Spatial Economic Analysis
Publication details, including instructions for authors and subscription information:
http://rsa.informaworld.com/srsa/title~content=t724921264

A Generalized Method of Moments Estimator for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Average Errors

Bernard Fingleton

To cite this Article Fingleton, Bernard(2008) 'A Generalized Method of Moments Estimator for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Average Errors', Spatial Economic Analysis, 3: 1, 27 — 44
To link to this Article: DOI: 10.1080/17421770701774922
URL: http://dx.doi.org/10.1080/17421770701774922

Spatial Economic Analysis, Vol. 3, No. 1, February 2008

Routledge
Taylor & Francis Group

A Generalized Method of Moments Estimator for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Average Errors

BERNARD FINGLETON

(Received January 2006; accepted August 2007)

ABSTRACT *This paper proposes a new generalized method of moments (GMM) estimator for spatial panel models with spatial moving average errors combined with a spatially autoregressive dependent variable. Monte Carlo results are given suggesting that the GMM estimator is consistent. The estimator is applied to English real estate price data.*

Une méthode généralisée d'estimateur de moments pour un modèle de panel spatial avec un décalage endogène spatial et des erreurs spatiales de type moyenne mobile

RÉSUMÉ *Cette étude propose un nouvel estimateur GMM pour des modèles de panel spatial avec des erreurs spatiales de type moyenne mobile combiné à une variable de dépendance spatiale autorégressive. Les résultats de Monte Carlo fournis suggèrent que l'estimateur GMM est cohérent. L'estimateur s'applique à des données sur des prix d'immobilier anglais.*

Un estimador de método generalizado de momentos para un modelo de panel espacial con retardo espacial endógeno y errores espaciales de media móvil

RESUMEN *Este estudio propone un nuevo estimador GMM para modelos de panel espacial con errores espaciales de media móvil combinado con una variable dependiente autorregresiva. Se indican los resultados de Monte Carlo que revelan la coherencia del estimador GMM. El estimador se aplica a los datos de precios en inmobiliarias inglesas.*

KEYWORDS: *Moving averages; GMM; real estate; spatial econometrics; panel data*

JEL CLASSSIFICATION: C21; R12; R31

Department of Economics, University of Strathclyde, Sir William Duncan Building, 130 Rottenrow, Glasgow G4 0GE, UK. Email: bf100@cam.ac.uk. The author is grateful to Michael Pfaffermayr and the other participants at the International Workshop on Spatial Econometrics and Statistics, Rome, 25–27 May 2006, and participants at the 13th International Conference on Panel Data, University of Cambridge, 7–9 July 2006, for their contributions to the discussion of this paper.

ISSN 1742-1772 print; 1742-1780 online/08/010027-18 DOI: 10.1080/17421770701774922

1. Introduction

There is a growing literature dedicated to the analysis of panel data with spatial dependence, with various different approaches suggested. Probably the most useful starting point in the spatial econometrics literature is Anselin (1988), and among some of the more recent contributions, such as Conley (1999), Chen & Conley (2001), Baltagi *et al.* (2003), Druska & Horrace (2003), Elhorst (2003), Baltagi (2005), Baltagi & Li (2006) and Pinkse *et al.* (2006), we highlight the work of Kapoor *et al.* (2007), which generalizes the generalized moments estimators of Kelejian & Prucha (1999) to a panel data model with spatially and temporally correlated error components, and which provides a feasible generalized least squares procedure for the regression parameters, and formal large sample results for their estimators.

This paper draws on their contribution, which provides the necessary theoretical, computational, and mathematical background for the present paper. Given this context, the specific innovatory aspects of the current paper are:

(i) the extension of the generalized moments estimators (GMM) estimation procedure to allow a spatial moving average (MA) error process rather than the spatial autoregressive process that has been the focus of attention thus far in the literature;
(ii) the extension of the methodology to incorporate an endogenous spatial lag, so that spatial dependence is not solely via the error process;
(iii) application of the method to real panel data involving real estate prices in England.

To summarize, the paper extends the scope of the approach suggested by Kapoor *et al.* (2007) by allowing different forms of spatial interaction for panel data.

2. The Model

Consider the N location cross-sectional time t regression specification

$$Y(t) = \lambda W^E Y(t) + H(t)\gamma + u(t) \tag{1}$$

in which $Y(t)$ is an $N \times 1$ vector of observations of the dependent variable, $H(t)$ is the $N \times k$ matrix of regressors with full column rank, γ is a $k \times 1$ vector of parameters, and $u(t)$ is an $N \times 1$ vector produced by a random error process. Also W^E is an $N \times N$ matrix of non-stochastic time constant weights which defines the interdependence of $Y(t)$ across areas, so that $W^E Y$ is an $N \times 1$ vector commonly referred to as an endogenous spatial lag and λ is a scalar parameter. Following Kapoor *et al.* (2007) and Kelejian & Prucha (1998), all of the diagonal elements of W^E are zero, and $(I - \rho W^E)$ is non-singular. Also W^E is uniformly bounded in absolute value, meaning that a constant c exists such that $\max_{1 \le i \le N} \Sigma_{j=1}^N |W_{ij}^E| \le c < \infty$ and $\max_{1 \le j \le N} \Sigma_{i=1}^N |W_{ij}^E| \le c < \infty$. Likewise, the elements of $H(t)$ are uniformly bounded in absolute value.

The most widely used approach to modelling spatial error dependence involving N locations is to assume that in each period $u(t) = \rho W u(t) + \xi(t)$, in which $u(t)$ is a vector of errors at time t, ρ is a parameter, W is also an $N \times N$ matrix of non-stochastic weights which defines the error interaction across areas and $\xi(t)$ is an $N \times 1$ vector of time t innovations. All the diagonal elements of W are zero,

$(I-\rho W)$ is non-singular and W is also uniformly bounded in absolute value. This is referred to as a spatial autoregressive (AR) process and implies complex interdependence between locations, so that a shock at location j is transmitted to all other locations, as indicated by the expansion of

$$u(t) = (I - \rho W)^{-1}\xi(t) \tag{2}$$

which, assuming $|\rho| < 1$and a row-standardized Wmatrix with row sums equal to 1, is

$$(I - \rho W)^{-1}\xi(t) = (\sum_{i=0}^{\infty} \rho^i W^i)\xi(t) = \xi(t) + \rho W\xi(t) + \rho^2 W^2\xi(t) + \rho^3 W^3\xi(t) + \ldots \tag{3}$$

in which $W^0 = \mathrm{I}$, W^2 is the matrix product of $W$and W, and W^i is the matrix product of W^{i-1} and W. The effect of shock at j is therefore felt directly at j, and there is an indirect effect due to $\rho W\xi(t)$ which affects only those location pairs for which there is a non-zero element on the Wmatrix. If Wwere a contiguity matrix we might think of these as local effects. The global effect of a shock occurs because it is transmitted also to locations that are 'neighbours of neighbours' via the powers of W. Note that the effect rebounds. A shock to j affects the neighbours, and the neighbours of the neighbours, and eventually works its way back to j. In other words, the full effect of a shock to j is not simply the shock itself, but the initial shock plus the feedback from the other locations.

In contrast, the MA error process,[1] which is the subject of this paper, is

$$u(t) = (I - \rho W)\xi(t) \tag{4}$$

so that a shock at location j will only affect the directly interacting locations as given by the non-zero elements in W. Hence shock-effects are local rather than global. Since we are considering a panel with T periods rather than purely cross-sectional data, we omit t to indicate that the observations are stacked. Hence

$$\begin{aligned} Y &= \lambda(I_T \otimes W^E)Y + H\gamma + u = Xb + u \\ X &= ((I_T \otimes W^E)Y, H) \\ b' &= (\lambda, \gamma') \end{aligned} \tag{5}$$

in which Y is a $TN \times 1$ vector of observations, X is a $TN \times (1 + k)$ matrix of regressors, comprising the $TN \times 1$ vector $(I_T \otimes W^E)Y$, which is the endogenous spatial lag, and H is a $TN \times k$ matrix of (exogenous) regressors. In addition, given ξ is an $NT \times 1$ vector of innovations, I_T is a $T \times T$ diagonal matrix with 1s on the main diagonal and 0s elsewhere, I_N is a similar $N \times N$ diagonal matrix, then $I_{TN} = I_T \otimes I_N$ is a $TN \times TN$ diagonal matrix with 1s on the main diagonal and 0s elsewhere and the $NT \times 1$ vector u is given by the MA process

$$u = (I_{TN} - \rho(I_T \otimes W))\xi = \xi - \rho\bar{\bar{\xi}}. \tag{6}$$

Regarding the error components in space-time, time dependency is introduced into the innovations ξ by specifying unobserved permanent unit-specific error components μ together with transient error components v, where

$$\mu \sim \mathrm{iid}(0, \sigma_\mu^2) \tag{7}$$

$$v \sim \text{iid}(0, \sigma_v^2) \tag{8}$$

and

$$\xi = (\iota_T \otimes I_N)\mu + v. \tag{9}$$

Thus μ is the $N \times 1$ vector of errors specific to each area, v is the $NT \times 1$ vector of errors specific to each area and time, ι_T is a $T \times 1$ matrix with 1s, and $\iota_T \otimes I_N$ is a $TN \times N$ matrix equal to T stacked I_N matrices.

For v, it is assumed that $0 < \underline{a}^{\sigma_v} \leq \sigma_v^2 \leq \overline{a}^{\sigma_v} < \infty$ and we also make the standard assumption that the errors have finite fourth moments ($Ev_j^4 < \infty$) to ensure a finite domain for estimation. Likewise for μ, $0 < \underline{a}^{\sigma_\mu} \leq \sigma_\mu^2 \leq \overline{a}^{\sigma_\mu} < \infty$ and $E\mu_j^4 < \infty$. Also, we assume that the error components are independent, hence $E[\mu_i v_{it}] = 0$, and each of the two error components μ and v is subject to the same spatial moving average process, since

$$u = (I_{TN} - \rho(I_T \otimes W))\xi = (I_{TN} - \rho(I_T \otimes W))((\iota_T \otimes I_N)\mu) + (I_{TN} - \rho(I_T \otimes W))v$$

$$= (\iota_T \otimes (I_N - \rho W))\mu + (I_T \otimes (I_N - \rho W))v. \tag{10}$$

For areas i, j and times t, s:

$$E(\xi'\xi) = [\sigma_v^2 + \sigma_\mu^2] \text{ if } i = j;\ t = s$$

$$E[\xi'\xi] = [\sigma_\mu^2] \text{ if } i = j;\ t \neq s$$

$$E[\xi'\xi] = [0] \text{ if } i \neq j;\ t \neq s. \tag{11}$$

For the purposes of estimation, it is useful to represent the $TN \times TN$ innovations variance–covariance matrix $\mathbf{\Omega}_\xi$ using the matrices Q_0 and Q_1 defined[2] as follows

$$Q_0 = (I_T - \frac{J_T}{T}) \otimes I_N, \tag{12}$$

in which J_T is a $T \times T$ matrix of 1s, and

$$Q_1 = \frac{J_T}{T} \otimes I_N. \tag{13}$$

It follows that $Q_0 + Q_1 = I_{TN}$ and

$$\mathbf{\Omega}_\xi = \sigma_\mu^2(J_T \otimes I_N) + \sigma_v^2 I_{TN}$$

$$\mathbf{\Omega}_\xi = \sigma_v^2 Q_0 + \sigma_1^2 Q_1 \tag{14}$$

in which

$$\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2. \tag{15}$$

3. The Moments Equations

Consider the $TN \times 1$ vector of residuals u

$$u = (I_{TN} - \rho(I_T \otimes W))\xi = \xi - \rho\bar{\xi} \tag{16}$$

and

$$\bar{u} = (I_{TN} - \rho(I_T \otimes W))\bar{\xi} = \bar{\xi} - \rho\bar{\bar{\xi}}. \tag{17}$$

Pre-multiplying by Q_0 gives

$$Q_0 u = Q_0(\xi - \rho\bar{\xi}) \tag{18}$$

and

$$\sum_j (Q_0 u)_j^2 = u' Q_0 u = (\xi - \rho\bar{\xi})' Q_0 (\xi - \rho\bar{\xi}) \tag{19}$$

with

$$(\xi - \rho\bar{\xi})' Q_0 (\xi - \rho\bar{\xi}) = \xi' Q_0 \xi + \rho^2 \bar{\xi}' Q_0 \bar{\xi} - 2\rho\xi' Q_0 \bar{\xi} = u' Q_0 u. \tag{20}$$

Likewise

$$(\bar{\xi} - \rho\bar{\bar{\xi}})' Q_0 (\bar{\xi} - \rho\bar{\bar{\xi}}) = \bar{\xi}' Q_0 \bar{\xi} + \rho^2 \bar{\bar{\xi}}' Q_0 \bar{\bar{\xi}} - 2\rho\bar{\xi}' Q_0 \bar{\bar{\xi}} = \bar{u}' Q_0 \bar{u} \tag{21}$$

and

$$(\xi - \rho\bar{\xi})' Q_0 (\bar{\xi} - \rho\bar{\bar{\xi}}) = \xi' Q_0 \bar{\xi} + \rho^2 \bar{\xi}' Q_0 \bar{\bar{\xi}} - \rho\bar{\xi}' Q_0 \bar{\xi} - \rho\xi' Q_0 \bar{\bar{\xi}} = u' Q_0 \bar{u}. \tag{22}$$

Similarly

$$(\xi - \rho\bar{\xi})' Q_1 (\xi - \rho\bar{\xi}) = \xi' Q_1 \xi + \rho^2 \bar{\xi}' Q_1 \bar{\xi} - 2\rho\bar{\xi}' Q_1 \xi = u' Q_1 u \tag{23}$$

$$(\bar{\xi} - \rho\bar{\bar{\xi}})' Q_1 (\bar{\xi} - \rho\bar{\bar{\xi}}) = \bar{\xi}' Q_1 \bar{\xi} + \rho^2 \bar{\bar{\xi}}' Q_1 \bar{\bar{\xi}} - 2\rho\bar{\xi}' Q_1 \bar{\bar{\xi}} = \bar{u}' Q_1 \bar{u} \tag{24}$$

$$(\xi - \rho\bar{\xi})' Q_1 (\bar{\xi} - \rho\bar{\bar{\xi}}) = \xi' Q_1 \bar{\xi} + \rho^2 \bar{\xi}' Q_1 \bar{\bar{\xi}} - \rho\bar{\xi}' Q_1 \bar{\xi} - \rho\xi' Q_1 \bar{\bar{\xi}} = u' Q_1 \bar{u}. \tag{25}$$

Also

$$(\xi - \rho\bar{\xi})'(\xi - \rho\bar{\xi}) = (\xi - \rho\bar{\xi})' Q_0 (\xi - \rho\bar{\xi}) + (\xi - \rho\bar{\xi})' Q_1 (\xi - \rho\bar{\xi}) \tag{26}$$

$$(\bar{\xi} - \rho\bar{\bar{\xi}})'(\bar{\xi} - \rho\bar{\bar{\xi}}) = (\bar{\xi} - \rho\bar{\bar{\xi}})' Q_0 (\bar{\xi} - \rho\bar{\bar{\xi}}) + (\bar{\xi} - \rho\bar{\bar{\xi}})' Q_1 (\bar{\xi} - \rho\bar{\bar{\xi}}) \tag{27}$$

$$(\xi - \rho\bar{\xi})'(\bar{\xi} - \rho\bar{\bar{\xi}}) = (\xi - \rho\bar{\xi})' Q_0 (\bar{\xi} - \rho\bar{\bar{\xi}}) + (\xi - \rho\bar{\xi})' Q_1 (\bar{\xi} - \rho\bar{\bar{\xi}}). \tag{28}$$

To obtain the expectations of these variables, we know (see Kapoor *et al.*, 2007) that

$$E(\xi' Q_0 \xi) = \sigma_v^2 Tr(Q_0) = \sigma_v^2 N(T-1) \tag{29}$$

$$E(\bar{\xi}' Q_0 \bar{\xi}) = E(v' Q_0 (I_T \otimes W'W)' Q_0 v) = \sigma_v^2 (T-1) Tr(W'W) \tag{30}$$

$$E(\xi' Q_0 \bar{\xi}) = 0 \tag{31}$$

and using similar arguments it is also possible to show that

$$E(\bar{\bar{\xi}}' Q_0 \bar{\bar{\xi}}) = \sigma_v^2 (T-1) Tr(W'WW'W) \tag{32}$$

since

$$\begin{aligned}
&E(\bar{\bar{\xi}}' Q_0 \bar{\bar{\xi}}) = E[v' Q_0 (I_T \otimes W'W)'(I_T \otimes W'W) Q_0 v] \\
&Q_0 (I_T \otimes W'W)'(I_T \otimes W'W) Q_0 = Q_0 (I_T \otimes W'W)'(I_T \otimes W'W) \\
&E(\bar{\bar{\xi}}' Q_0 \bar{\bar{\xi}}) = \sigma_v^2 Tr(Q_0 (I_T \otimes W'W)'(I_T \otimes W'W)) \\
&E(\bar{\bar{\xi}}' Q_0 \bar{\bar{\xi}}) = \sigma_v^2 (T-1) Tr(W'WW'W).
\end{aligned} \tag{33}$$

Likewise[3]

$$E(\bar{\xi}' Q_0 \bar{\bar{\xi}}) = \sigma_v^2 (T-1) Tr(W'WW). \tag{34}$$

Also

$$E(\xi' Q_0 \bar{\bar{\xi}}) = Tr(E\xi' Q_0 WW\xi) = \sigma_v^2 (T-1) Tr(WW). \tag{35}$$

By analogy, and following Kapoor *et al.* (2007),

$$E(\xi' Q_1 \xi) = \sigma_1^2 Tr(Q_1) = \sigma_1^2 N \tag{36}$$

$$E(\bar{\xi}' Q_1 \bar{\xi}) = \sigma_1^2 Tr(W'W) \tag{37}$$

$$E(\xi' Q_1 \bar{\xi}) = 0 \tag{38}$$

$$E(\bar{\bar{\xi}}' Q_1 \bar{\bar{\xi}}) = \sigma_1^2 Tr(W'WW'W) \tag{39}$$

$$E(\xi' Q_1 \bar{\bar{\xi}}) = \sigma_1^2 Tr(WW) \tag{40}$$

$$E(\bar{\xi}' Q_1 \bar{\bar{\xi}}) = \sigma_1^2 Tr(W'WW). \tag{41}$$

Ignoring the expectations, we put these equations together using the 3×3 matrices Γ and $\tilde{\Gamma}$, the 3×1 vectors ϕ and $\tilde{\phi}$, and the 3×1 vectors γ and $\tilde{\gamma}$, using $t_1 = Tr(W'W), t_2 = Tr(W'WW'W)$, $t_3 = Tr(W'WW)$, and $t_4 = Tr(WW)$so that

$$\Gamma\phi - \gamma = 0 \tag{42}$$

and

$$\tilde{\Gamma}\tilde{\phi} - \tilde{\gamma} = 0, \tag{43}$$

where

$$\Gamma = \begin{bmatrix} N(T-1) & t_1(T-1) & 0 \\ t_1(T-1) & t_2(T-1) & 2(T-1)t_3 \\ 0 & (T-1)t_3 & (T-1)(t_1+t_4) \end{bmatrix} \quad \phi = \begin{bmatrix} \sigma_v^2 \\ \rho^2\sigma_v^2 \\ -\rho\sigma_v^2 \end{bmatrix} \qquad \gamma = \begin{bmatrix} u'Q_0u \\ \bar{u}'Q_0\bar{u} \\ u'Q_0\bar{u} \end{bmatrix}$$

$$\tilde{\Gamma} = \begin{bmatrix} N & t_1 & 0 \\ t_1 & t_2 & 2t_3 \\ 0 & t_3 & (t_1 + t_4) \end{bmatrix} \qquad \tilde{\phi} = \begin{bmatrix} \sigma_1^2 \\ \rho^2\sigma_1^2 \\ -\rho\sigma_1^2 \end{bmatrix} \qquad \tilde{\gamma} = \begin{bmatrix} u'Q_1 u \\ \bar{u}'Q_1\bar{u} \\ u'Q_1\bar{u} \end{bmatrix}.$$

4. Estimation

The estimation procedure comprises three stages. At stage 1, because of the presence of the spatial lag, we obtain[4] IV estimates of b and hence residuals $\hat{u} = Y - X\hat{b}$. In stage 2 we use these IV residuals to obtain the estimates g and $\tilde{g}$ of γ and $\tilde{\gamma}$, and denoting Γ and $\tilde{\Gamma}$ by G and $\tilde{G}$ we have the sample counterpart of equation (42), which is

$$G[\sigma_v^2 \quad \rho^2\sigma_v^2 \quad -\rho\sigma_v^2]' - g = \zeta(\rho \quad \sigma_v^2) \tag{44}$$

in which $\zeta(\rho \quad \sigma_v^2)$ is a vector of residuals, and the non-linear least squares estimators are given by

$$(\hat{\rho}, \hat{\sigma}_v^2) = \arg\min\{\zeta(\rho, \sigma_v^2)'\zeta(\rho, \sigma_v^2)\}.$$

From the estimated ρ and σ_v^2 one can obtain the estimate of σ_1^2 indirectly, using the fact that, for MA errors, $u = (I_{TN} - \rho(I_T \otimes W))\xi$, hence $\xi = (I_{TN} - \rho(I_T \otimes W))^{-1}u$. Since $E(\xi' Q_1 \xi) = \sigma_1^2 Tr(Q_1) = \sigma_1^2 N$, then

$$\sigma_1^2 = \frac{1}{N}((I_{TN} - \rho(I_T \otimes W))^{-1}u)'Q_1(I_{TN} - \rho(I_T \otimes W))^{-1}u \tag{45}$$

and therefore

$$\hat{\sigma}_1^2 = \frac{1}{N}((I_{TN} - \hat{\rho}(I_T \otimes W))^{-1}\hat{u})'Q_1(I_{TN} - \hat{\rho}(I_T \otimes W))^{-1}\hat{u}. \tag{46}$$

In practice, to obtain a direct estimate of σ_1^2, we also use the sample counterpart of equation (43), which is

$$\tilde{G}[\sigma_1^2 \quad \rho^2\sigma_1^2 \quad -\rho\sigma_1^2]' - \tilde{g} = \tilde{\zeta}(\rho \quad \sigma_1^2), \tag{47}$$

in which $\tilde{\zeta}(\rho \quad \sigma_1^2)$ is a vector of residuals, and obtain $\hat{\rho}, \hat{\sigma}_v^2, \hat{\sigma}_1^2$ as the minimum[5] of $F_1 + F_2$, where $F_1 = \zeta(\rho, \sigma_v^2)'\zeta(\rho, \sigma_v^2)$ and $F_2 = \tilde{\zeta}(\rho, \sigma_1^2)'\tilde{\zeta}(\rho, \sigma_1^2)$.

In general, the variances associated with F_1 and F_2 differ, and Kapoor *et al.* (2007) suggest weighting to allow for this. However, in the Monte Carlo simulations that follow, for simplicity we have not introduced differential weighting. In the analogous situation examined by Kapoor *et al.* (2007) they note that giving equal weight to all six moments equations does give consistent estimates. While the small sample behaviour in the AR case is the worse of the alternative weighting schemes they examine, it seems appropriate commencing with MA errors to initially explore the behaviour of the simplest approach prior to more elaborate methods, which could be the subject of further research.

In the third stage, because the errors $\mathbf{\Omega}_\xi = \sigma_v^2 Q_0 + \sigma_1^2 Q_1$ are not constant, the appropriate method is generalized least squares (GLS), estimated by IV to also allow for the presence of the endogenous spatial lag. The estimated error covariance matrix $\hat{\mathbf{\Omega}}_\xi$ is obtained using $\hat{\sigma}_v^2, \hat{\sigma}_1^2$ from stage 2, but first $\hat{\rho}$ is used to perform a Cochrane–Orcutt (C-O)-type transformation to account for the spatial dependence in the residuals.

Normally with C-O the assumption is an autoregressive error process, hence $u = (I_{TN} - \rho(I_T \otimes W))^{-1}\xi$, in which case one pre-multiplies through by $I_{TN} - \rho(I_T \otimes W)$ to obtain the innovations ξ. However, the MA error process $u = (I_{TN} - \rho(I_T \otimes W))\xi$ requires pre-multiplication[6] by the inverse[7] to obtain ξ, thus

$$\begin{aligned} Y^* &= (I_T \otimes (I_N - \hat{\rho} W))^{-1} Y \\ X^* &= (I_T \otimes (I_N - \hat{\rho} W))^{-1} X \\ \xi &= (I_T \otimes (I_N - \hat{\rho} W))^{-1} u. \end{aligned} \tag{48}$$

In both the first and third stages, to carry out the IV estimation, as instruments we employ a linearly independent subset of the exogenous variables, so that Z is a $TN \times f \geq (k+1)$ matrix of instruments. Assume matrices X and Z are full column rank with $f \geq (k+1)$, and following what is evidently a comparatively robust approach for IV estimation with non-spherical disturbances (Bowden & Turkington, 1984), calculate $P_z = Z(Z'\hat{\Omega}_\xi Z)^{-1} Z'$, which is a symmetric matrix ($P_z \Omega_\xi$ is idempotent) and hence

$$\begin{aligned} \hat{b}^* &= [(X^{*\prime} Z)(Z'\hat{\Omega}_\xi Z)^{-1}(Z'X^*)]^{-1}(X^{*\prime}Z)(Z'\hat{\Omega}_\xi Z)^{-1}(Z'Y^*) \\ &= (X^{*\prime} P_z X^*)^{-1} X^{*\prime} P_z Y^*. \end{aligned} \tag{49}$$

The estimated variance–covariance matrix of the parameters is given by

$$\hat{C} = [(X^{*\prime}Z)(Z'\hat{\Omega}_\xi Z)^{-1}(Z'X^*)]^{-1} = (X^{*\prime} P_z X^*)^{-1}. \tag{50}$$

Greene (2003) also gives the equivalent of (49) and (50) as generalized methods of moments (instrumental variables) estimators with non-spherical disturbances. The standard errors of the $\hat{b}$ are given by the squares roots of the values on the main diagonal of $\hat{C}$, which allows 't-ratios' to be calculated for purposes of inference.

5. Example 1: the Data-generating Process

In this first example the data are purely artificial, and correspond to model (6), which is repeated here for convenience:

$$Y = \lambda I_T \otimes W^E Y + H\gamma + u \tag{51}$$

$$\begin{aligned} X &= (I_T \otimes W^E Y, H) \\ b' &= (\lambda, \gamma') \\ Y &= Xb + u \\ u &= (I_{TN} - \rho(I_T \otimes W))\xi. \end{aligned} \tag{52}$$

In the MA error process, W is a contiguity matrix[8] on a $\sqrt{N} \times \sqrt{N}$ square. Matrix W is standardized by dividing each row cell by its row total, so that the maximum and minimum eigenvalues are 1 and –1. Also in practice, for simplicity we assume that $W^E = W$.

Matrix H had columns equal to the $TN \times 1$ vectors ι_{TN}, H_1, H_2, and H_3, in which ι_{TN} is a $TN \times 1$ vector with 1s. To obtain each H, first generate time $t = 0$,

$N \times 1$ vectors $H_1(0)$, $H_2(0)$, $H_3(0)$ by sampling at random from a uniform (rectangular) distribution with minimum equal to 0 and maximum equal to 1. Then for t equal to $1 \ldots T$, $H_1(t) = H_1(t-1) + \pi_1$, in which $\pi_1 \sim N(0,1)$, and likewise for $H_2(t)$, $H_3(t)$ using $\pi_2 \sim N(0,1)$ and $\pi_3 \sim N(0,1)$. Then stacking these $N \times 1$ vectors we obtain H_1, H_2 and H_3. In his way the exogenous variables in H have some time dependency, as seems reasonable for panel data. Once generated, the variables H_1, H_2 and H_3 remain fixed. Also, in practice $T=2$ and $T=4$ below.

Given the exogenous variables, we next obtain the innovations vector $NT \times 1$ vector ξ. The innovations vector depends on the $N \times 1$ vector μ obtained by sampling from an $N(0, \sigma_\mu^2)$ distribution and on the $NT \times 1$ vector v obtained by sampling from an $N(0, \sigma_v^2)$ distribution, so that $\xi = (\iota_T \otimes I_N)\mu + v$. This is repeated for each iteration $k = 1 \ldots K$, to obtain

$$Y_k = (I_{TN} - \lambda(I_T \otimes W))^{-1} H\gamma + (I_{TN} - \lambda(I_T \otimes W))^{-1}(I_{TN} - \rho(I_T \otimes W))\xi_k. \quad (53)$$

Given Y, W and H, K estimates are obtained of the known parameters $\rho, \lambda, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \sigma_v^2$ and σ_1^2 using the three-stage method outlined above. This is achieved by using instruments Z for the endogenous spatial lag comprising the exogenous variables, H, together with the $TN \times 1$ vector comprising T stacked identical time 'zero' $N \times 1$ spatial lag vectors $WY(0)$, which is assumed to be exogenous with respect to the endogenous lag $(I_T \otimes W)Y$.

5.1. Monte Carlo Results

Monte Carlo results are given both here and in more detail in Appendix B. Those given here are illustrative, while those in Appendix B provide more substantive empirical evidence of the consistency of the estimator. In this first example, the values $\rho = -0.25$, $\lambda = 0.75$, $\gamma_0 = 1$, $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 10$, $\sigma_\mu^2 = 1$ and $\sigma_v^2 = 1$ are used to generate Y, and the three-stage estimation method employed $K = 100$ times[9] gives a set of K estimates $\hat{\lambda}_k, \hat{\gamma}_{0k}, \hat{\gamma}_{1k}, \hat{\gamma}_{2k}, \hat{\gamma}_{3k}, \hat{\rho}_k, \sigma_{vk}^2$ and $\hat{\sigma}_{1k}^2$ $(k = 1 \ldots K)$ of these parameters. Table 1 summarizes the parameter estimate distributions. It is evident that the parameter estimate means are close to the true values, although it is shown later (see Appendix B) that there is evidence of small sample bias in the estimator of ρ, although it is apparently consistent. The distributions are relatively symmetrical, and on the whole have a degree of kurtosis consistent with the normal distribution. To formally test the null of normality, the K estimates are divided into $\sqrt{K} = 10$ groups with upper and lower bounds defined so that each group has approximately the same observed frequency (O_i). These observed frequencies are then compared to expected frequencies (E_i) calculated using the data to obtain maximum-likelihood estimates of the normal mean and variance.[10] The test statistic $X^2 = \Sigma_{i=1}^{10}((O_i - E_i)^2/E_i)$ is then referred to the χ_7^2 distribution. It is apparent that none of the distributions differs significantly from normal, using the upper 5% point (14.07) of the χ_7^2 distribution.

Table 2 summarizes the estimate distributions obtained with $T = 4$ time periods and assuming a different set of parameter values. In this case the true values are $\rho = -0.5$, $\lambda = 0.25$, $\gamma_0 = 1$, $\gamma_1 = 2$, $\gamma_2 = 4$, $\gamma_3 = 6$, $\sigma_\mu^2 = 0.1$ and $\sigma_v^2 = 1$. Once again the distribution means are all quite close to expectation, in most cases with low levels of skewness and kurtosis and acceptable approximations to normality, although the σ_1^2 estimates are clearly furthest from normality.

Table 1. Estimated parameter distributions

$T=2$	$\sqrt{N}=15$	$\mu \sim N(0,\sigma_\mu^2)$ $\sigma_\mu^2=1$ Mean	$v \sim N(0,\sigma_v^2)$ $\sigma_v^2=1$ SD	$W^E=W$ (Rook's case) Skewness	$H_0 \sim$ U[0,1] Kurtosis	$K=100$ Normality
$\lambda=0.75$	$\hat{\lambda}_k$	0.7486	0.0071	−0.04	−0.21	4.44
$\gamma_0=1$	$\hat{\gamma}_{0k}$	1.0960	0.4594	0.19	−0.11	8.58
$\gamma_1=10$	$\hat{\gamma}_{1k}$	10.0013	0.0667	−0.16	−0.39	13.57
$\gamma_2=10$	$\hat{\gamma}_{2k}$	10.0089	0.0680	0.08	−0.47	3.28
$\gamma_3=10$	$\hat{\gamma}_{3k}$	9.9992	0.0651	0.13	0.09	8.05
$\rho=-0.25$	$\hat{\rho}_k$	−0.2373	0.0902	−0.04	−0.75	11.22
$\sigma_v^2=1$	$\hat{\sigma}_{vk}^2$	1.0167	0.0953	0.73	1.76	11.15
$\sigma_1^2=3$	$\hat{\sigma}_{1k}^2$	2.9282	0.2878	0.53	−0.18	4.08

Notes: T = time periods, N = sample size, K = number of replications, W is a standardized $N \times N$ contiguity matrix.

H_0 denotes time 0 distribution of exogenous variables H.

U[0,1] denotes a uniform (rectangular) distribution with minimum equal to 0 and maximum equal to 1.

Skewness is calculated as $\Sigma(x_i - m)^3/(n-1)s^3$.

Kurtosis is calculated as $\Sigma\{(x_i - m)^4/(n-1)s^4\} - 3$ in which $m = \Sigma x_i/n$.

The goodness of fit to the normal distribution is indicated by the residual deviance which has an asymptotic chi-squared distribution with the specified degrees of freedom. The table is formed by dividing the data into groups of approximately equal observed frequency. The degrees of freedom $c-p-1$, where c is the number of cells in the table of fitted values and p is the number of parameters (2) estimated in the model. Here there are 7 df.

The more detailed Monte Carlo results given in Appendix B (Tables A2 to A8) use both Rook's (edge touching) and Queen's (edge and corner touching) definitions of contiguity on a lattice, and also a torus, with opposite sides of the lattice treated as being contiguous so as to eliminate edges. Measures of bias and of a single indicator combining both precision (variance) and accuracy (bias), as given by a variant of the RMSE statistic (see Appendix B), are calculated from 1,000 Monte Carlo replications of equation (53). Summarizing the outcomes obtained under the various alternative assumptions detailed in Tables A2 to A8, it is evident that there is a small sample bias in $\hat{\rho}$. Attention is focused on positive dependence (negative ρ), which gives positive bias, and these outcomes are mirrored in the case of negative dependence,[11] which gives negative bias, so that in both cases the estimated parameter is closer to zero than the true value. The bias is increasing in $\hat{\rho}$, but the most significant result is the clear evidence that as the sample size (N) increases, the bias in $\hat{\rho}$ diminishes and the RMSE falls, suggesting consistency.

Table 2. Estimated parameter distributions

$T=4$	$\sqrt{N}=15$	$\mu \sim N(0,\sigma_\mu^2)$ $\sigma_\mu^2=0.1$ Mean	$v \sim N(0,\sigma_v^2)$ $\sigma_v^2=1$ SD	$W^E=W$ (Rook's case) Skewness	$H_0 \sim$ U[0,1] Kurtosis	$K=100$ Normality
$\lambda=0.25$	$\hat{\lambda}_k$	0.2481	0.0291	−0.37	−0.15	9.03
$\gamma_0=1$	$\hat{\gamma}_{0k}$	1.0254	0.2868	0.35	−0.37	10.68
$\gamma_1=2$	$\hat{\gamma}_{1k}$	1.9996	0.0250	−0.36	0.11	10.05
$\gamma_2=4$	$\hat{\gamma}_{2k}$	3.9950	0.0229	0.02	0.35	16.17
$\gamma_3=6$	$\hat{\gamma}_{3k}$	6.0001	0.0212	0.08	0.61	8.69
$\rho=-0.5$	$\hat{\rho}_k$	−0.4879	0.0590	0.06	−0.46	5.28
$\sigma_v^2=1$	$\hat{\sigma}_{vk}^2$	1.0227	0.0678	0.16	0.64	11.54
$\sigma_1^2=1.4$	$\hat{\sigma}_{1k}^2$	1.5026	0.2832	2.16	7.42	20.59

These indications of consistency are precisely what one might anticipate on the basis of the theoretical results given by Kapoor *et al.* (2007) (see also Kelejian & Prucha, 1998, 1999). An essential difference between their analysis and what is done here is, of course, that here we are assuming a spatial moving average process rather than spatially autoregressive errors. In addition, in this paper we also introduce an endogenous spatial lag in the panel context, a feature absent from their analysis. Consistency of the generalized moments estimators of ρ and σ^2 is maintained by utilizing IV estimates of b leading to consistent disturbances. Thus, although the formal proofs given by Kapoor *et al.* (2007) are in the context of exogenous regressors (no spatial lag) and autoregressive rather than moving average errors, it is clear that their results carry through to the present set-up. Finally, it seems that although there is a small sample positive bias in the estimator $\hat{\rho}$, in many applied situations $\hat{\rho}$ will be effectively unbiased. One advantage of GMM estimation is its comparative simplicity and computational efficaciousness[12] in applications in which the number of locations is far in excess of those subject to Monte Carlo exploration in this paper. It is clear from the results presented here that as the number of locations rises into the thousands, as for example with the 3,000 plus counties in the USA, small sample bias in the estimator $\hat{\rho}$ should be minimal.

6. Example 2: Real Estate Prices

In this example the GMM estimator is applied to a panel of average house prices in $N = 353$ small areas[13] of England in the $T = 2$ years 2000, 2001, denoted by the $NT \times 1$ vector p. If the price at j is comparatively high, then demand may be displaced to nearby location k. On the other hand, supply may be displaced from k to j as investors in property seek higher returns. We therefore assume that price in area j interacts contemporaneously with price in area k, and model this interaction by the presence of an endogenous spatial lag Wp. In this case we again use the row normalized contiguity matrix W for both the endogenous spatial lag and the MA error process. The other explanatory variables[14] are income from local jobs (wE), equal to the local wage rate (w) times the local employment level (E), and income from wages and employments within commuting distance ($w^c E^c$). In order to be able to treat these as exogenous variables, 1 year lags are introduced, so that year 2001 prices are a function of income in 2000, and year 2000 prices are a function of income in 1999.

There are many other variables that one might wish to introduce were panel data available, such as air quality, the quality of local schooling, the size of the existing stock of properties, demand coming from non-wage earners such as the retired and students, and the effects of criminality, social quality of the neighbourhood, amenity, local taxes, the nature of the housing stock, planning and building regulations, vulnerability to flooding and therefore the additional insurance premiums for areas on flood plains, and various other social, demographic, labour market, environmental and cultural differences. These omitted variables are likely to be spatially autocorrelated, the net effect of which is to induce an organized residual pattern (Dubin, 1988). We model these omitted variables by the spatial MA error process. While displaced demand or supply may causes price interactions that cascade outwards in an autoregressive process, I assume no such chain reaction for these variables, so that a shock, on its own, has a limited spatial extent, which is a property of the spatial MA process.

Table 3. GMM estimates for the real estate price panel data with spatial moving average errors

	Parameter	Par. est.	SE	*t*-ratio
Constant	γ_0	44.4949 (43.6422)	12.3936 (12.5396)	3.59 (3.48)
Wp	λ	0.3941 (0.4011)	0.1174 (0.1196)	3.36 (3.35)
wE	γ_1	0.2221 (0.2494)	0.0976 (0.0978)	2.27 (2.55)
w^cE^c	γ_2	0.0507 (0.0493)	0.0066 (0.0067)	7.68 (7.34)
MA	ρ	$-$0.49279 (0.34998)		
	σ_v^2	42.9822 (37.4959)		
	σ_1^2	2722.83133 (2566.9143)		

Note: Estimates using AR errors are given in parentheses.

In order to obtain the estimates given in Table 3, the exogenous variables wE and w^cE^c and their first spatial lags, obtained by pre-multiplying these vectors by W, were used as instruments for the endogenous spatial lag in the first stage of the three-stage estimation process. This then provided estimated IV residuals $\hat{u}$ which facilitated the second stage, enabling ρ, σ_v^2 and σ_1^2 estimates to be obtained. The third-stage estimates are given in Table 3, showing that there is a significant endogenous spatial lag effect, so that prices are directly positively related to contemporaneous prices in contiguous areas, and there are significant effects due to income from local jobs and jobs within commuting distance.

7. Conclusion

This paper considers panel data in which spatial interaction comes from the effects of an endogenous lag and also from the MA error process. Monte Carlo results are given suggesting that the GMM estimator is consistent. It appears that this is the first paper to consider panel analysis with spatial MA errors, and also to jointly consider an endogenous lag together with spatial and temporal correlation in the error components, although much of this has been presaged in the earlier spatial econometrics literature (Anselin, 1988), and also in the time series context (Harvey, 1990). Indeed, in the conclusion to their paper, Kapoor *et al.* (2007) suggest that they would like to extend their results to models containing spatially lagged dependent variables. The present paper raises many issues which should be the subject of further study, such as the choice of appropriate instruments, the most efficient optimization method, and the small sample properties of the estimator, but the evidence presented here does suggest that there is scope for the practical implementation via GMM of panel data models with an endogenous spatial lag and spatial error processes.

Notes

1. An early detailed account of the MA spatial process is given by Haining (1978).
2. Pre-multiplication of a $TN \times 1$ vector θ by Q_0 creates a $TN \times 1$ vector of deviations from the mean, where the mean is obtained by averaging θ over time. Pre-multiplication of a $TN \times 1$ vector θ by Q_1 creates a $TN \times 1$ vector, comprising N across time area-specific means stacked for each T.
3. Note that $Tr(W'WW) = 0$ for the Rook's case contiguity matrix.
4. So that we can use equation (49) in both stage 1 and stage 3, it is assumed that $\sigma_v^2 = 1$ and $\sigma_1^2 = 1$ (so that Ω_ξ is a diagonal matrix of 1s) and that $\rho = 0$. The result is that at stage 1 we simply obtain IV estimates.
5. Using unconstrained non-linear least squares estimation. The method is a modified Newton–Raphson method which is suitable for minimizing any non-linear function, and which depends on numerical differences rather than derivatives.

6. In contrast, at stage 1, ρ is assumed to equal 0, so that in that case $Y^* = Y$, $X^* = X$.
7. Moore–Penrose generalized inverses are used to avoid singularities.
8. In the main body of the text W is a Rook's case contiguity matrix based on a 15×15 lattice. In the Monte Carlo simulations described in Appendix B, the lattice size is varied and an irregular spatial partitioning is also considered. Also, alternative contiguity definitions, namely the Queen's case and torus, are also implemented.
9. Appendix B gives the results obtained using $K = 1{,}000$ replications.
10. These procedures were carried out using the DISTRIBUTION directive of the programming language GENSTAT. The DISTRIBUTION directive is used to fit an observed sample of data to a theoretical distribution function, in order to obtain maximum-likelihood estimates of the parameters of the distribution and test the goodness of fit.
11. To save space these results have not been reported here.
12. With the MA error process the C-O transform involves the inverse. See Smirnov & Anselin (2001) for a discussion of the use of the power expansion to approximate the matrix inverse with large matrices.
13. Unitary authority and local authority districts, or UALADs.
14. Appendix A gives details of the sources and construction of these variables.
15. Small administrative areas, with median area equal to 250.77 km^2.
16. Available on the NOMIS website (the ONS online labour market statistics database).
17. 1991 Census of Population—Special Workplace Statistics, available from NOMIS.
18. Total employees and self-employed with a workplace coded, tabulated by residents in each zone (10% sample).
19. Minimum of the sum of the squared deviations of the observed proportions in each distance band up to 40 km and the proportions of the sum of the function $\exp(-\delta_i d_{ij})$ calculated using the upper limit of each distance band.

References

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Dordrecht, Kluwer.

Baltagi, B. H. (2005) *Econometric Analysis of Panel Data*, 3rd edn, Chichester, Wiley.

Baltagi, B. H. & Li, D. (2006) Prediction in the panel data model with spatial correlation: the case of liquor, *Spatial Economic Analysis*, 1, 175–185.

Baltagi, B. H., Song, S. H. & Koh, W. (2003) Testing panel data regression models with spatial error correlation, *Journal of Econometrics*, 117, 123–150.

Bowden, R. J. & Turkington, D. A. (1984) *Instrumental Variables*, Cambridge, Cambridge University Press.

Brueckner, J. K. (2003) Strategic interaction among governments: an overview of empirical studies, *International Regional Science Review*, 26, 175–188.

Chen, X. & Conley, T. G. (2001) A new semiparametric spatial model for panel time series, *Journal of Econometrics*, 105, 59–83.

Conley, T. G. (1999) GMM estimation with cross sectional dependence, *Journal of Econometrics*, 92, 1–45.

Druska, V. & Horrace, W. C. (2003) *Generalized Moments Estimation for Spatial Panel Data*, Technical Working Paper 291, National Bureau of Economic Research, Cambridge, MA.

Dubin, R. A. (1988) Estimation of regression coefficients in the presence of spatially autocorrelated error terms, *Review of Economics and Statistics*, 70, 466–474.

Elhorst, J. P. (2003) Specification and estimation of spatial panel data models, *International Regional Science Review*, 26, 244–268.

Greene, W. H. (2003) *Econometric Analysis*, 5th edn, Upper Saddle River, NJ, Prentice Hall.

Haining, R. P. (1978) The moving average model for spatial interaction, *Transactions of the Institute of British Geographers*, 3, 202–225.

Harvey, A. C. (1990) *The Economic Analysis of Time Series*, 2nd edn, Cambridge, MA, MIT Press.

Kapoor, M., Kelejian, H. H. & Prucha, I. (2007) Panel data models with spatially correlated error components, *Journal of Econometrics*, 140, 97–130.

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

Pinkse, J., Slade, M. & Shen, L. (2006) Dynamic spatial discrete choice using one-step GMM: an application to mine operating decisions, *Spatial Economic Analysis*, 1, 53–99.

Smirnov, O. & Anselin, L. (2001) Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach, *Computational Statistics and Data Analysis*, 35, 301–319.

Appendix A

The dependent variable p is the mean transaction price (all types of residential property) by area for the period July–September 2000, and July–September 2001 for the $N = 353$ English Unitary Authority and Local Authority Districts[15] (UALADs). The data were provided by the Land Registry. The wage rate (w) is the gross weekly pay for all occupations and both males and females taken from the Office for National Statistics'[16] (ONS) New Earnings Survey. The employment level for the years 1999 and 2000 is based on the annual business enquiry employee analysis, also carried out by the ONS and available on the NOMIS database.

Total earnings in an area is the product of the average wage rate (w) in 1999 and 2000 and the total level of employment in 1999 and 2000, denoted by wE. These are assumed to be predetermined with respect to year 2000 and 2001 price levels.

The vector w^cE^c denotes total earnings within commuting distance of a UALAD. This is equal to the matrix product of the $n \times n$ matrix C and the $n \times 1$ vector wE. Matrix C is defined as follows:

$$C_{ij} = \exp(-\delta_i d_{ij}) \qquad i \neq j$$

$$C_{ij} = 0 \qquad i = j$$

$$C_{ij} = 0 \quad d_{ij} > 100 \text{ km.}$$

Cell (i, j) of the C matrix is a function of the (straight line) distance (d_{ij}) between areas i and j and an area-specific coefficient δ_i. This allows for the different levels of transport infrastructure and commuting in different areas, with the choice of exponent δ_i based on empirical comparisons with observed census data[17] on travel-to-work patterns. Table A1 shows the overall proportion of workers[18] living in England and Wales travelling various distances from home to work. Given observed travel percentages comparable to Table A1 for each area, the exponent δ_i for each area was chosen by iterating the function $\exp(-\delta_i d_{ij})$ through a range of values to obtain the value giving the closest fit[19] to each area's commuting data.

Table A1. Commuting distances to work in England and Wales

	Distance (km)							
	<2	2–4	5–9	10–19	20–29	30–39	>40	
%	26.63	25.28	20.93	15.90	4.96	2.05	4.25	100

Appendix B: Monte Carlo Investigation

Bias = median − true parameter value

$$\mathrm{RMSE} = \left[\mathrm{bias}^2 + \left(\frac{\mathrm{IQ}}{1.35}\right)^2\right]^{0.5},$$

where IQ is the interquantile range, equal to the difference between the 0.75 and 0.25 quantile. While this approximation is based on IQ rather than the variance, under normality the median is equal to the mean and, apart from slight rounding, IQ/1.35 is the standard deviation, so this measure reduces to the standard RMSE statistic (see Kapoor *et al.*, 2007).

In all cases W is normalized to row totals equal to 1 and the bias is based on 1,000 Monte Carlo replications.

Table A2. Bias and RMSE with increasing lattice size, positive dependence, Rook's case. $\rho = -0.25$, $\lambda = 0.75$, $\gamma_0 = 1$, $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 10$, $\sigma_\mu^2 = 1$, $\sigma_\nu^2 = 1$. Matrix W is a Rook's case contiguity matrix. There are two layers ($T = 2$)

Bias

$T = 2$		$\mu \sim N(0, \sigma_\mu^2)$	$\nu \sim N(0, \sigma_\nu^2)$	$W^E = W$ Rook's case	$H \sim \mathrm{U}[0,10]$	$K = 1{,}000$
Lattice size	5 × 5	7 × 7	9 × 9	11 × 11	13 × 13	15 × 15
γ_0	0.31131	0.25106	0.10097	0.007300	−0.09312	0.010786
λ	−0.00031	−0.00048	−0.00006	0.000041	0.00012	−0.000092
γ_1	0.00009	−0.00078	−0.00121	0.000398	0.00051	0.000472
γ_2	−0.00385	0.00209	−0.00120	−0.000090	0.00043	0.001027
γ_3	0.00647	−0.00226	−0.00037	0.002743	0.00251	−0.000860
ρ	0.09523	0.07720	0.05598	0.039236	0.02294	0.012190

RMSE

$T = 2$				$\mu \sim N(0, \sigma_\mu^2)$ $H \sim \mathrm{U}[0,10]$	$\nu \sim N(0, \sigma_\nu^2)$ $K = 1{,}000$	$W^E = W$ Rook's case
Lattice size	5 × 5	7 × 7	9 × 9	11 × 11	13 × 13	15 × 15
γ_0	5.489	3.035	2.519	1.265	1.665	1.368
λ	0.008289	0.004939	0.004075	0.002079	0.002743	0.002337
γ_1	0.07910	0.06366	0.04657	0.03653	0.02868	0.02700
γ_2	0.06890	0.06400	0.05137	0.03187	0.02646	0.02800
γ_3	0.09746	0.05135	0.04204	0.03859	0.03179	0.02603
ρ	0.2180	0.1744	0.1391	0.1192	0.1029	0.08351

Table A3. Bias and RMSE with increasing lattice size, with negative dependence, Rook's case. $\rho = 0.5$, $\lambda = 0.75$, $\gamma_0 = 1$, $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 10$, $\sigma^2_\mu = 1$, $\sigma^2_v = 1$. Matrix W is a Rook's case contiguity matrix. There are two layers ($T = 2$)

Bias

$T = 2$	$\mu \sim N(0, \sigma^2_\mu)$	$v \sim N(0, \sigma^2_v)$	$W^E = W$ Rook's case	$H \sim$ U[0,10]	$K = 1{,}000$
Lattice size	5 × 5	7 × 7	9 × 9	11 × 11	13 × 13
γ_0	− 0.05185	0.09947	− 0.004240	− 0.007804	0.04937
λ	0.00007	− 0.00014	0.000012	− 0.000016	− 0.00009
γ_1	0.00082	0.00102	0.000088	− 0.000737	0.00235
γ_2	− 0.00586	− 0.00128	− 0.000204	− 0.001198	0.00009
γ_3	− 0.01132	− 0.00224	− 0.001816	− 0.001393	0.00060
ρ	− 0.04482	− 0.05278	− 0.019962	− 0.022022	− 0.01821

RMSE

	$\mu \sim N(0, \sigma^2_\mu)$	$v \sim N(0, \sigma^2_v)$	$W^E = W$ Rook's case	$H \sim$ U[0,10]	$K = 1{,}000$
Lattice size	5 × 5	7 × 7	9 × 9	11 × 11	13 × 13
γ_0	2.644	2.202	0.8189	0.4571	0.8922
λ	0.005466	0.004334	0.001814	0.000999	0.001852
γ_1	0.09958	0.05722	0.05082	0.03614	0.03045
γ_2	0.08222	0.06473	0.04433	0.03452	0.02953
γ_3	0.1147	0.05843	0.04015	0.03460	0.03073
ρ	0.2690	0.1859	0.1483	0.1337	0.1168

Table A4. Bias with increases in the time dimension, positive dependence, Rook's case. $\rho = -0.25$, $\lambda = 0.75$, $\gamma_0 = 1$, $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 10$, $\sigma^2_\mu = 1$, $\sigma^2_v = 1$. Matrix W is a Rook's case contiguity matrix for a 7 × 7 square, hence the number of locations is $N = 49$, and W is of dimension 49 × 49. The number of layers is from 3 up to 10

		$\mu \sim N(0, \sigma^2_\mu)$ $H \sim$ U[0,10]	$v \sim N(0, \sigma^2_v)$ $K = 1{,}000$	$W^E = W$ Rook's case
T	3	4	5	10
γ_0	− 0.11550	− 0.03176	− 0.15351	− 0.04148
λ	0.00017	0.00006	0.00029	0.00013
γ_1	0.00238	0.00174	0.00289	− 0.00148
γ_2	− 0.00289	0.00141	0.00103	0.00096
γ_3	0.00143	− 0.00157	0.00088	0.00352
ρ	0.05291	0.05349	0.05032	0.07608

Table A5. Bias with increasing positive dependence, Rook's case. $\lambda = 0.75$, $\gamma_0 = 1$, $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 10$, $\sigma_\mu^2 = 1$, $\sigma_v^2 = 1$, $T = 2$. Matrix W is a Rook's case contiguity matrix for a 7×7 square, hence the number of locations is $N = 49$, and W is of dimension 49×49. There are two layers ($T = 2$)

	$\rho = 0$	$\rho = -0.25$	$\rho = -0.5$	$\rho = -0.75$	$\rho = -0.95$
γ_0	−0.005857	−0.08888	0.02602	0.06492	0.01550
λ	0.000164	0.00010	0.00010	0.00003	0.00000
γ_1	0.000023	0.00168	0.00042	−0.00056	0.00125
γ_2	0.000085	−0.00157	−0.00111	−0.00005	−0.00230
γ_3	−0.005534	−0.00041	−0.00445	−0.00084	−0.00028
ρ	0.044455	0.06901	0.12872	0.19019	0.25428

Table A6. Bias with increasing positive dependence, Queen's case. $\lambda = 0.75$, $\gamma_0 = 1$, $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 10$, $\sigma_\mu^2 = 1$, $\sigma_v^2 = 1$, $T = 2$. Matrix W is a Queen's case contiguity matrix for a 9×9 square, hence the number of locations is $N = 81$, and W is of dimension 81×81. There are two layers ($T = 2$)

	$\rho = 0$	$\rho = -0.25$	$\rho = -0.5$	$\rho = -0.75$	$\rho = -0.95$
γ_0	−0.13457	−0.10588	−0.17422	−0.17873	−0.26565
λ	0.00019	0.00025	0.00026	0.00025	0.00029
γ_1	−0.00075	−0.00077	−0.00063	−0.00051	−0.00073
γ_2	−0.00024	0.00001	0.00029	0.00056	0.00021
γ_3	−0.00053	0.00020	−0.00024	−0.00065	−0.00036
ρ	0.07173	0.10511	0.13368	0.17192	0.19795

Table A7. Bias and RMSE with increasing lattice size, with positive dependence, Queen's case. $\rho = -0.25$, $\lambda = 0.75$, $\gamma_0 = 1$, $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 10$, $\sigma_\mu^2 = 1$, $\sigma_v^2 = 1$. Matrix W is a Queen's case contiguity matrix. There are two layers ($T = 2$)

Bias

$T=2$		$\mu \sim N(0, \sigma_\mu^2)$	$v \sim N(0, \sigma_v^2)$	$W^E = W$ Queen's case	$H \sim$ U[0,10]	$K = 1{,}000$
Lattice size	5 × 5	7 × 7	9 × 9	11 × 11	13 × 13	15 × 15
γ_0	0.15583	0.40579	0.05725	0.00685	−0.07037	0.07598
λ	−0.00002	−0.00077	−0.00001	−0.00001	0.00012	−0.00018
γ_1	0.00077	−0.00088	−0.00048	−0.00058	0.00105	0.00057
γ_2	−0.00188	0.00266	−0.00157	0.00240	0.00067	0.00131
γ_3	0.00056	0.00049	−0.00163	0.00119	0.00364	−0.00119
ρ	0.27036	0.14180	0.06174	0.06889	0.04701	0.03344

RMSE

				$\mu \sim N(0, \sigma_\mu^2)$ $H \sim$ U[0,10]	$v \sim N(0, \sigma_v^2)$ $K = 1{,}000$	$W^E = W$ Queen's case
Lattice size	5 × 5	7 × 7	9 × 9	11 × 11	13 × 13	15 × 15
γ_0	5.117	4.296	4.441	1.678	2.345	1.794
λ	0.008873	0.007006	0.007649	0.002720	0.003938	0.003026
γ_1	0.06928	0.06651	0.04646	0.03442	0.02895	0.02723
γ_2	0.1033	0.06679	0.04325	0.03177	0.02564	0.02739
γ_3	0.1116	0.05538	0.04431	0.03356	0.03217	0.02701
ρ	0.3914	0.2835	0.2160	0.1789	0.1529	0.1297

Table A8. Bias with increasing lattice size, with positive dependence, torus. $\rho = -0.75$, $\lambda = 0.25$, $\gamma_0 = 1$, $\gamma_1 = 10$, $\gamma_2 = 10$, $\gamma_3 = 10$, $\sigma_\mu^2 = 1$, $\sigma_v^2 = 1$. Matrix W is a torus, in other words a Rook's case contiguity matrix with no edges. There are three layers ($T = 3$)

$T=3$		$\mu \sim N(0, \sigma_\mu^2)$ $H \sim$ U[0,10]	$v \sim N(0, \sigma_v^2)$ $K = 1{,}000$	$W^E =$ W Rook's case
Lattice size	5 × 5	7 × 7	9 × 9	11 × 11
γ_0	−0.01686	−0.04101	−0.02184	−0.01271
λ	−0.00008	0.00018	0.00022	−0.00020
γ_1	−0.00450	0.00127	−0.00189	0.00103
γ_2	−0.00061	0.00089	−0.00009	−0.00149
γ_3	−0.00013	0.00046	0.00175	−0.00044
ρ	0.20835	0.14658	0.09758	0.08673