0166
−0.5	0.2	0.6	0.0120	0.6586	0.0168	0.0711	0.2258	0.1957	0.0267	0.0121	0.6775	0.0167
−0.5	0.4	0.4	0.0117	0.6854	0.0163	0.0695	0.3241	0.1375	0.0264	0.0119	0.6914	0.0164
−0.5	0.6	0.2	0.0114	0.6848	0.0152	0.0667	0.4177	0.0780	0.0263	0.0118	0.7069	0.0151
−0.9	0.2	0.6	0.0138	0.8382	0.0155	0.0622	0.2395	0.2037	0.0277	0.0134	0.8183	0.0149
−0.9	0.4	0.4	0.0135	0.8183	0.0149	0.0595	0.3178	0.1393	0.0267	0.0133	0.8399	0.0147
−0.9	0.6	0.2	0.0131	0.8424	0.0141	0.0578	0.4061	0.0766	0.0273	0.0134	0.8536	0.0137
$J = 6$												
−0.2	0.2	0.6	0.0119	0.5562	0.0185	0.1406	0.2194	0.1853	0.0267	0.0120	0.5638	0.0185
−0.2	0.4	0.4	0.0120	0.5777	0.0179	0.1377	0.3439	0.1309	0.0271	0.0118	0.5682	0.0182
−0.2	0.6	0.2	0.0119	0.5661	0.0169	0.1293	0.4169	0.0730	0.0271	0.0119	0.5689	0.0169
−0.5	0.2	0.6	0.0137	0.6976	0.0181	0.1597	0.2237	0.1889	0.0273	0.0141	0.6946	0.0180
−0.5	0.4	0.4	0.0135	0.7005	0.0175	0.1544	0.3403	0.1311	0.0272	0.0140	0.7025	0.0176

Table 2 continued

Parameter values			RMSE									
λ	γ_1	γ_2	$\hat{\rho}^{ML}$	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\lambda}^{ML}$	$\hat{\sigma}_\alpha^{2,ML}$	$\hat{\sigma}_\mu^{2,ML}$	$\hat{\sigma}_v^{2,ML}$	$\hat{\rho}$	$\hat{\beta}_0$	$\hat{\beta}_1$
−0.5	0.6	0.2	0.0139	0.6968	0.0158	0.1434	0.4252	0.0741	0.0272	0.0138	0.7042	0.0160
−0.9	0.2	0.6	0.0166	0.8740	0.0172	0.1439	0.2198	0.1887	0.0275	0.0171	0.8518	0.0174
−0.9	0.4	0.4	0.0166	0.8852	0.0168	0.1433	0.3345	0.1321	0.0272	0.0166	0.8796	0.0166
−0.9	0.6	0.2	0.0166	0.8864	0.0148	0.1338	0.4256	0.0719	0.0272	0.0163	0.8896	0.0154
$J = 10$												
−0.2	0.2	0.6	0.0124	0.5857	0.0198	0.1732	0.2279	0.1858	0.0270	0.0125	0.5677	0.0196
−0.2	0.4	0.4	0.0124	0.5687	0.0193	0.1724	0.3396	0.1336	0.0281	0.0124	0.5716	0.0190
−0.2	0.6	0.2	0.0125	0.5768	0.0175	0.1605	0.4249	0.0738	0.0271	0.0121	0.5813	0.0178
−0.5	0.2	0.6	0.0146	0.6982	0.0192	0.2010	0.2292	0.1853	0.0274	0.0147	0.7121	0.0192
−0.5	0.4	0.4	0.0146	0.7110	0.0185	0.1942	0.3453	0.1354	0.0276	0.0145	0.7223	0.0184
−0.5	0.6	0.2	0.0144	0.7341	0.0173	0.1819	0.4256	0.0715	0.0277	0.0144	0.7257	0.0172
−0.9	0.2	0.6	0.0177	0.8861	0.0187	0.1849	0.2253	0.1864	0.0278	0.0178	0.8985	0.0185
−0.9	0.4	0.4	0.0172	0.8901	0.0180	0.1848	0.3385	0.1321	0.0276	0.0175	0.9124	0.0181
−0.9	0.6	0.2	0.0172	0.9025	0.0165	0.1730	0.4296	0.0707	0.0273	0.0175	0.9129	0.0170
Averages			0.0136	0.7106	0.0172	0.1300	0.3274	0.1329	0.0271	0.0137	0.7144	0.0172

Table 2 continued

Parameter values			RMSE										
λ	γ_1	γ_2	$\hat{\lambda}$	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_\mu^2$	$\hat{\sigma}_v^2$	$\hat{\rho}^{(1)}$	$\hat{\beta}_0^{(1)}$	$\hat{\beta}_1^{(1)}$	$\hat{\lambda}^{(1)}$	$\hat{\sigma}_\alpha^{2(1)}$	$\hat{\sigma}_\mu^{2(1)}$	$\hat{\sigma}_v^{2(1)}$
$J=2$													
-0.2	0.2	0.6	0.1051	0.2357	0.1918	0.0297	0.0113	0.5439	0.0184	0.0815	0.2412	0.1978	0.0269
-0.2	0.4	0.4	0.1146	0.3568	0.1410	0.0312	0.0111	0.5692	0.0175	0.0812	0.3737	0.1372	0.0269
-0.2	0.6	0.2	0.1247	0.4925	0.0829	0.0327	0.0105	0.5619	0.0164	0.0105	0.4937	0.0766	0.0268
-0.5	0.2	0.6	0.1258	0.2354	0.2107	0.0297	0.0121	0.6682	0.0165	0.0966	0.2373	0.2021	0.0277
-0.5	0.4	0.4	0.0960	0.3708	0.1399	0.0277	0.0119	0.6923	0.0162	0.1353	0.3605	0.1553	0.0318
-0.5	0.6	0.2	0.1501	0.4944	0.0952	0.0343	0.0118	0.7026	0.0153	0.0952	0.4955	0.0750	0.0277
-0.9	0.2	0.6	0.1219	0.2350	0.2051	0.0336	0.0136	0.8081	0.0148	0.1111	0.2292	0.2092	0.0433
-0.9	0.4	0.4	0.1215	0.3615	0.1400	0.0338	0.0136	0.8444	0.0145	0.1147	0.3689	0.1419	0.0488
-0.9	0.6	0.2	0.1256	0.4982	0.0797	0.0552	0.0133	0.8564	0.0139	0.1197	0.4961	0.0756	0.0337
$J=6$													
-0.2	0.2	0.6	0.2282	0.2229	0.1993	0.0283	0.0121	0.5622	0.0187	0.1530	0.2389	0.1924	0.0268
-0.2	0.4	0.4	0.2779	0.3512	0.1427	0.0296	0.0118	0.5793	0.0180	0.1529	0.3657	0.1352	0.0269
-0.2	0.6	0.2	0.3170	0.4932	0.0816	0.0316	0.0120	0.5691	0.0169	0.1526	0.4989	0.0746	0.0270
-0.5	0.2	0.6	0.2762	0.2209	0.2036	0.0298	0.0137	0.6989	0.0180	0.1871	0.2352	0.1965	0.0270
-0.5	0.4	0.4	0.3398	0.3468	0.1404	0.0328	0.0136	0.7123	0.0173	0.1862	0.3653	0.1354	0.0271

Table 2 continued

Parameter values			RMSE										
λ	γ_1	γ_2	$\hat{\lambda}$	$\hat{\sigma}^2_\alpha$	$\hat{\sigma}^2_\mu$	$\hat{\sigma}^2_v$	$\hat{\rho}^{(1)}$	$\hat{\beta}_0^{(1)}$	$\hat{\beta}_1^{(1)}$	$\hat{\lambda}^{(1)}$	$\hat{\sigma}^{2(1)}_\alpha$	$\hat{\sigma}^{2(1)}_\mu$	$\hat{\sigma}^{2(1)}_v$
−0.5	0.6	0.2	0.3915	0.4818	0.0788	0.0365	0.0140	0.7073	0.0160	0.1860	0.4978	0.0744	0.0271
−0.9	0.2	0.6	0.1604	0.2208	0.1894	0.0430	0.0169	0.8775	0.0175	0.2191	0.2344	0.1962	0.0293
−0.9	0.4	0.4	0.1327	0.3597	0.1321	0.0560	0.0166	0.8925	0.0167	0.2192	0.3648	0.1331	0.0293
−0.9	0.6	0.2	0.1193	0.4855	0.0759	0.0659	0.0166	0.8870	0.0151	0.2188	0.4992	0.0726	0.0294
$J = 10$													
−0.2	0.2	0.6	0.2682	0.2126	0.2035	0.0279	0.0123	0.5763	0.0196	0.2004	0.2406	0.1954	0.0272
−0.2	0.4	0.4	0.3473	0.3446	0.1397	0.0293	0.0123	0.5727	0.0188	0.2009	0.3702	0.1340	0.0271
−0.2	0.6	0.2	0.4088	0.4760	0.0794	0.0299	0.0122	0.5946	0.0178	0.2016	0.5035	0.0731	0.0271
−0.5	0.2	0.6	0.3301	0.2156	0.2051	0.0293	0.0147	0.7228	0.0192	0.2507	0.2376	0.1973	0.0272
−0.5	0.4	0.4	0.4202	0.3412	0.1397	0.0318	0.0146	0.7225	0.0185	0.2507	0.3659	0.1340	0.0271
−0.5	0.6	0.2	0.4515	0.4663	0.0765	0.0341	0.0147	0.7382	0.0173	0.2491	0.5049	0.0728	0.0270
−0.9	0.2	0.6	0.1769	0.2178	0.1950	0.0367	0.0177	0.8980	0.0185	0.2905	0.2380	0.1964	0.0290
−0.9	0.4	0.4	0.1510	0.3477	0.1328	0.0473	0.0175	0.9170	0.0180	0.2904	0.3612	0.1341	0.0290
−0.9	0.6	0.2	0.1328	0.4719	0.0750	0.0547	0.0175	0.9200	0.0167	0.2903	0.4977	0.0724	0.0290
Averages			0.2228	0.3540	0.1399	0.0364	0.0137	0.7183	0.0171	0.1758	0.3673	0.1367	0.0294

Table 3 RMSEs of the estimators of ρ, β_0, β_1, λ, σ_α^2, σ_μ^2 and σ_v^2 for pattern P_3 considering $(N, T) = (20, 5)$, $\rho = 0.3$, 1,000 replications

Parameter values			RMSE									
λ	γ_1	γ_2	$\hat{\rho}^{ML}$	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\lambda}^{ML}$	$\hat{\sigma}_\alpha^{2,ML}$	$\hat{\sigma}_\mu^{2,ML}$	$\hat{\sigma}_v^{2,ML}$	$\hat{\rho}$	$\hat{\beta}_0$	$\hat{\beta}_1$
$J = 2$												
−0.2	0.2	0.6	0.0117	0.5538	0.0200	0.0709	0.2391	0.1925	0.0272	0.0122	0.5454	0.0198
−0.2	0.4	0.4	0.0119	0.5599	0.0189	0.0690	0.3404	0.1339	0.0270	0.0119	0.5627	0.0187
−0.2	0.6	0.2	0.0110	0.5556	0.0173	0.0698	0.4091	0.0721	0.0271	0.0109	0.5664	0.0176
−0.5	0.2	0.6	0.0129	0.6738	0.0175	0.0715	0.2460	0.1896	0.0267	0.0130	0.6695	0.0179
−0.5	0.4	0.4	0.0171	0.6809	0.0171	0.0703	0.3376	0.1366	0.0265	0.0124	0.6815	0.0173
−0.5	0.6	0.2	0.0117	0.6925	0.0161	0.0700	0.4092	0.0722	0.0263	0.0118	0.7026	0.0162
−0.9	0.2	0.6	0.0139	0.8270	0.0164	0.0636	0.2292	0.2023	0.0286	0.0140	0.8482	0.0162
−0.9	0.4	0.4	0.0133	0.8302	0.0157	0.0592	0.3287	0.1388	0.0273	0.0137	0.8611	0.0158
−0.9	0.6	0.2	0.0131	0.8281	0.0145	0.0588	0.4078	0.0737	0.0273	0.0133	0.8648	0.0145
$J = 6$												
−0.2	0.2	0.6	0.0136	0.5834	0.0205	0.1348	0.2414	0.1934	0.0277	0.0136	0.5874	0.0206
−0.2	0.4	0.4	0.0135	0.6009	0.0197	0.1309	0.3420	0.1346	0.0275	0.0135	0.5973	0.0200
−0.2	0.6	0.2	0.0131	0.6093	0.0187	0.1286	0.4089	0.0731	0.0278	0.0132	0.5936	0.0187
−0.5	0.2	0.6	0.0154	0.7249	0.0199	0.1494	0.2453	0.1924	0.0274	0.0158	0.7347	0.0199
−0.5	0.4	0.4	0.0154	0.7405	0.0191	0.1479	0.3469	0.1383	0.0274	0.0156	0.7298	0.0194

Table 3 continued

Parameter values			RMSE									
λ	γ_1	γ_2	$\hat{\rho}^{ML}$	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\lambda}^{ML}$	$\hat{\sigma}_\alpha^{2,ML}$	$\hat{\sigma}_\mu^{2,ML}$	$\hat{\sigma}_v^{2,ML}$	$\hat{\rho}$	$\hat{\beta}_0$	$\hat{\beta}_1$
−0.5	0.6	0.2	0.0151	0.7367	0.0178	0.1427	0.4131	0.0714	0.0273	0.0155	0.7421	0.0181
−0.9	0.2	0.6	0.0182	0.9051	0.0187	0.1380	0.2413	0.1906	0.0279	0.0184	0.8988	0.0191
−0.9	0.4	0.4	0.0181	0.9058	0.0183	0.1393	0.3431	0.1363	0.0277	0.0184	0.9043	0.0189
−0.9	0.6	0.2	0.0179	0.9067	0.0171	0.1328	0.4170	0.0732	0.0280	0.0180	0.9081	0.0174
$J = 10$												
−0.2	0.2	0.6	0.0144	0.6160	0.0215	0.1848	0.2469	0.1906	0.0277	0.0147	0.6265	0.0211
−0.2	0.4	0.4	0.0145	0.6184	0.0202	0.1805	0.3395	0.1363	0.0274	0.0146	0.6394	0.0202
−0.2	0.6	0.2	0.0138	0.6245	0.0189	0.1707	0.4219	0.0727	0.0273	0.0141	0.6432	0.0186
−0.5	0.2	0.6	0.0169	0.7610	0.0207	0.2115	0.2459	0.1910	0.0278	0.0170	0.7541	0.0209
−0.5	0.4	0.4	0.0169	0.7877	0.0196	0.2067	0.3446	0.1350	0.0275	0.0168	0.7929	0.0196
−0.5	0.6	0.2	0.0163	0.7971	0.0187	0.1942	0.4124	0.0722	0.0275	0.0165	0.8043	0.0183
−0.9	0.2	0.6	0.0204	0.9487	0.0200	0.2052	0.2436	0.1936	0.0279	0.0204	0.9378	0.0201
−0.9	0.4	0.4	0.0204	0.9612	0.0194	0.2007	0.3455	0.1335	0.0276	0.0203	0.9946	0.0197
−0.9	0.6	0.2	0.0197	0.9865	0.0180	0.1886	0.4220	0.0716	0.0276	0.0200	1.0134	0.0183
Averages			0.0152	0.7413	0.0185	0.1330	0.3322	0.1338	0.0274	0.0152	0.7483	0.0186

Table 3 continued

Parameter values			RMSE										
λ	γ_1	γ_2	$\hat{\lambda}$	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_\mu^2$	$\hat{\sigma}_v^2$	$\hat{\rho}^{(1)}$	$\hat{\beta}_0^{(1)}$	$\hat{\beta}_1^{(1)}$	$\hat{\lambda}^{(1)}$	$\hat{\sigma}_\alpha^{2(1)}$	$\hat{\sigma}_\mu^{2(1)}$	$\hat{\sigma}_v^{2(1)}$
$J = 2$													
−0.2	0.2	0.6	0.1077	0.2479	0.1960	0.0315	0.0119	0.5447	0.0196	0.0820	0.2620	0.1943	0.0274
−0.2	0.4	0.4	0.1229	0.4006	0.1421	0.0345	0.0119	0.5665	0.0190	0.0819	0.4131	0.1310	0.0273
−0.2	0.6	0.2	0.1378	0.5555	0.0808	0.0365	0.0110	0.5561	0.0176	0.0813	0.5782	0.0707	0.0273
−0.5	0.2	0.6	0.1331	0.2485	0.2030	0.0332	0.0130	0.6749	0.0176	0.0964	0.2624	0.1975	0.0277
−0.5	0.4	0.4	0.1490	0.3982	0.1612	0.0359	0.0125	0.6833	0.0173	0.0964	0.4181	0.1347	0.0276
−0.5	0.6	0.2	0.1643	0.5585	0.0945	0.0373	0.0118	0.6956	0.0162	0.0958	0.5832	0.0727	0.0276
−0.9	0.2	0.6	0.1179	0.2513	0.2056	0.0457	0.0138	0.8498	0.0161	0.1247	0.2665	0.1975	0.0338
−0.9	0.4	0.4	0.1189	0.4029	0.1393	0.0530	0.0136	0.8523	0.0156	0.1247	0.4124	0.1350	0.0337
−0.9	0.6	0.2	0.1242	0.5636	0.0804	0.0602	0.0129	0.8520	0.0145	0.1239	0.5805	0.0735	0.0336
$J = 6$													
−0.2	0.2	0.6	0.2364	0.2473	0.2044	0.0312	0.0137	0.5833	0.0208	0.1528	0.2626	0.2011	0.0274
−0.2	0.4	0.4	0.2976	0.3995	0.1431	0.0340	0.0136	0.6007	0.0199	0.1524	0.4231	0.1366	0.0274
−0.2	0.6	0.2	0.3489	0.5464	0.0819	0.0361	0.0132	0.5966	0.0186	0.1525	0.5994	0.0704	0.0274
−0.5	0.2	0.6	0.2915	0.2448	0.2062	0.0336	0.0157	0.7159	0.0200	0.1831	0.2645	0.2006	0.0272
−0.5	0.4	0.4	0.3647	0.3967	0.1418	0.0376	0.0156	0.7365	0.0194	0.1837	0.4192	0.1375	0.0273

Table 3 continued

Parameter values			RMSE										
λ	γ_1	γ_2	$\hat{\lambda}$	$\hat{\sigma}^2_\alpha$	$\hat{\sigma}^2_\mu$	$\hat{\sigma}^2_v$	$\hat{\rho}^{(1)}$	$\hat{\beta}^{(1)}_0$	$\hat{\beta}^{(1)}_1$	$\hat{\lambda}^{(1)}$	$\hat{\sigma}^{2(1)}_\alpha$	$\hat{\sigma}^{2(1)}_\mu$	$\hat{\sigma}^{2(1)}_v$
−0.5	0.6	0.2	0.4129	0.5429	0.0799	0.0427	0.0154	0.7396	0.0180	0.1837	0.5938	0.0719	0.0273
−0.9	0.2	0.6	0.1467	0.2424	0.1967	0.0488	0.0185	0.9084	0.0188	0.2156	0.2640	0.2022	0.0297
−0.9	0.4	0.4	0.1234	0.3902	0.1346	0.0613	0.0184	0.9129	0.0186	0.2150	0.4180	0.1340	0.0297
−0.9	0.6	0.2	0.1113	0.5531	0.0737	0.0763	0.0181	0.9132	0.0169	0.2153	0.5961	0.0725	0.0297
$J = 10$													
−0.2	0.2	0.6	0.3112	0.2391	0.2004	0.0308	0.0145	0.6175	0.0212	0.2017	0.2727	0.1979	0.0271
−0.2	0.4	0.4	0.3926	0.3866	0.1361	0.0322	0.0143	0.6349	0.0197	0.2014	0.4334	0.1323	0.0271
−0.2	0.6	0.2	0.4822	0.5327	0.0785	0.0356	0.0137	0.6454	0.0189	0.2024	0.5997	0.0730	0.0271
−0.5	0.2	0.6	0.3770	0.2396	0.1977	0.0331	0.0167	0.7539	0.0209	0.2561	0.2722	0.1968	0.0274
−0.5	0.4	0.4	0.4345	0.3847	0.1364	0.0367	0.0168	0.7909	0.0195	0.2554	0.4235	0.1331	0.0275
−0.5	0.6	0.2	0.4527	0.5307	0.0758	0.0415	0.0164	0.7991	0.0185	0.2531	0.5980	0.0723	0.0275
−0.9	0.2	0.6	0.1625	0.2430	0.1975	0.0436	0.0207	0.9427	0.0202	0.2930	0.2751	0.1977	0.0294
−0.9	0.4	0.4	0.1365	0.3815	0.1309	0.0545	0.0204	0.9902	0.0191	0.2945	0.4287	0.1314	0.0294
−0.9	0.6	0.2	0.1145	0.5456	0.0744	0.0642	0.0199	1.0134	0.0179	0.2937	0.6118	0.0726	0.0294
Averages			0.2360	0.3953	0.1405	0.0423	0.0151	0.7470	0.0185	0.1782	0.4271	0.1348	0.0286

smaller λ, the difference is less stark. For example, when $\lambda = -0.2$, the difference is equal to approximately 16%.

In Table 3, the RMSEs are affected differently because of the way we have treated unbalancedness in P_3. Focusing especially on the differences between ML and iterated GM-S-IV, we note that the regression parameters are estimated efficiently in both cases as is ρ. As for the variances, the differences between ML and iterated GM-S-IV are higher for P_3 than was the case P_2 when one consider σ_α^2 (28% for P_3 compared to 12% for P_2). Conversely, the differences are smaller for σ_μ^2 (0.74% for P_3 compared to 2.86% for P_2) and σ_v^2 (4.38% for P_3 compared to 8.49% for P_2). The estimates of the spatial error parameter are affected in a similar way under P_3 as was the case for P_2 : the RMSE of $\hat{\lambda}^{(1)}$ is 33% higher than the RMSE of $\hat{\lambda}^{\text{ML}}$ with especially high differences for $\lambda = 0.9$.

We have also performed the simulations with $\rho = 0.6$ considering the same patterns P_1, P_2 and P_3. The results are provided in an online appendix and the conclusions remain identical to those described above.

5 Empirical application

In this section, we consider the relationship between log employment ($\ln E$), log output ($\ln Q$) and an indicator of (log) capital investment ($\ln K$) across $S = \sum_{i=1}^{N} M_i = 255$ NUTS2 regions nested within $N = 25$ countries of the EU. $\ln Q$ is measured by gross value added, or GVA, and $\ln K$ is gross fixed capital formation, or GFCF. These annual regional data series are based on Cambridge Econometrics' European Regional Economic Data Base. As an illustration, Fig. 1 shows the distribution of log employment in the year 2010 across the 255 regions. Similar maps but with varying regional employment levels covering the period 1999–2010 constitute the dependent variable. Our model endeavours to explain the spatio-temporal variation in $\ln E$ as a function of $\ln Q$ and $\ln K$ organized on the same basis as Fig. 1, and also as an outcome of unobservable region-specific random effects nested within country-specific random effects. Accordingly, the model specification is

$$\ln E_t = \rho W_S \ln E_t + \beta_0 \iota_t + \beta_1 \ln Q_t + \beta_2 \ln K_t + \varepsilon_t, \tag{79}$$

in which $\ln E_t$ is an $(S \times 1)$ vector of levels of (log) employment at time t, with exogenous variables $\ln Q_t$ and $\ln K_t$, and ι_t is a vector of ones of dimension $(S \times 1)$. The compound errors ε_t are an $(S \times 1)$ vector of spatially dependent unobservables comprising time-invariant national effects, one for each of N countries and denoted by $\alpha_i, i = 1, \ldots, N$, together with time-invariant regional effects with the region j effect, where j is nested within country i, denoted by $\mu_{ij} = \mu_k, k = 1, \ldots, S$. In addition, there are remainders of dimension S which vary across regions and time, and the remainder effect for region j within country i at time t is denoted by v_{ijt}. Thus, repeating for convenience Eqs. (5) and (6) in vector and matrix notation, we have
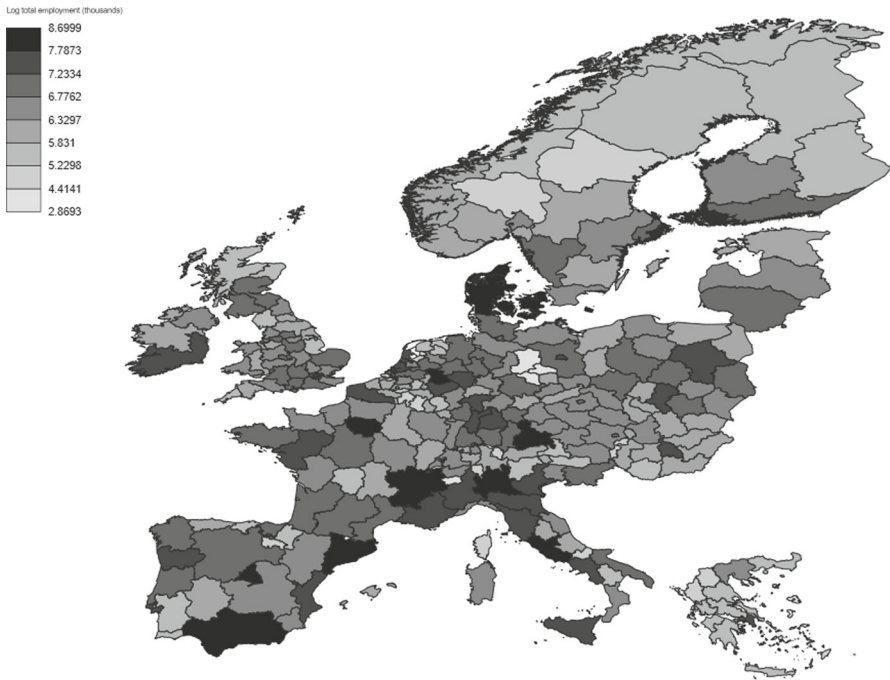
$$\varepsilon_t = u_t - \lambda M_S u_t, \tag{80}$$

Fig. 1 Distribution of log employment in the year 2010

and

$$u_t = \text{diag}\left(\iota_{M_i}\right)\alpha + \mu + v_t. \tag{81}$$

So, the estimation procedure takes into account two different spatial interaction processes: one for the endogenous spatial lag and the other for the errors. For the spatial lag at time t, $W_S \ln E_t$, the matrix W_S is based on interregional trade flows between the 255 EU regions in the year 2000. The method of estimating these trade flows has been discussed elsewhere, for example by Polasek et al. (2010), Vidoli and Mazziotta (2010) and Fingleton et al. (2015), so here we simply note that the method employed bases interregional trade on data for international trade using a spatial version of the method for the construction of quarterly time series from annual series introduced by Chow and Lin (1971). The resulting matrix of bilateral interregional trade flows W_S^* is scaled following the approach of Ord (1975), so that

$$W_S = \widetilde{D}^{-0.5} W_S^* \widetilde{D}^{-0.5}, \tag{82}$$

in which $\widetilde{D}$ is a diagonal matrix with each cell on the leading diagonal containing the corresponding row total from W_S^*. This normalization means that the most positive real eigenvalue of W_S is equal to $\max(\text{eig}) = 1.0$, and the continuous range for which $(I_S - \rho W_S)$ is non-singular is $1/\min(\text{eig}) < \rho < 1$. Thus, we require estimated ρ to occur within this range to ensure stationarity.

For GM-S-IV, the assumption is that the compound errors are interrelated according to an SMA process (80). In this case, the spatial matrix M_S is based on a contiguity matrix of dimension $(S \times S)$ with 1 in cell (m, n) indicating that regions m and n share a common border and 0 indicating otherwise, although for nine isolated regions it has been necessary to create artificial, contiguous neighbours. The resulting contiguity matrix has been subsequently standardized to give M_S in which the rows sum to 1. This means that stationary region for λ is also given by $1/\min(\text{eig}) < \lambda < 1/\max(\text{eig}) = 1$. The key feature of an SMA process is that shocks to the unobservables have local rather than global effects. Note that all components of the compound errors are assumed to be subject to this same spatial error dependence processes.

Table 4 gives the resulting non-iterated estimates. The ratios of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ to their respective standard errors indicate that $\ln Q$ and $\ln K$ are significantly positively related to $\ln E$, and there is also a significant positive effect due to the endogenous spatial lag, since $\rho > 0$ with t ratio equal to 34.2374. The significant positive effect due to the endogenous spatial lag means that we should interpret the effects of these variables via the true derivatives, following LeSage and Pace (2009) and Elhorst (2014). These show that, allowing for both the direct and indirect effects of spatial interaction across regions, the total effect of a 1% change in Q is associated with a 0.3795% change in employment. The total effect of 1% change in K leads to 0.0459% change in employment.

We find that the null hypothesis that $\lambda = 0$ is rejected in favour of positive residual spatial dependence (which is indicated by a negative estimate of λ). The distribution of λ_{null}, which is λ under a null hypothesis of no spatial dependence among the errors, is based on the residuals from the nested error model assuming no spatial error dependence, but which includes a spatial lag (as described in Baltagi et al. (2014)). We refer to this by the acronym NRE-IV. The residuals on which the null distribution is based have the same moments as the NRE-IV residuals, and are assumed to be normally distributed, but they are randomly assigned to regions in order to eliminate spatial dependence. Given randomly assigned residuals, the same GM estimation method used to obtain $\widehat{\lambda}$ is applied to obtain λ_{null}, and this estimation is repeated 100 times to obtain 100 estimates of λ_{null}. We find that the estimate $\widehat{\lambda}$ is not a typical member of this λ_{null} distribution, since

$$t = \frac{\widehat{\lambda} - \overline{\lambda}_{\text{null}}}{\sqrt{\text{var}(\lambda_{\text{null}})}} = \frac{-0.8641 - (-0.0021)}{0.0252} = -34.24, \tag{83}$$

in which $\overline{\lambda}_{\text{null}}$ is the mean of the empirical null distribution, and $\text{var}(\lambda_{\text{null}})$ is the variance.

The estimated variance $\widehat{\sigma}_\alpha^2 = 0.0577$ of unobserved country effects is larger than estimated regional effects variance, which is $\widehat{\sigma}_\mu^2 = 0.0483$, and both of these are large relative to the remainder variance $\widehat{\sigma}_\nu^2 = 0.0008$. In the generation of the λ_{null} distribution, we also generate null distributions of $\sigma_{\alpha,\text{null}}^2$, $\sigma_{\mu,\text{null}}^2$ and $\sigma_{\nu,\text{null}}^2$. Because the null distributions are based on a random pattern of errors, this also breaks up any effects due to country or region. This is evident from the means of the resulting distribution, hence $\overline{\sigma}_{\alpha,\text{null}}^2 = 0.0131$, and $\overline{\sigma}_{\mu,\text{null}}^2 = 0.0118$. In contrast, the remainder null variance is comparatively large, hence $\overline{\sigma}_{\nu,\text{null}}^2 = 0.1719$. Using also the standard

deviations of the null distributions, the resulting t-ratios given in Table 4 indicate that $\widehat{\sigma}_\alpha^2$ and $\widehat{\sigma}_\mu^2$ are significantly greater than expected under the null hypothesis, suggesting significant effects due to countries and regions. In contrast is $\widehat{\sigma}_\nu^2$ significantly below the null mean indicating that unexplained remainder effects varying across regions and across time are much less than one would expect were the errors distributed at random. Note that these interpretations are informal because generally the null distributions may be asymmetrical. The t-ratios give the distance from the mean of the null distributions is units of standard error, and the ratios given in Table 4 are far outside the range of outcomes observed in the null distributions.

For comparison, we also estimate the parameters of three different models which also assume a nested error structure, but which differ in the way spatial dependence is treated. The first estimator is the above-mentioned NRE-IV, which assumes that there is a spatial lag but no spatial error dependence, hence $\lambda = 0$. The second estimator, referred to by the acronym GM-S-FGLS, in line with the published literature (Fingleton et al. 2016), does assume spatial error dependence, but the dependence is an autoregressive process not moving average. Also, it assumes there is no spatial lag effect, hence $\rho = 0$. The third comparator introduces a SAR structure for the disturbances instead of a SMA one, but is otherwise identical to GM-S-IV. We refer to this as GM-S-IV*.

The resulting estimates are given in Table 4. It is evident that on the whole the outcomes with regard to the effects of ln Q and ln K are similar to those produced by GM-S-IV. One difference, however, is that the effect of ln Q under GM-S-FGLS is larger than the effects reported under the other estimators (note that with ρ constrained to zero the estimate $\widehat{\beta}_1$ is directly comparable to the total effects under the other estimators). The introduction of a SAR process under estimator GM-S-IV* moderates the apparent effect of ln Q, which is now quite similar to the outcomes from the other estimators including a spatial lag. Elimination of spatial dependence among the errors, due to the restriction $\lambda = 0$, gives the NRE-IV estimates. The effect of this is to increase $\widehat{\sigma}_\alpha^2$. This suggest that there is more intercountry heterogeneity than under the other estimators. Using the same null reference distribution as for GM-S-IV, $\widehat{\sigma}_\alpha^2$ is 50.84 standard errors above the mean of the null distribution, suggesting a significant country effect. In contrast, $\widehat{\sigma}_\mu^2$ is very close to the mean of the null reference distribution for region effects, suggesting that employment is not subject to a region effect. However, these interpretations are not taking into account the presence of error dependence, due to either a spatial moving average or a spatial autoregressive process.

Spatial error dependence is shown by the other estimators to be highly significant and should be taken into account in interpreting the nested effects embodied in the errors. It is noteworthy that smallest estimate occurs under the GM-S-IV estimator, indicating that this model explains more of the overall variance using ln Q, ln K, the spatial lag, the spatial moving average error process and the national and regional effects in the errors. The other specifications leave more of overall the variance as an unexplained remainder component. It is apparent that controlling for localized spillovers via the spatial moving average process for the errors produces superior outcomes to assuming spatially autoregressive errors, as under GM-S-FGLS and GM-S-IV*, and provides a more appropriate interpretation of the magnitude of country and regional effects than is given by the NRE-IV estimator.

Table 4 Estimates for spatial panel models with nested random effects

	GM-S-IV	NRE-IV	GM-S-FGLS	GM-S-IV*
β_1	0.3025	0.3074	0.5940	0.3338
s.e.	0.0096	0.0080	0.0111	0.0155
t-ratio	31.67	38.28	53.61	21.60
β_2	0.0366	0.0277	0.0288	0.0270
s.e.	0.0031	0.0037	0.0051	0.0059
t-ratio	11.83	7.47	5.62	4.55
ρ	0.2256	0.2250	n.a.	0.2165
s.e.	0.0066	0.0048	n.a.	0.0085
t-ratio	34.24	46.44	n.a.	25.49
λ	-0.8641	n.a.	0.5223[a]	0.3826[a]
s.e.	0.0252	n.a.	0.0256	0.0272
t-ratio	-34.24	n.a.	20.44	14.15
σ_α^2	0.0577	0.1250	0.0411	0.0658
s.e.	0.0022	0.0022	0.0007	0.0065
t-ratio	20.63	50.84	54.60	2.84
σ_μ^2	0.0483	0.0138	0.0397	0.0310
s.e.	0.0037	0.0037	0.0008	0.0014
t-ratio	9.84	0.5522	47.52	21.57
σ_ν^2	0.0008	0.0009	0.0014	0.0020
s.e.	0.0179	0.0179	0.0211	0.0178
t-ratio	-9.58	-9.55	-9.56	-9.55
Direct effect ln Q	0.3033	0.3083	n.a.	0.3347
Total effect ln Q	0.3795	0.3854	n.a.	0.4144
Direct effect ln K	0.0367	0.0277	n.a.	0.0270
Total effect ln K	0.0459	0.0347	n.a.	0.0335

[a]This coefficient is related to the SAR error process

6 Conclusion

In this paper, we focus on estimation methods for a multidimensional spatial lag panel data model with SMA nested random effects errors. The introduction of spatial effects *via* the spatial lag and the errors is an extension of previously published work by Fingleton et al. (2016) in which the spatial effects are only due to the SAR nested random effects errors. The SMA structure constitutes an alternative to incorporating spatial lags of the exogenous variables (X), and potentially avoids the weak instrument problem related to the use of spatial lags of X, given that it embodies local spillovers that one would otherwise control *via* the spatial lags.

We derived GM estimators for the SMA error coefficient and the variance components of the error process. Using a spatial counterpart to the Cochrane–Orcutt transformation, the regression parameters are estimated through a spatial IV esti-

mator. Compared to the ML estimators, the GM-S-IV estimators are computationally feasible even for large sample sizes, and are robust to deviation from distributional assumptions (normality) typifying ML estimation. The Monte Carlo simulations show that RMSE magnitudes of the ML and GM-S-IV estimators of the regression coefficients are globally similar. This means that the benefits associated with the use of the GM-S-IV approaches do not seem to have much cost in terms of efficiency, although it seems that less efficient outcomes for λ and σ_α^2 may be due to our reliance on unweighted GM. It is also evident that there are benefits in terms of efficiency as a result of iterated estimation. The results of the empirical example indicate that in the context of EU regions nested within countries, the assumption of a SMA error process with spatial dependence is preferable to assuming no spatial error dependence, or SAR error dependence. This may reflect the fact that exogenous spatial lags are comparatively unrepresented in the latter, whereas the SMA error process picks up local spillovers explicitly. For future research, possibilities include the investigation of the performance of the weighted GM approach, a study of formal large sample results of the estimators, and the inclusion of dynamic effects in the model.

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht

Antweiler W (2001) Nested random effects estimation in unbalanced panel data. J Econ 101:295–313

Baltagi BH (2013) Econometric analysis of panel data. Wiley, Chichester

Baltagi BH, Song SH, Jung BC (2001) The unbalanced nested error component regression model. J Econ 101:357–381

Baltagi BH, Fingleton B, Pirotte A (2014) Spatial lag models with nested random effects: an instrumental variable procedure with an application to English house prices. J Urban Econ 80:76–86

Chow GC, Lin A-L (1971) Best linear unbiased interpolation, distribution and extrapolation of time series by related series. Rev Econ Stat 53:372–375

Elhorst J (2014) Spatial econometrics: from cross-sectional data to spatial panels. Springer, Heidelberg

Fingleton B (2008) A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. Spat Econ Anal 3:27–44

Fingleton B, Garretsen H, Martin RL (2015) Shocking aspects of monetary union: the vulnerability of regions in Euroland. J Econ Geogr 15:907–934

Fingleton B, Le Gallo J, Pirotte A (2016) Panel data models with spatially dependent nested random effects. J Reg Sci. https://doi.org/10.1111/jors.12327

Kapoor M, Kelejian HH, Prucha IR (2007) Panel data models with spatially correlated error components. J Econ 140:97–130

Kelejian HH, Prucha IR (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J Real Estate Finance Econ 17:99–121

Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. Int Econ Rev 40:509–533

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press, Boca Raton

Nerlove M (1971) Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections. Econometrica 39:359–383

Ord K (1975) Estimation methods for models of spatial interaction. J Am Stat Assoc 70:120–126

Pace R K, LeSage J P, Zhu S (2012) Spatial Dependence in Regressors. In: Terrell D, Millimet D (eds) Advances in econometrics volume 30, Thomas B. Fomby, R. Carter Hill, Ivan Jeliazkov, Juan Carlos Escanciano and Eric Hillebrand. Emerald Group Publishing Limited, Bingley, pp 257–295

Polasek W, Verduras C, Sellner R (2010) Bayesian methods for completing data in spatial models. Rev Econ Anal 2:194–214

Upton G, Fingleton B (1985) Spatial data analysis by example, volume 1: point pattern and quantitative data. Wiley, Chichester

Vidoli F, Mazziotta C (2010) Spatial composite and disaggregate indicators: Chow–Lin methods and application. In: Proceedings of the 45th scientific meeting of the Italian Statistical Society. Padua