A Time-Space Dynamic Panel Data Model with Spatial Moving Average Errors

Badi H. Baltagi and Bernard Fingleton and Alain Pirotte

18 April 2018

A Time-Space Dynamic Panel Data Model with Spatial Moving Average Errors

Badi H. Baltagi[a], Bernard Fingleton[b] and Alain Pirotte[c]

[a] *Department of Economics and Center for Policy Research, Syracuse University, United States*

[b] *Department of Land Economy, University of Cambridge, United Kingdom*

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

April 18, 2018

Abstract

This paper focuses on the estimation and predictive performance of several estimators for the time-space dynamic panel data model with Spatial Moving Average Random Effects (SMA-RE) structure of the disturbances. A dynamic spatial Generalized Moments (GM) estimator is proposed which combines the approaches proposed by Baltagi, Fingleton and Pirotte (2014) and Fingleton (2008). The main idea is to mix non-spatial and spatial instruments to obtain consistent estimates of the parameters. Then, a forecasting approach is proposed and a linear predictor is derived. Using Monte Carlo simulations, we compare the short-run and long-run effects and evaluate the predictive efficiencies of optimal and various suboptimal predictors using the Root Mean Square Error (RMSE) criterion. Last, our approach is illustrated by an application in geographical economics which studies the employment levels across 255 NUTS regions of the EU over the period 2001-2012, with the last two years reserved for prediction.

Keywords: Panel data; Spatial lag; Error components; Time-space; Dynamic; OLS; Within; GM; Spatial autocorrelation; Direct and indirect effects; Moving average; Prediction; Simulations, Rook contiguity, Interregional trade.

JEL classification: C23.

---

*Corresponding author. Badi H. Baltagi. Tel.: + 1 315 443 1630; Fax: + 1 315 443 1081. *E-mail:* bbaltagi@maxwell.syr.edu.

1 Introduction

There is a growing literature dedicated to dynamic spatial panels, see Elhorst (2012, 2014), Lee and Yu (2010, 2015) for a review. Dynamic spatial panels are able to deal with unobservable spatial, individual and/or time specific effects. They also tackle more efficiently endogeneity problems, such as the potential bias in the coefficient of the spatial lag of the dependent variable. Baltagi, Fingleton and Pirotte (2014) propose a dynamic spatial model that allows for time dependence as well as spatial dependence, but restricts the time-space covariate to zero along the lines of Franzese and Hays (2007), Kukenova and Monteiro (2009), Jacobs, Ligthart and Vrijburg (2009), Korniotis (2010), Elhorst (2010), and Brady (2011) to mention a few. They focus on several estimators for the dynamic and autoregressive spatial lag panel data model with spatially correlated disturbances. In the spirit of Arellano and Bond (1991) and Mutl (2006), a dynamic spatial Generalized Method of Moments (GM) is proposed including the Kapoor, Kelejian and Prucha (2007, hereafter KKP) approach for the Spatial AutoRegressive (SAR) error term. The remainder term of the latter is assumed to have a random effects structure. This means that the disturbances are characterised by a SAR-RE process. The main idea is to mix non-spatial and spatial instruments to obtain consistent estimates of the parameters. Their Monte Carlo study finds that when the true model is a dynamic first order spatial autoregressive specification with SAR-RE disturbances, estimators that ignore the endogeneity of the spatial lag of the dependent variable and endogeneity of the temporally lagged dependent variable perform badly in terms of bias and RMSE. Accounting for spatial correlation in the disturbances also reduces bias and RMSE. Thus, ignoring both sources of spatial dependence leads to a huge bias in the estimated coefficients.

A more general specification has been suggested by Yu, de Jong and Lee (2008, 2012), Lee and Yu (2010, 2014) who implement a model that allows for time and spatial dependence as well as component mixing of time-space dependence that can be interpreted as spatial diffusion that takes place over time. If all these forms of dependence are included, Anselin (2001, p. 318) suggests calling this a time-space dynamic model. Parent and LeSage (2010, 2011, 2012), Debarsy, Ertur and LeSage (2012) extend this approach introducing a space-time filter that implies a constraint on the mixing term that reflects spatial diffusion. This type of constraint induces a separability of time and space dependence that simplifies the estimation procedure, especially in the Bayesian Markov Chain Monte Carlo (MCMC) approach. Lee and Yu (2016) consider a spatial Durbin dynamic model which includes simultaneously time dependence, spatial dependence and time-space dependence on the explanatory variables. They focus on identification issues and show that parameters are generally identified *via* Two-Stage Least Square (2SLS) moment relations.

Following this literature, this paper employs a time-space specification as defined by Anselin (2001, p. 318), i.e. including a temporal lag (capturing time dependence), a spatial lag (that accounts for spatial dependence), a cross-product term reflecting the time-space diffusion of the dependent variable. Many theoretical and/or applied papers are based on this structure, building on *a priori* theory which does not call for **WX**s in the model, see Yu, de Jong and Lee (2008), Parent and LeSage (2010, 2011, 2012), Lee and Yu (2014) or Yang (2017) among others. Additionally, the disturbances are assumed to follow a Spatial Moving Average (SMA) process (local spatial spillover effects) in the spirit of Fingleton (2008*a*). Fingleton (2008*b*) shows that the sharp cut-off of the SMA disturbances specification is moderated by the spatial lag element of the model. In the cross-section 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 a full column rank for the matrix of instruments, and thus recommend that $(\mathbf{X}, \mathbf{W}_N\mathbf{X})$ should be used, if the number of regressors is large. Inclusion of spatial lags on the explanatory variables could have a major impact on the performance of the estimation procedure if one were to keep to this recommendation. Pace, LeSage and Zhu (2012) show that the instrumental variables estimation suffers greatly in situations where spatial lags of the explanatory variables $(\mathbf{W}_N\mathbf{X})$ are included in the model specification. The reason is that this requires the use of $(\mathbf{W}_N^2\mathbf{X}, \mathbf{W}_N^3\mathbf{X}, \ldots)$ as instruments, in place of the conventional instruments that rely on $\mathbf{W}_N\mathbf{X}$, and this appears to result in a weak instruments problem. However the SMA process avoids this problem, given that it embodies the same $\mathbf{W}_N$ matrix, by including omitted spatial lags of the explanatory variables implicitly as part of the error process. This means we can use the recommended instrument set without having exogenous spatial lags among the set of regressors. Instead, we assume that the disturbances are characterized by a SMA-RE structure which purposefully captures these local spillovers. The adoption of a SMA specification of the error process can mitigate against the problem for instrumental variables estimation identified by Pace, LeSage and Zhu (2012). Naturally the choice of this specification would carefully examine the nature of the local spillovers in order to establish their appropriateness for the empirical application at hand.

This paper proposes a spatial GM estimator following the work of Arellano and Bond (1991), Mutl (2006) and KKP (2007). Using Monte Carlo simulations, we compare the empirical performance of our GM spatial estimator with that of OLS, Within and GMM *à la* 'Arellano and Bond'. The latter estimators are panel data estimators that take no account of the spatial structure of the disturbances, especially on short-run and long-run effects. We also compare our spatial estimator to other misspecified spatial GM estimators, such as that of Mutl (2006). Moreover, forecasting with spatial panel has become recently an integral part of the empirical work in economics, see Baltagi and Li (2004, 2006), Longhi and Nijkamp (2007), Kholodilin, Siliverstovs, and Kooths (2008), Fingleton (2009), Schanne, Wapler and Weyh (2010), Girardin and Kholodilin (2011) and Baltagi, Fingleton and Pirotte (2014) among others. We develop a dynamic spatial predictor and evaluate the predictive efficiencies of various suboptimal predictors relative to the Root Mean Square Error (RMSE) criterion along the lines of Kelejian and Prucha (2007). The plan of the paper is as follows: Section 2 presents the model, section 3 focuses on our spatial GM estimator. Section 4 derives a linear predictor. Section 5 describes the Monte Carlo design. Section 6 presents the simulation results, Section 7 illustrates our approach using an application in geographical economics which studies employment levels across 255 NUTS regions of the EU over the period 2001-2012, with estimation to 2010 and out-of-sample prediction for 2011 and 2012. The last section concludes.

2 The Time-Space Dynamic Panel Model with SMA Errors

Unlike many expositions of spatial econometric models, we give logically consistent reasons behind the presence of the spatial lag, temporal lag and spatial lag of the temporal lag which are an outcome of assuming a tendency towards equilibrium. In our exposition the model specification depends on an assumption that disparities in the dependent variable will persist as an equilibrium outcome to unchanging and fundamental causes. We therefore assume that the $(N \times 1)$ vector $\mathbf{y}_t$, where N is the number of individuals or regions, at time t will persist at dynamically stable levels so that $\mathbf{y}_t = \mathbf{y}_{t-1}$ unless there are changes in factors that affect the level of $\mathbf{y}_t$. For example there may be changes in explanatory variables $\mathbf{x}_t$, where $\mathbf{x}_t = (x_{1t}, \ldots, x_{Nt})^\top$ is an $(N \times K)$ matrix of explanatory (exogenous) variables, or other changes, such as in unobservable effects. If such a disturbance occurs at time t and is ephemeral, then $\mathbf{y}_t \neq \mathbf{y}_{t-1}$ but given a subsequent period of quiescence as $t \to T$ then once again we expect $\mathbf{y}_t$ to converge on a new equilibrium at which $\mathbf{y}_T = \mathbf{y}_{T-1}$. Assume data are observed where $\mathbf{y}_t \neq \mathbf{y}_{t-1}$ but tending to converge, so that $\mathbf{y}_t = f(\mathbf{y}_{t-1})$, and an autoregressive process is assumed, hence

$$\mathbf{y}_t = \boldsymbol{\varsigma} + \gamma \mathbf{y}_{t-1}, \tag{1}$$

in which $\boldsymbol{\varsigma}$ is an $(N \times 1)$ vector and γ is a scalar parameter. In the long-run with $|\gamma| < 1$, and with no subsequent disturbances, the process converges to $\mathbf{y}_T = \frac{\boldsymbol{\varsigma}}{1-\gamma}$.

Consider next connectivity between individuals or regions in the form of a matrix $\mathbf{W}_N^*$, which is a time-invariant $(N \times N)$ matrix. For purposes of interpreting parameter estimates we normalize $\mathbf{W}_N^*$. This can be done in several ways, for instance by dividing $\mathbf{W}_N^*$ by the maximum eigenvalue of $\mathbf{W}_N^*$ to give[1] $\mathbf{W}_N$, or by dividing each element of $\mathbf{W}_N^*$ by its row sum. Either of these normalizations gives the maximum eigenvalue of $\mathbf{W}_N$ equal to 1, and the continuous range for which $(\mathbf{I}_N - \rho_1 \mathbf{W}_N)$ is nonsingular is $\frac{1}{\min(eig)} < \rho_1 < \frac{1}{\max(eig)} = 1$, in which ρ_1 is a scalar spatial autoregressive parameter.

Given (1), logic dictates that

$$\rho_1 \mathbf{W}_N \mathbf{y}_t = \rho_1 \mathbf{W}_N \boldsymbol{\varsigma} + \rho_1 \mathbf{W}_N \gamma \mathbf{y}_{t-1}. \tag{2}$$

Subtracting (2) from (1) leads to another logically consistent expression in which the spatial dependence implied by (2) can be seen in (3) as an explicit cause of variation in $\mathbf{y}_t$. Thus

$$\mathbf{y}_t - \rho_1 \mathbf{W}_N \mathbf{y}_t = \boldsymbol{\varsigma} + \gamma \mathbf{y}_{t-1} - (\rho_1 \mathbf{W}_N \boldsymbol{\varsigma} + \rho_1 \mathbf{W}_N \gamma \mathbf{y}_{t-1}),$$

$$(\mathbf{I}_N - \rho_1 \mathbf{W}_N)\, \mathbf{y}_t = (\gamma \mathbf{I}_N - \rho_1 \gamma \mathbf{W}_N)\, \mathbf{y}_{t-1} + (\mathbf{I}_N - \rho_1 \mathbf{W}_N)\, \boldsymbol{\varsigma}.$$

[1] The matrix $\mathbf{W}_N$ comprises fixed (non-stochastic) non-negative values with zeros on the leading diagonal and its row and column sums are uniformly bounded in absolute value.

Writing $\theta = -\rho_1 \gamma$ gives

$$\mathbf{y}_t = \mathbf{B}_N^{-1} \left[\mathbf{C}_N \mathbf{y}_{t-1} + \mathbf{B}_N \boldsymbol{\varsigma} \right], \tag{3}$$

in which $\mathbf{B}_N = (\mathbf{I}_N - \rho_1 \mathbf{W}_N), \mathbf{C}_N = (\gamma \mathbf{I}_N + \theta \mathbf{W}_N)$ in which γ is the autoregressive time dependence parameter, ρ_1 is the spatial lag coefficient, θ is the time-space diffusion parameter and $\mathbf{I}_N$ is an identity matrix of order N. In order to solve equation (3), given appropriate parameter restrictions, equation (3) converges to $\mathbf{y}_T = (\mathbf{B}_N - \mathbf{C}_N)^{-1} \mathbf{B}_N \boldsymbol{\varsigma}$.

Introducing the additional covariates by writing $\mathbf{B}_N \boldsymbol{\varsigma} = (\mathbf{x}\boldsymbol{\beta})$, in which $\boldsymbol{\beta}$ is a $(k \times 1)$ vector, gives

$$\mathbf{y}_t = \mathbf{B}_N^{-1} \left[\mathbf{C}_N \mathbf{y}_{t-1} + \mathbf{x}\boldsymbol{\beta} \right].$$

In order to maintain dynamically stable simulations, following Elhorst (2001, 2014, p. 98), Parent and LeSage (2011, p. 478, 2012, p. 731) and Debarsy, Ertur and LeSage (2012, p. 162), requires the largest characteristic root ($eig_{\max}$) of $\mathbf{B}_N^{-1}\mathbf{C}_N$ to be less than 1. This restriction ensures that $\mathbf{y}_t$ converges to equilibrium levels $\mathbf{y}_T = (\mathbf{B}_N - \mathbf{C}_N)^{-1} (\mathbf{x}\boldsymbol{\beta})$.

Additional realism is introduced as follows. First, the restriction that $\theta = -\rho_1 \gamma$ is removed since ρ_1 and γ are unknown, so that θ is free to vary. However we anticipate that $\widehat{\theta} \approx -\widehat{\rho}_1 \widehat{\gamma}$. Second, the time invariant matrix $\mathbf{x}$ is replaced by time-varying matrix[2] $\mathbf{x}_t$. Third, unobservables are represented by the error term ε_t. Although the system may, depending on $\mathbf{B}_N^{-1}\mathbf{C}_N$, still tend towards equilibrium, equilibrium will be continuously disturbed and new equilibrium levels established as t varies. For simplicity of estimation, inter-regional connectivity is assumed to remain constant over the estimation period. These considerations lead to the time-space dynamic panel data model for $i = 1, \ldots, N$; $t = 1, \ldots, T$, which takes the form

$$y_{it} = \gamma y_{it-1} + \rho_1 \mathbf{w}_i \mathbf{y}_t + \theta \mathbf{w}_i \mathbf{y}_{t-1} + \mathbf{x}_{it} \boldsymbol{\beta} + \varepsilon_{it}, \tag{4}$$

where ε_{it} is an error term for region i at time t and $\mathbf{w}_i = (w_{i1}, \ldots, w_{iN})$ is a $(1 \times N)$ vector which corresponds to the ith row of the matrix $\mathbf{W}_N$. In contrast to the classical literature on panel data, grouping the data by periods rather than units is more convenient when we consider the spatial dimension. For each period t, we have

$$\begin{aligned} \mathbf{y}_t &= \gamma \mathbf{y}_{t-1} + \rho_1 \mathbf{W}_N \mathbf{y}_t + \theta \mathbf{W}_N \mathbf{y}_{t-1} + \mathbf{x}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \\ \left[-\mathbf{C}_N L + \mathbf{B}_N \right] \mathbf{y}_t &= \mathbf{x}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{5}$$

[2] We assume that the elements of $\mathbf{X}_t$ are uniformly bounded in absolute value.

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top$ is an $(N \times 1)$ vector and L is the time-lag operator, i.e. $L\mathbf{y}_t = \mathbf{y}_{t-1}$. $\mathbf{B}_N$ is a nonsingular matrix, and $\mathbf{B}_N^{-1}$ is uniformly bounded. For (5) and following Elhorst (2001, p. 131), Parent and LeSage (2011, p. 478, 2012, p. 731) and Debarsy, Ertur and LeSage (2012, p. 162), stationarity conditions are satisfied only if $\left|\mathbf{C}_N\mathbf{B}_N^{-1}\right| < 1$, which requires

$$\gamma + (\rho_1 + \theta)\, r_{\max} < 1 \quad \text{if } \rho_1 + \theta \geq 0, \tag{6}$$

$$\gamma + (\rho_1 + \theta)\, r_{\min} < 1 \quad \text{if } \rho_1 + \theta < 0, \tag{7}$$

$$\gamma - (\rho_1 - \theta)\, r_{\max} > -1 \quad \text{if } \rho_1 - \theta \geq 0, \tag{8}$$

$$\gamma - (\rho_1 - \theta)\, r_{\min} > -1 \quad \text{if } \rho_1 - \theta < 0, \tag{9}$$

where $r_{\min}$ and $r_{\max}$ are the minimum and maximum eigenvalues of $\mathbf{W}_N$.

We allow for a general spatial-autoregressive-moving-average errors process

$$\boldsymbol{\varepsilon}_{it} = \alpha\boldsymbol{\varepsilon}_{it-1} + \rho_2\mathbf{m}_i\boldsymbol{\varepsilon}_t + u_{it} - \lambda\mathbf{m}_i\mathbf{u}_t, \tag{10}$$

where ρ_2 and λ are respectively autoregressive and moving average parameters, and $\mathbf{m}_i = (m_{i1}, \ldots, m_{iN})$ is a $(1 \times N)$ vector which corresponds to the ith row of the spatial matrix $\mathbf{M}_N$. $\mathbf{M}_N$ is similar to $\mathbf{W}_N$ in that it defines the interaction assumed between the disturbances attributed to different regions, and it is often the case that $\mathbf{M}_N = \mathbf{W}_N$ is assumed. Estimating α, ρ_2 and λ jointly could be prohibitively difficult, and the most widely-used approach to modelling spatial error dependence is a restricted version of this in which $\alpha = 0$ and $\lambda = 0$, so that

$$\varepsilon_{it} = \rho_2\mathbf{w}_i\boldsymbol{\varepsilon}_t + u_{it}, \tag{11}$$

in which the autoregressive parameter space must be defined so that $(\mathbf{I}_N - \rho_2\mathbf{W}_N)$ is non-singular, see LeSage and Pace (2009, pp. 88-89). (11) is referred to as a Spatial AutoRegressive (SAR) process and implies complex interdependence between locations, so that a shock at location j is transmitted to all other locations. The SAR process is known to transmit the shocks globally.

In contrast, the Spatial Moving Average (SMA) process, which is the focus of this paper, is obtained by the restrictions $\alpha = 0$ and $\rho_2 = 0$, hence

$$\varepsilon_{it} = u_{it} - \lambda\mathbf{w}_i\mathbf{u}_t, \tag{12}$$

so that a shock at location j will only affect the directly interacting locations given by the non-zero elements in $\mathbf{W}_N$. Hence shock-effects are *local* rather than *global.* Regarding the error components, time dependency is introduced in the innovation u_{it} by specifying an unobserved permanent unit-specific error component μ_i together with the transient error component v_{it}. Thus, u_{it} follows an error component structure

$$u_{it} = \mu_i + v_{it}, \tag{13}$$

where μ_i is an individual specific time-invariant effect which is assumed to be $iid\left(0, \sigma_\mu^2\right)$, and v_{it} is a remainder effect which is assumed to be $iid\left(0, \sigma_v^2\right)$. μ_i and v_{it} are independent of each other and among themselves. Combining (12) and (13), we obtain the SMA-RE specification of the disturbance ε_{it}. For a cross-section t, we have

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_N \mathbf{u}_t, \tag{14}$$

with

$$\mathbf{u}_t = \boldsymbol{\mu} + \mathbf{v}_t, \tag{15}$$

where $\mathbf{H}_N = (\mathbf{I}_N - \lambda \mathbf{W}_N)$, and $\mathbf{u}_t = (u_{1t}, \ldots, u_{Nt})^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^\top$, $\mathbf{v}_t = (v_{1t}, \ldots, v_{Nt})^\top$ are three vectors of dimension $(N \times 1)$.

By rewriting the model (5) as

$$\mathbf{y}_t = \mathbf{B}_N^{-1} \mathbf{C}_N \mathbf{y}_{t-1} + \mathbf{B}_N^{-1} \mathbf{x}_t \boldsymbol{\beta} + \mathbf{B}_N^{-1} \boldsymbol{\varepsilon}_t, \tag{16}$$

the matrix of partial derivatives of $\mathbf{y}_t$ with respect to the kth explanatory variable of $\mathbf{x}_t$ in unit 1 up to unit N at time t is given by

$$\begin{bmatrix} \frac{\partial \mathbf{y}}{\partial x_{1k}} & \cdots & \frac{\partial \mathbf{y}}{\partial x_{Nk}} \end{bmatrix}_t = \beta_k \mathbf{B}_N^{-1}, \tag{17}$$

where

$$\mathbf{B}_N^{-1} = (\mathbf{I}_N - \rho_1 \mathbf{W}_N)^{-1} = \mathbf{I}_N + \rho_1 \mathbf{W}_N + \rho_1^2 \mathbf{W}_N^2 + \rho_1^3 \mathbf{W}_N^3 + \cdots. \tag{18}$$

The expression (17) denotes the effect of a change of an explanatory variable in a particular spatial unit on the dependent variable of all other units in the short-term. Similarly, the long-term effects are obtained by

$$\begin{aligned} \begin{bmatrix} \frac{\partial \mathbf{y}}{\partial x_{1k}} & \cdots & \frac{\partial \mathbf{y}}{\partial x_{Nk}} \end{bmatrix} &= \beta_k \left[-\mathbf{C}_N + \mathbf{B}_N\right]^{-1} \\ &= \beta_k \left[(1-\gamma) \mathbf{I}_N - (\rho_1 + \theta) \mathbf{W}_N\right]^{-1} \\ &= \beta_k^* \mathbf{B}_N^{*-1}, \end{aligned} \tag{19}$$

where

$$\mathbf{B}_N^{*-1} = (\mathbf{I}_N - \rho_1^* \mathbf{W}_N)^{-1}, \; \rho_1^* = \frac{\rho_1 + \theta}{1-\gamma} \text{ and } \beta_k^* = \frac{\beta_k}{1-\gamma}. \tag{20}$$

The expressions in (17) and (19) illustrate that they depend respectively on one and two global spatial multiplier matrices, respectively. Moreover, short-term indirect effects do not occur if $\rho_1 = 0$, while long-term indirect effects do not occur if $\rho_1 = -\theta$.

3 A Four Stage Spatial GM Estimator

The presence of a spatial lag, a time-lagged and a time-space lag dependent variable renders the usual panel data estimators that ignore this spatial correlation biased and inconsistent. In this context, Instrumental Variables (IV) or GM estimators are required. These estimators assume much weaker assumptions about the initial conditions compared to those of Maximum Likelihood (ML). The consistency of ML estimators depends on the initial conditions and on the way in which the time dimension T and number of cross-sections N tend to infinity. For dynamic panel data models, Bond (2002) argues that the distribution of the dependent variable depends in a non-negligible way on what is assumed about the distribution of the initial conditions. For example, the initial condition could be stochastic or non-stochastic, correlated or uncorrelated with the individual effects, or have to satisfy stationarity properties. Different assumptions about the nature of the initial conditions will lead to different likelihood functions, and the resulting ML estimators can be inconsistent when the assumptions on the initial conditions are misspecified, see Hsiao (2003, pp. 80-135) for more details. IV or GM estimators require much weaker assumptions. Following Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991), the individual effect μ_i in (13), which is correlated with the spatial lag and time-lagged dependent variable, is eliminated by first-differencing the model (4) yielding

$$\Delta y_{it} = \gamma \Delta y_{it-1} + \rho_1 \mathbf{w}_i \Delta \mathbf{y}_t + \theta \mathbf{w}_i \Delta \mathbf{y}_{t-1} + \mathbf{\Delta x}_{it} \boldsymbol{\beta} + \Delta \varepsilon_{it}, \tag{21}$$

$t = 2, \ldots, T$, and for a cross-section t, we have

$$\mathbf{\Delta y}_t = \gamma \mathbf{\Delta y}_{t-1} + \rho_1 \mathbf{W}_N \mathbf{\Delta y}_t + \theta \mathbf{W}_N \mathbf{\Delta y}_{t-1} + \mathbf{\Delta x}_t \boldsymbol{\beta} + \mathbf{\Delta} \boldsymbol{\varepsilon}_t. \tag{22}$$

Following (14) and (15), (22) can be written as

$$\mathbf{\Delta y}_t = \gamma \mathbf{\Delta y}_{t-1} + \rho_1 \mathbf{W}_N \mathbf{\Delta y}_t + \theta \mathbf{W}_N \mathbf{\Delta y}_{t-1} + \mathbf{\Delta x}_t \boldsymbol{\beta} + \mathbf{H}_N \mathbf{\Delta v}_t. \tag{23}$$

Following Arellano and Bond (1991), we can define a GM estimator based on the assumption of no correlation between the first-differenced disturbances and earlier time-lagged levels of the dependent variable, y_{it-1}. This yields the following moment conditions:

$$E(y_{il}\Delta v_{it}) = 0, \quad \forall i,\ l = 0, 1, \ldots, t-2;\ t = 2, \ldots, T. \tag{24}$$

If we assume that the explanatory variables $x_{k,im}$ are strictly exogenous[3], the following additional moment conditions can be used

$$E(x_{k,im}\Delta v_{it}) = 0, \quad \forall i, k,\ m = 1, \ldots, T;\ t = 2, \ldots, T. \tag{25}$$

If, $x_{k,im}$ is weakly exogenous then the associated moment conditions are

$$E(x_{k,im}\Delta v_{it}) = 0, \quad \forall i, k,\ m = 1, \ldots, t-1;\ t = 2, \ldots, T. \tag{26}$$

Moreover, to take into account the endogeneity of the spatial lag, we can use spatially weighted earlier time-lagged levels of the dependent and explanatory variables as instruments. This strategy is associated with the following moments conditions:

$$\begin{aligned} E(\mathbf{w}_i\mathbf{y}_l\Delta v_{it}) &= 0, \quad l = 0, 1, \ldots, t-2;\ t = 2, \ldots, T, & (27)\\ E(\mathbf{w}_i^*\mathbf{y}_l\Delta v_{it}) &= 0, \quad l = 0, 1, \ldots, t-2;\ t = 2, \ldots, T, & (28)\\ E(\mathbf{w}_i\mathbf{x}_{k,m}\Delta v_{it}) &= 0, \quad \forall i, k,\ m = 1, \ldots, T;\ t = 2, \ldots, T, & (29)\\ E(\mathbf{w}_i^*\mathbf{x}_{k,m}\Delta v_{it}) &= 0, \quad \forall i, k,\ m = 1, \ldots, T;\ t = 2, \ldots, T, & (30) \end{aligned}$$

where $\mathbf{w}_i^* = (w_{i1}^*, \ldots, w_{iN}^*)$ is a $(1 \times N)$ vector which corresponds to the ith row of the matrix $\mathbf{W}_N^2$. Considering the weakly exogenous assumption of $x_{k,im}$, the moment conditions (29) and (30) are replaced by the following:

$$\begin{aligned} E(\mathbf{w}_i\mathbf{x}_{k,m}\Delta v_{it}) &= 0, \quad \forall i, k,\ m = 1, \ldots, t-1;\ t = 2, \ldots, T, & (31)\\ E(\mathbf{w}_i^*\mathbf{x}_{k,m}\Delta v_{it}) &= 0, \quad \forall i, k,\ m = 1, \ldots, t-1;\ t = 2, \ldots, T. & (32) \end{aligned}$$

Let us define the matrix $\mathbf{Z}$ which contains the non-spatial instruments (i.e. related to the conditions (24) and (25)) as

$$\mathbf{Z} = diag\left(\mathbf{Z}_2, \mathbf{Z}_3, \ldots, \mathbf{Z}_T\right), \tag{33}$$

[3]For other cases, see Bond (2002, p. 152), Bouayad-Agha and Védrine (2010, p. 211).

where

$$\mathbf{Z}_t = (\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{t-2}, \mathbf{x}_1, \ldots, \mathbf{x}_T) \quad (34)$$

is an $(N \times [(t-1) + KT])$ matrix of instruments at time t, $\mathbf{y}_l$ is a vector of dimension $(N \times 1)$ and $\mathbf{x}_r$ is a matrix of dimension $(N \times K)$. Moreover, we can define a matrix $\mathbf{Z}^s$ which contains the spatial instruments (i.e. related to the conditions (27), (28), (29) and (30)) as

$$\mathbf{Z}^s = diag\left(\mathbf{Z}_2^s, \mathbf{Z}_3^s, \ldots, \mathbf{Z}_T^s\right), \quad (35)$$

with

$$\mathbf{Z}_t^s = \left(\mathbf{y}_t^s, \mathbf{y}_t^{s*}, \mathbf{x}^s, \mathbf{x}^{s*}\right), \quad (36)$$

where $\mathbf{y}_t^s = \left(\mathbf{y}_0^s, \mathbf{y}_1^s, \ldots, \mathbf{y}_{t-2}^s\right)$, $\mathbf{y}_t^{s*} = \left(\mathbf{y}_0^{s*}, \mathbf{y}_1^{s*}, \ldots, \mathbf{y}_{t-2}^{s*}\right)$, $\mathbf{x}^s = (\mathbf{x}_1^s, \ldots, \mathbf{x}_T^s)$, $\mathbf{x}^{s*} = (\mathbf{x}_1^{s*}, \ldots, \mathbf{x}_T^{s*})$, and

$$\mathbf{y}_l^s = \begin{pmatrix} \mathbf{w}_1\mathbf{y}_l \\ \mathbf{w}_2\mathbf{y}_l \\ \vdots \\ \mathbf{w}_N\mathbf{y}_l \end{pmatrix} = \mathbf{W}_N\mathbf{y}_l,\ \mathbf{y}_l^{s*} = \begin{pmatrix} \mathbf{w}_1^*\mathbf{y}_l \\ \mathbf{w}_2^*\mathbf{y}_l \\ \vdots \\ \mathbf{w}_N^*\mathbf{y}_l \end{pmatrix} = \mathbf{W}_N^2\mathbf{y}_l, \quad (37)$$

where $\mathbf{y}_l = (y_{1l}, \ldots, y_{Nl})^\top$ and

$$\mathbf{x}_r^s = \begin{pmatrix} \mathbf{w}_1\mathbf{x}_{1r} & \mathbf{w}_1\mathbf{x}_{2r} & \cdots & \mathbf{w}_1\mathbf{x}_{Kr} \\ \mathbf{w}_2\mathbf{x}_{1r} & \mathbf{w}_2\mathbf{x}_{2r} & \cdots & \mathbf{w}_2\mathbf{x}_{Kr} \\ \vdots & \vdots & & \vdots \\ \mathbf{w}_N\mathbf{x}_{1r} & \mathbf{w}_N\mathbf{x}_{2r} & \cdots & \mathbf{w}_N\mathbf{x}_{Kr} \end{pmatrix} = \mathbf{W}_N\mathbf{x}_r, \quad (38)$$

$$\mathbf{x}_r^{s*} = \begin{pmatrix} \mathbf{w}_1^*\mathbf{x}_{1r} & \mathbf{w}_1^*\mathbf{x}_{2r} & \cdots & \mathbf{w}_1^*\mathbf{x}_{Kr} \\ \mathbf{w}_2^*\mathbf{x}_{1r} & \mathbf{w}_2^*\mathbf{x}_{2r} & \cdots & \mathbf{w}_2^*\mathbf{x}_{Kr} \\ \vdots & \vdots & & \vdots \\ \mathbf{w}_N^*\mathbf{x}_{1r} & \mathbf{w}_N^*\mathbf{x}_{2r} & \cdots & \mathbf{w}_N^*\mathbf{x}_{Kr} \end{pmatrix} = \mathbf{W}_N^2\mathbf{x}_r, \quad (39)$$

where $\mathbf{x}_{kr} = (x_{k1r}, \ldots, x_{kNr})^\top$, $k = 1, \ldots, K$. If we stack the matrices $\mathbf{Z}$ and $\mathbf{Z}^s$, we obtain the valid instruments for the model (21), namely $\mathbf{Z}^*$. Moreover, we use the weight matrix of moments

$$\mathbf{A}_N = \left[E\left[\mathbf{Z}^{*\top}\left(\boldsymbol{\Delta\varepsilon}\right)\left(\boldsymbol{\Delta\varepsilon}\right)^\top\mathbf{Z}^*\right]\right]^{-1}, \quad (40)$$

with

$$E\left[(\boldsymbol{\Delta\varepsilon})(\boldsymbol{\Delta\varepsilon})^{\top}\right] = \sigma_v^2\left(\boldsymbol{\Lambda}\otimes\mathbf{H}_N\mathbf{H}_N^{\top}\right), \tag{41}$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}, \tag{42}$$

is the variance-covariance matrix of the remainder MA(1) unit roots error which is of dimension $(T-1\times T-1)$ used by Arellano and Bond (1991) in their one-step GMM estimator.

A consistent estimate of the parameters λ and σ_v^2 can be obtained using a GM approach in the spirit of Fingleton (2008a) for the static spatial lag model including a SMA-RE process on the disturbances. In fact, Fingleton (2008a) extended the GM procedure from panel data proposed by KKP (2007) for the SAR-RE case to the SMA-RE one. Here, the main difference is that we base the estimation of λ and σ_v^2 on the first differences of the errors to account for the dynamics and to get rid of the individual effects. Following the derivation of the moment conditions and ignoring the expectations of each term, the system involving λ and σ_v^2 can be expressed as

$$\boldsymbol{\Gamma}_N\boldsymbol{\phi} - \mathbf{g}_N = \mathbf{0}, \tag{43}$$

where

$$\begin{aligned} \boldsymbol{\Gamma}_N &= \begin{pmatrix} 2N(T-1) & 2(T-1)t_1 & 0 \\ 2(T-1)t_1 & 2(T-1)t_2 & 4(T-1)t_3 \\ 0 & 2(T-1)t_3 & 2(T-1)(t_1+t_4) \end{pmatrix}, \boldsymbol{\phi} = \begin{pmatrix} \sigma_v^2 \\ \lambda^2\sigma_v^2 \\ -\lambda\sigma_v^2 \end{pmatrix}, \\ \mathbf{g}_N &= \begin{pmatrix} \boldsymbol{\Delta\varepsilon}^{\top}\boldsymbol{\Delta\varepsilon} \\ \boldsymbol{\Delta\overline{\varepsilon}}^{\top}\boldsymbol{\Delta\overline{\varepsilon}} \\ \boldsymbol{\Delta\overline{\varepsilon}}^{\top}\boldsymbol{\Delta\varepsilon} \end{pmatrix}, \end{aligned} \tag{44}$$

with $t_1 = \mathrm{tr}\left(\mathbf{W}_N^{\top}\mathbf{W}_N\right)$, $t_2 = \mathrm{tr}\left((\mathbf{W}_N^2)^{\top}\mathbf{W}_N^2\right)$, $t_3 = \mathrm{tr}\left((\mathbf{W}_N^2)^{\top}\mathbf{W}_N\right)$, $t_4 = \mathrm{tr}(\mathbf{W}_N^2)^{\top}$. The GM estimators of λ and σ_ν^2 are the solution of the sample moments using nonlinear least squares on equation (43). See Fingleton

(2008a) for the static spatial lag model including a SMA-RE process for the disturbances. Our spatial GM procedure comprises of the following four steps:

- In the first step, we use an IV or GM estimator to get consistent estimates of γ, ρ_1, θ and $\boldsymbol{\beta}$.

- In the second step, the IV or GM residuals are used to obtain consistent estimates of the moving average parameter λ and the variance σ_v^2.

- In the third step, we compute the preliminary one-stage consistent Spatial GM estimator which is given by

$$\widehat{\boldsymbol{\delta}}_1 = (\boldsymbol{\Delta}\widetilde{\mathbf{X}}^\top \mathbf{Z}^* \widehat{\mathbf{A}}_N \mathbf{Z}^{*\top} \boldsymbol{\Delta}\widetilde{\mathbf{X}})^{-1} \boldsymbol{\Delta}\widetilde{\mathbf{X}}^\top \mathbf{Z}^* \widehat{\mathbf{A}}_N \mathbf{Z}^{*\top} \boldsymbol{\Delta}\mathbf{y}, \tag{45}$$

  where $\boldsymbol{\Delta}\widetilde{\mathbf{X}} = \left(\boldsymbol{\Delta}\mathbf{y}_{-1}, (\mathbf{I}_{T-1} \otimes \mathbf{W}_N)\, \boldsymbol{\Delta}\mathbf{y}, (\mathbf{I}_{T-1} \otimes \mathbf{W}_N)\, \boldsymbol{\Delta}\mathbf{y}_{-1}, \boldsymbol{\Delta}\mathbf{x}\right)$, $\boldsymbol{\delta}_1^\top = \left(\gamma, \rho_1, \theta, \boldsymbol{\beta}^\top\right)$ and

$$\widehat{\mathbf{A}}_N = \left[\mathbf{Z}^{*\top} \left(\boldsymbol{\Lambda} \otimes \widehat{\mathbf{H}}_N \widehat{\mathbf{H}}_N^\top\right) \mathbf{Z}^*\right]^{-1}, \tag{46}$$

  with $\widehat{\mathbf{H}}_N = (\mathbf{I}_N - \widehat{\lambda}\mathbf{W}_N)$.

- In the fourth step, following Arellano and Bond (1991), we replace (46) by its robust version

$$\mathbf{V}_N = \left[\mathbf{Z}^{*\top} \left(\mathbf{I}_{T-1} \otimes \widehat{\mathbf{H}}_N\right) \boldsymbol{\Phi} \left(\mathbf{I}_{T-1} \otimes \widehat{\mathbf{H}}_N^\top\right) \mathbf{Z}^*\right]^{-1} \tag{47}$$

  where

$$\boldsymbol{\Phi} = \left[(\boldsymbol{\Delta}\mathbf{v})\,(\boldsymbol{\Delta}\mathbf{v})^\top\right] \odot (\mathbf{J}_{T-1} \otimes \mathbf{I}_N)\,, \tag{48}$$

  and $\mathbf{J}_{T-1} = \left(\boldsymbol{\iota}_{T-1}\boldsymbol{\iota}_{T-1}^\top\right)$, $\boldsymbol{\iota}_{T-1}$ is a vector of ones of dimension $(T-1 \times 1)$. To operationalize this estimator, $\boldsymbol{\Delta}\mathbf{v}$ is replaced by differenced residuals obtained from the preliminary one-stage consistent Spatial GM estimator (45). The resulting estimator is the two-stage Spatial GM estimator

$$\widehat{\boldsymbol{\delta}}_2 = (\boldsymbol{\Delta}\widetilde{\mathbf{X}}^\top \mathbf{Z}^* \widehat{\mathbf{V}}_N \mathbf{Z}^{*\top} \boldsymbol{\Delta}\widetilde{\mathbf{X}})^{-1} \boldsymbol{\Delta}\widetilde{\mathbf{X}}^\top \mathbf{Z}^* \widehat{\mathbf{V}}_N \mathbf{Z}^{*\top} \boldsymbol{\Delta}\mathbf{y}. \tag{49}$$

4 Prediction

In 2014, Baltagi, Fingleton and Pirotte argued that the derivation of the predictor for a dynamic autoregressive spatial panel data model is more complicated because the time/space lags of the dependent variable are correlated with the disturbances. Thus, the Goldberger (1962) framework, which assumes no correlation between the regressors and the error term, is not applicable in this context. The predictor proposed by Baltagi, Fingleton and Pirotte (2014), under the restrictive assumption of no time-space dependence and SAR-RE disturbances structure, depends on the process that generates the initial values and this may be difficult to handle in practice. So, here, considering (4) without any restriction on time-space dependence and a SMA-RE process on the disturbances $\boldsymbol{\varepsilon}$ which captures local spillovers, we propose a tractable approach which does not depend explicitly on the initial values. In the first step, the individual effects are estimated from the residuals observed over time. To obtain this, we commence by using a single cross-section equation at time t, in particular

$$\mathbf{y}_t = \rho_1 \mathbf{W}_N \mathbf{y}_t + \mathbf{C}_N \mathbf{y}_{t-1} + \mathbf{x}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t. \tag{50}$$

So that

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_N \mathbf{u}_t = \mathbf{y}_t - \rho_1 \mathbf{W}_N \mathbf{y}_t - \mathbf{C}_N \mathbf{y}_{t-1} - \mathbf{x}_t \boldsymbol{\beta}, \tag{51}$$

and

$$\mathbf{u}_t = \boldsymbol{\mu} + \mathbf{v}_t = \mathbf{H}_N^{-1} \left(\mathbf{y}_t - \rho_1 \mathbf{W}_N \mathbf{y}_t - \mathbf{C}_N \mathbf{y}_{t-1} - \mathbf{x}_t \boldsymbol{\beta} \right), \tag{52}$$

$$\boldsymbol{\mu} = \mathbf{H}_N^{-1} \left(\mathbf{B}_N \mathbf{y}_t - \mathbf{C}_N \mathbf{y}_{t-1} - \mathbf{x}_t \boldsymbol{\beta} \right) - \mathbf{v}_t, \tag{53}$$

with $\mathbf{v}_t \sim N\left(\mathbf{0}, \sigma_v^2 \mathbf{I}_N\right)$. In order to calculate $\widehat{\boldsymbol{\mu}}$, one uses observed data for the sequence of $\mathbf{y}$s in equation (53) together with the parameter estimates $(\widehat{\gamma},\widehat{\rho}_1,\widehat{\theta},\widehat{\boldsymbol{\beta}},\widehat{\lambda},\widehat{\sigma}_v^2,\widehat{\sigma}_\mu^2)$, using the data for the period $t = 2, \ldots, T$ (assuming that $\mathbf{y}_0$ is not available) on each occasion drawing an $(N \times 1)$ vector $\mathbf{v}_t$ at random from the $N\left(\mathbf{0}, \widehat{\sigma}_v^2 \mathbf{I}_N\right)$ distribution. This gives $T-1$ different estimates of $\widehat{\boldsymbol{\mu}}$, so we take the time mean as an estimate of the time-invariant $(N \times 1)$ vector $\boldsymbol{\mu}$, given $E[\mathbf{v}_t] = \mathbf{0}$. Also, the estimate is scaled so that its variance is equal to $\widehat{\sigma}_\mu^2$. So, here, it is necessary to estimate σ_μ^2 and not only λ and σ_v^2. These two parameters are sufficient to compute the variance-covariance matrix of the GM estimator. This means that instead of using the first differences moment conditions approach, see (43) and (44), we have to consider the moment conditions in levels described by Fingleton (2008a). The IV or GM level residuals are used to obtain consistent estimates of the moving average parameter λ and the variances σ_v^2 and σ_μ^2. In the second step, this estimated $\boldsymbol{\mu}$, denoted by $\overline{\boldsymbol{\mu}}$, is then used in a bootstrap forecast approach considering observed values $\mathbf{y}_T$ for the first forecast $\widehat{\mathbf{y}}_{T+1}$, and estimates of future $\widehat{\mathbf{y}}$ from then on. Thus, the predictor is given by

$$\widehat{\mathbf{y}}_t = \widehat{\mathbf{B}}_N^{-1}\left[\widehat{\mathbf{C}}_N\widehat{\mathbf{y}}_{t-1} + \mathbf{x}_t\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{H}}_N\overline{\boldsymbol{\mu}}\right], \tag{54}$$

which solves recursively over $t = T+1, T+2, \ldots, T+\tau$, $\tau \geq 1$ with $\mathbf{y}_T$ replacing $\widehat{\mathbf{y}}_T$ for $t = T+1$. In the Monte Carlo experiments that follow, this predictor (54) is compared to other suboptimal predictors which correspond to misspecified dynamic estimators.

5 Monte Carlo Design

We assume that the dependent variable y_{it}, $i = 1, \ldots, N$; $t = 1, \ldots, T$, is given from a spatial dynamic panel data model of the form

$$y_{it} = a + \gamma y_{it-1} + \rho_1 \mathbf{w}_i \mathbf{y}_t + \theta \mathbf{w}_i \mathbf{y}_{t-1} + \beta x_{it} + \varepsilon_{it}, \tag{55}$$

where the disturbance ε_{it} follows a SMA process

$$\varepsilon_{it} = u_{it} - \lambda \mathbf{w}_i \mathbf{u}_t, \tag{56}$$

$\mathbf{w}_i = (w_{i1}, \ldots, w_{iN})$ is a $(1 \times N)$ vector which corresponds the ith row of the matrix $\mathbf{W}_N$, $\lambda = -0.4$, since this equates to positive dependence, and u_{it} has an error component structure

$$u_{it} = \mu_i + v_{it}, \tag{57}$$

with $\mu_i \sim iid.N\left(0, \sigma_\mu^2\right)$, $v_{it} \sim iid.N\left(0, \sigma_v^2\right)$ and $\left(\sigma_\mu^2, \sigma_v^2\right) = (0.8, 0.2), (0.2, 0.8)$. The explanatory variable x_{it} is generated as

$$x_{it} = \delta x_{it-1} + \xi_{it}, \tag{58}$$

with $\delta = 0.8$, $\xi_{it} \sim iid.N\left(0, \sigma_\xi^2\right)$ and $\sigma_\xi^2 = 10$. We set $x_{i,-50} = 0$ and generate x_{it} for $t = -49, -48, \ldots, T$. For y_{it}, we also set $y_{i,-50} = 0$ and discarded the first 51 observations, using the observations $t = 1$ through T for estimation (we assume that y_0 is non available in order to be closer to the empirical applications). Different sample sizes are considered $N = 100$, 200 and $T = 12$.

For the coefficients of (55), we take $(a, \gamma, \rho_1, \theta, \beta) = (1, 0.2, 0.8, -0.2, 1)$, $(1, 0.8, 0.4, -0.3, 1)$, $(1, 0.7, 0.8, -0.6, 1)$ and $(1, 0.8, 0.8, -0.7, 1)$. Moreover, following Baltagi and Yang (2013), the spatial matrix $\mathbf{W}_N$ has a rook contiguity structure[4]. This matrix is generated as follows: (i) index the N spatial units by $1, ..., N$. Randomly permute these indices and then allocate them into a lattice of $R \times M(\geq N)$ squares. (ii) let $w_{i,j} = 1$ if the index j is in a square which is on the immediate left, or right, or above, or below the square which contains the index i, otherwise $w_{i,j} = 0$; and (iii) divide each element of $\mathbf{W}_N$ by its row sum. For all experiments, $5,000$ replications are performed. We compute the mean, standard deviation, bias and RMSE of the coefficients $\widehat{\gamma}$, $\widehat{\rho}_1$, $\widehat{\theta}$, $\widehat{\beta}$, $\widehat{\lambda}$ and the average direct, indirect and total short-term and long-term effects using respectively (17) and (19). Following KKP (2007), we adopt a measure of dispersion which is closely related to the standard measure of root mean square error (RMSE), but is based on quantiles. It is defined as

$$\text{RMSE} = \left[bias^2 + \left(\frac{IQ}{1.35} \right)^2 \right]^{1/2}, \tag{59}$$

where $bias$ is the difference between the median and the true value of the parameter, and IQ is the interquantile range defined as $c_1 - c_2$ where c_1 is the 0.75 quantile and c_2 is the 0.25 quantile. Clearly, if the distribution is normal the median is the mean and, aside from a slight rounding error, $IQ/1.35$ is the standard deviation. In this case, the measure (59) reduces to the standard RMSE.

We compare the performance of 5 estimators in our Monte Carlo experiments. These are as follows:

1. Ordinary Least Squares (OLS) which does not deal with the endogeneity of the spatial lag $\mathbf{W}_N\mathbf{y}$, the time lag $\mathbf{y}_{-1}$ and the time-space lag $\mathbf{W}_N\mathbf{y}_{-1}$. OLS also ignores the SMA-RE process generating the disturbances.

2. The Within estimator which wipes out the individual effects, but otherwise does not deal with the endogeneity of the spatial lag $\mathbf{W}_N\mathbf{y}$, the time lag $\mathbf{y}_{-1}$ and the time-space lag $\mathbf{W}_N\mathbf{y}_{-1}$ nor the SMA process for the disturbances.

[4]Following Kelejian and Prucha (1999), we have also considered the spatial matrix which is labelled as "j ahead and j behind" with the non-zero elements being $1/2j$, $j = 2$, i.e. $\mathbf{W}(2, 2)$. The results are similar.

3. The Arellano and Bond (1991) GMM estimator which eliminates the individual effects by differencing, and handles the presence of the lagged dependent variable by using the orthogonality conditions (24) and (25). However, this estimator ignores the spatial lag $\mathbf{W}_N\mathbf{y}$, the time-space lag $\mathbf{W}_N\mathbf{y}_{-1}$ and the SMA process for the disturbances.

4. GM-TS-RE is an estimator that uses the orthogonality conditions (24) and (25) of Arellano and Bond (1991) as well as the spatial orthogonality conditions (27), (28), (29) and (30). However, this estimator ignores the SMA process for the disturbances.

5. GM-TS-SMA-RE is an estimator that uses the orthogonality conditions (24) and (25) of Arellano and Bond (1991) as well as the spatial orthogonality conditions (27), (28), (29) and (30) used by GM-TS-RE, but it also accounts for the SMA structure of the disturbances using a similar approach to that of Fingleton (2008*a)*, see Section 3.

Last, for each experiment, we use the post sample RMSE criterion and compute the out of sample forecast errors for each predictor associated with the alternative estimators for one to five step ahead forecasts. In order to do this, we generate 5 more time periods for each individual (i.e. $T+\tau$, $\tau=1,\ldots,5$) which are not used in the estimation. An average RMSE is also calculated across all individuals at different forecast horizons.

6 Monte Carlo Results

The estimates obtained *via* the various estimators are summarised in Tables 1 to 4 ($N=100$) and 6 to 9 ($N=200$). They report the mean, bias and RMSE of $\widehat{\gamma}$, $\widehat{\rho}_1$, $\widehat{\lambda}$, $\widehat{\theta}$ and $\widehat{\beta}$ given various assumed true values for the data generating process. The difference between Tables 1 and 2 (resp. Tables 5 and 6) is that in Table 2 (resp. Table 6) we assume greater individual heterogeneity where $\sigma^2_\mu=0.8$ rather than 0.2. This is done holding the total variance of the disturbances constant, i.e., $\sigma^2_\mu+\sigma^2_v=1$ in both Tables. Tables 1 and 2 (resp. Tables 5 and 6) show clearly that the GM-TS-SMA-RE estimator has the lowest RMSEs for all experiments considered. This is especially true when the individual heterogeneity and is high, i.e. when $\left(\sigma^2_\mu,\sigma^2_v\right)=(0.8,0.2)$ rather than $\left(\sigma^2_\mu,\sigma^2_v\right)=(0.2,0.8)$. When N increases, comparing Tables 1 to 4 to their respective counterparts Tables 6 to 9, the spatial estimators perfom better in terms of RMSE. Table 11 reports the mean RMSE and it clearly shows that GM-TS-SMA-RE performs well compared to the other estimators. Tables 3 and 4 (resp. Tables 7 and 8) give the RMSE variation for direct, indirect and total long and short-term effects in the case of the two GM-based spatial estimators.

[**INSERT TABLES 1 TO 10**]

Table 11 - Mean RMSE for the non-spatial and spatial estimators

	Mean (RMSE)	
	(Tables 1 & 2)	(Tables 6 & 7)
Non-spatial estimators		
- *OLS*	0.0343	0.0299
- *Within*	0.0165	0.0188
- *GMM**	0.1102	0.1643
Spatial estimators		
- *GM-TS-RE*	0.0189	0.0134
- *GM-TS-SMA-RE*	0.0172	0.0109
Total	0.0394	0.0475

* *This estimator does not consider RMSEs of the coefficients* $\widehat{\rho}_1$ *and* $\widehat{\theta}$.

In order to illustrate the comparative prediction performance, Figures 1a to 5a and 1b to 5b show outcomes from specific parameter assumptions considering $(N, T) = (200, 12)$. In particular, Figures 1a to 5a, use $(a, \gamma, \rho_1, \theta, \beta, \lambda) = (1, 0.2, 0.8, -0.2, 1, -0.4)$, $\left(\sigma^2_\mu, \sigma^2_v\right) = (0.2, 0.8)$ and $\tau = 1, \ldots, 5$ whereas Figures 1b to 5b use the same values of $(a, \gamma, \rho_1, \theta, \beta, \lambda)$, but allow for a higher level of individual heterogeneity, i.e. $\left(\sigma^2_\mu, \sigma^2_v\right) = (0.8, 0.2)$. Figure 1$b$ shows the good performance associated with the GM-TS-SMA-RE compared to the other figures. This figure also shows that the higher the individual heterogeneity, i.e. $\left(\sigma^2_\mu, \sigma^2_v\right) = (0.8, 0.2)$, the better the forecasting performance comparing to Figure 1a. This is shown more formally in Table 10 which summarizes the forecasting performance results in RMSE terms and shows that GM-TS-SMA-RE gives the lowest RMSE in all cases considered.

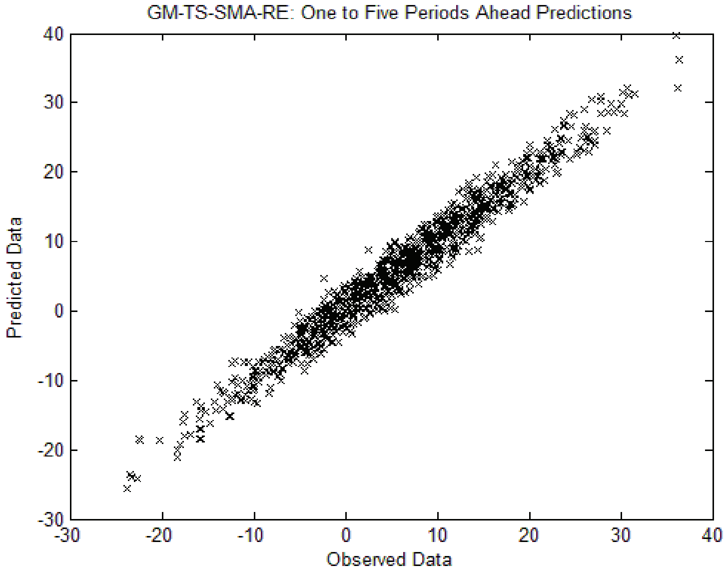

Figure 1a: $(\sigma_\mu^2, \sigma_v^2) = (0.2, 0.8)$

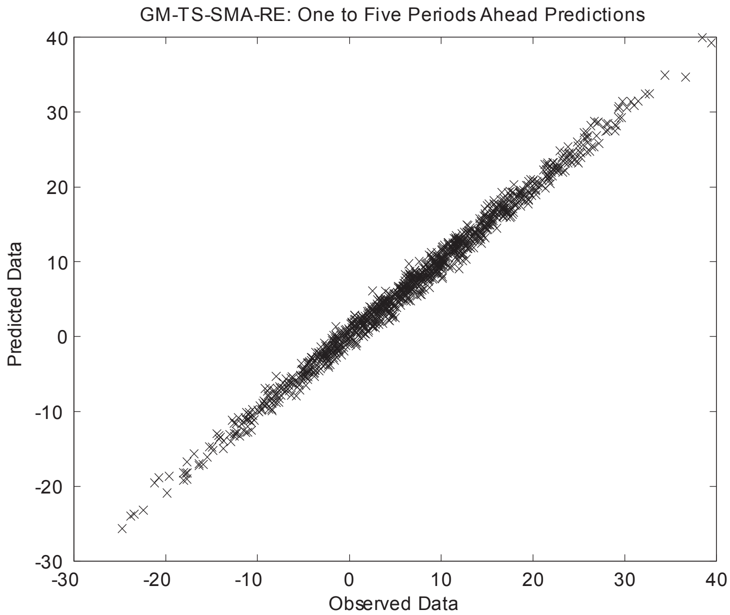

Figure 1b: $(\sigma_\mu^2, \sigma_v^2) = (0.8, 0.2)$

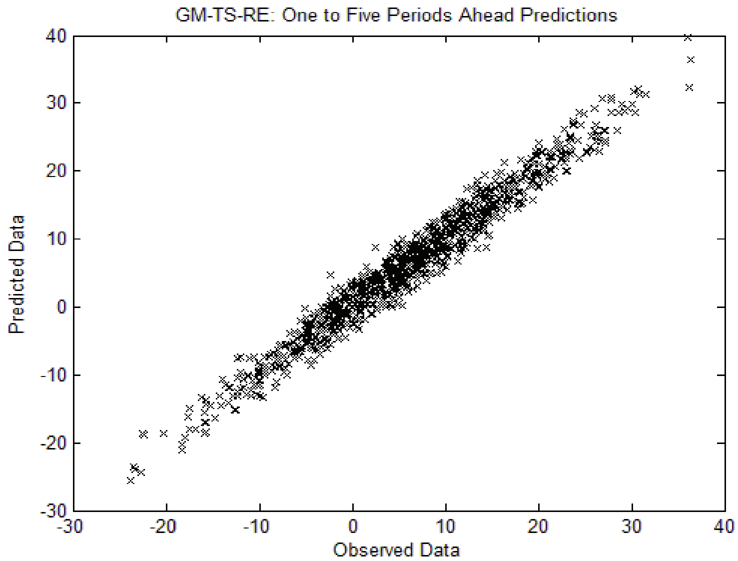

Figure 2a: $\left(\sigma_\mu^2, \sigma_v^2\right) = (0.2, 0.8)$

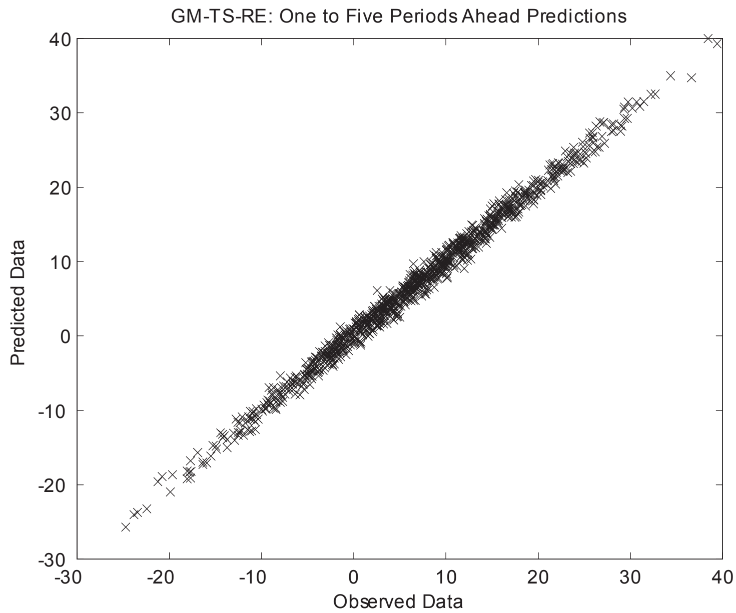

Figure 2b: $\left(\sigma_\mu^2, \sigma_v^2\right) = (0.8, 0.2)$

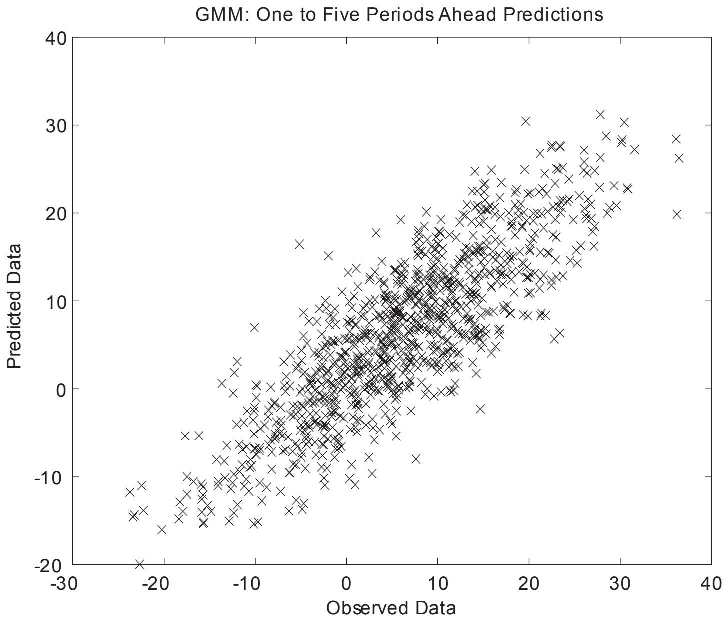

Figure 3a: $\left(\sigma_\mu^2, \sigma_v^2\right) = (0.2, 0.8)$

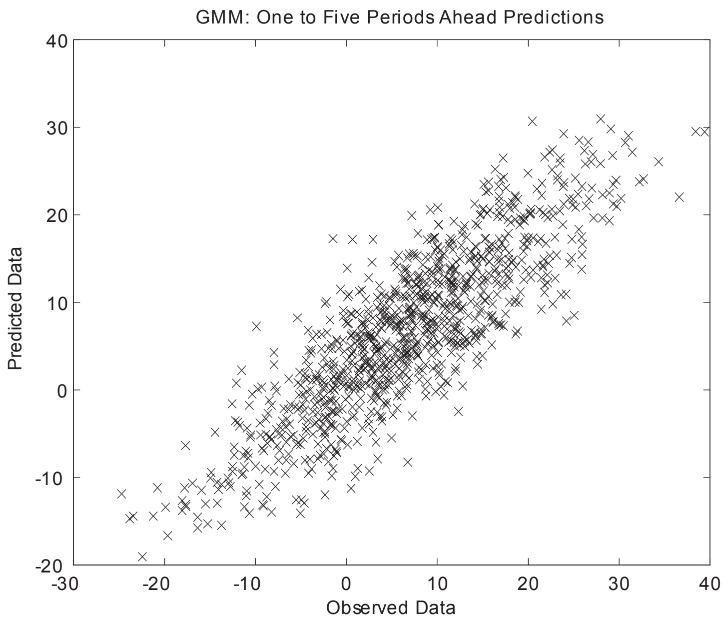

Figure 3b: $\left(\sigma_\mu^2, \sigma_v^2\right) = (0.8, 0.2)$

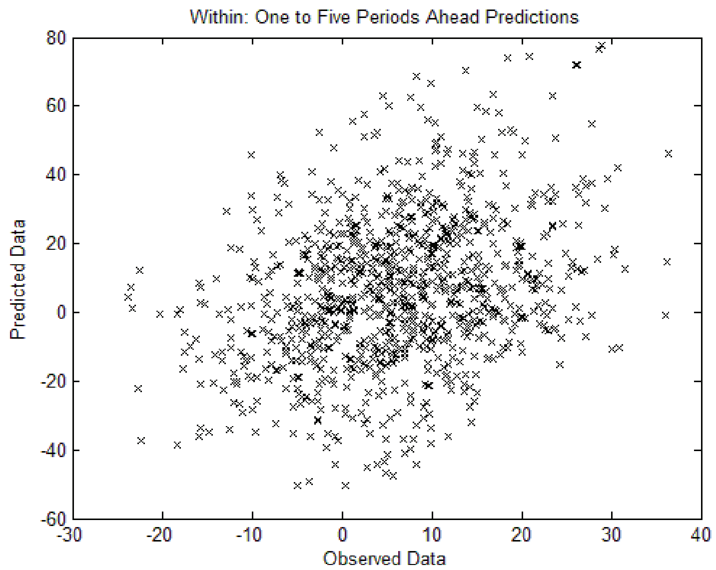

Figure 4a: $\left(\sigma_\mu^2, \sigma_v^2\right) = (0.2, 0.8)$

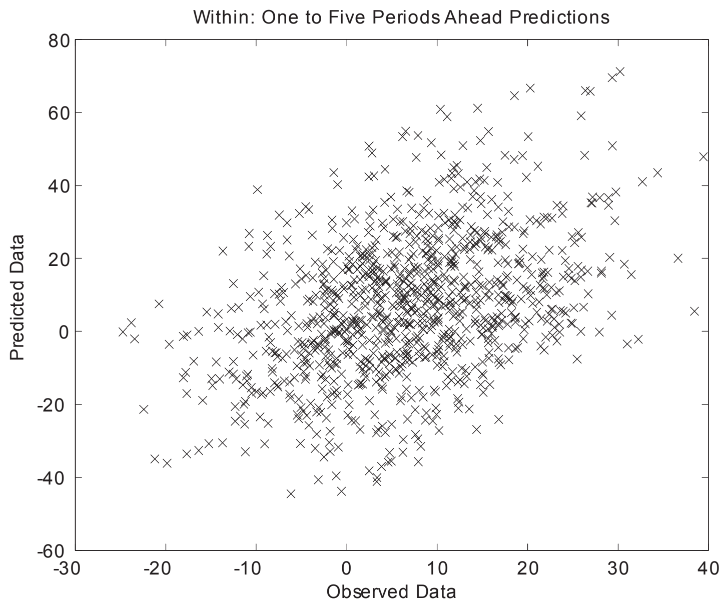

Figure 4b: $\left(\sigma_\mu^2, \sigma_v^2\right) = (0.8, 0.2)$

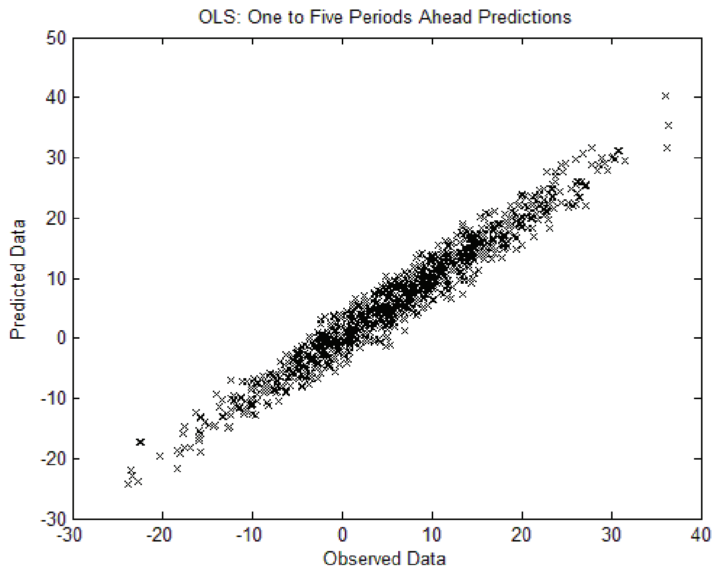

Figure 5a: $\left(\sigma_{\mu}^{2}, \sigma_{v}^{2}\right) = (0.2, 0.8)$

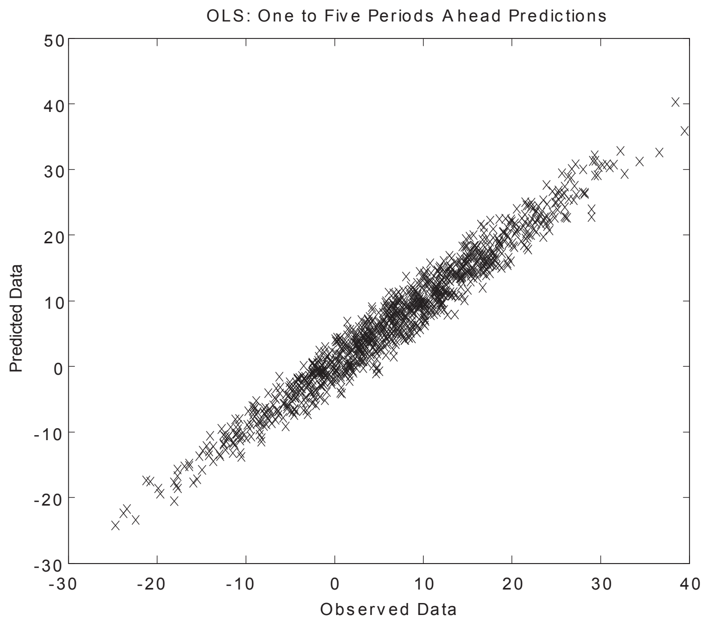

Figure 5b: $\left(\sigma_{\mu}^{2}, \sigma_{v}^{2}\right) = (0.8, 0.2)$

7 Empirical Illustration

In this section, we apply the estimation and prediction methods outlined above to estimate a time-space dynamic panel data model in which the level of employment across EU regions observed over recent years is the dependent variable and levels of output and capital are hypothesized causes of employment variation, controlling for spatial and temporal interactions involving employment. We analyze total employment over the period 2001-2010 across $N = 255$ NUTS2 regions of the EU[5]. The model specification we adopt is derived from standard urban economics as given by Abdel-Rahman and Fujita (1990), Ciccone and Hall (1996) and Fujita and Thisse (2002), among many others. From this it is possible to derive a model which shows that the drivers of (log) employment ($\ln \mathbf{e}$) are (log) output denoted by $\ln \mathbf{q}$ and a measure of (log) capital investment denoted by $\ln \mathbf{k}$. Following the arguments made at the start of Section 2, we assume an underlying trend toward equilibrium in employment levels in the absence of any disturbances. In reality equilibrium will be disturbed by other factors, but we maintain an assumption of a tendency towards equilibrium following a shock, so that $\ln \mathbf{e}_t = f(\ln \mathbf{e}_{t-1})$ and this leads logically to our specification with temporal and spatial wage interdependence, thus

$$\ln \mathbf{e}_t = c\boldsymbol{\iota}_t + \gamma \ln \mathbf{e}_{t-1} + \rho_1 \mathbf{W}_N \ln \mathbf{e}_t + \theta \mathbf{W}_N \ln \mathbf{e}_{t-1} + \beta_1 \ln \mathbf{q}_t + \beta_2 \ln \mathbf{k}_t + \boldsymbol{\varepsilon}_t \quad (60)$$

in which $\boldsymbol{\iota}_t$ is a vector of ones of dimension $(N \times 1)$, $\mathbf{e}_t$, $\mathbf{e}_{t-1}$, $\mathbf{q}_t$ and $\mathbf{k}_t$ are $(N \times 1)$ vectors of observations in levels, with $t = 2001, \ldots, 2010$. The data series are based on Cambridge Econometrics' European Regional Economic Data Base, in which $\mathbf{e}_t$ is the annual regional employment series, $\mathbf{q}_t$ is output (Gross Value Added, or GVA) and $\mathbf{k}_t$ is a measure of capital investment (Gross Fixed Capital Formation, or GFCF). The error term $\boldsymbol{\varepsilon}_t$ captures all other unobservable effects influencing the level of employment, especially interregional heterogeneity.

Written in first difference terms, in other words as exponential growth rates, the estimating equation is

$$\begin{aligned} \Delta \ln \mathbf{e}_t &= \gamma \Delta \ln \mathbf{e}_{t-1} + \rho_1 \mathbf{W}_N \Delta \ln \mathbf{e}_t + \theta \mathbf{W}_N \Delta \ln \mathbf{e}_{t-1} + \beta_1 \Delta \ln \mathbf{q}_t \\ &\quad + \beta_2 \Delta \ln \mathbf{k}_t + \Delta \boldsymbol{\varepsilon}_t. \end{aligned} \quad (61)$$

[5] We use 'regions of the EU' as a convenience, since some regions are located in closely intergrated but non-EU countries Norway and Switzerland.

The matrix $\mathbf{W}_N$ is based on estimated bilateral trade flows between EU NUTS2 regions. The data come from the PBL (the Netherlands Environmental Assessment Agency)[6] who developed a new methodology which is close to that of Simini et al. (2012). Details of the methodology are given in Thiessen et al. (2013*a*, *b*, *c*), see also Gianelle et al. (2014). The method follows a top-down approach and therefore is consistent with the national accounts of the different countries. Given the total international exports and imports on the country level, interregional trade flows are derived using data on business travel (services) and on freight transport (goods). Trade flows involving regions of Switzerland and Norway were obtained on the basis of interregional trade flows estimated by the best linear disaggregation method of Chow and Lin (1971), which was initially used to break down annual time series into quarterly series (see Abeysinghe and Lee, 1998, Doran and Fingleton, 2014). In this, commencing with aggregate trade values[7] between 21 EU counties, these were allocated to the NUTS2 regions. A parallel approach has been used by Polasek, Verduras and Sellner (2010), Vidoli and Mazziotta (2010), and Fingleton, Garretsen and Martin (2015), who provide more detail. Finally, OLS regression of the log PBL trade flows on log Chow-Lin trade flows produced parameters used to predict the missing PBL regional trade flows for Switzerland and Norway using the values for these regions obtained *via* the Chow-Lin approach. We subsequently normalize the trade matrix by dividing by its maximum eigenvalue, thus giving the matrix $\mathbf{W}_N$. This normalization ensures that the most positive real eigenvalue of $\mathbf{W}_N$ is equal to $\max(eig) = 1.0$, and the continuous range for ρ_1 for which $\mathbf{B}_N = (\mathbf{I}_N - \rho_1 \mathbf{W}_N)$ is nonsingular is $\frac{1}{\min(eig)} < \rho_1 < 1$. This normalization retains real trade magnitude differences between regions so that $\mathbf{W}_N \ln \mathbf{e}_t$, for example, depend on the size of the trade flow between regions. In contrast, normalization by row standardization would make interregional interaction depend on trade shares.

Assuming an SMA process for the compound errors, the spatial error dependence is $\varepsilon_{it} = u_{it} - \lambda \mathbf{m}_i u_t$ in which $\mathbf{m}_i$ is the i'th row of $\mathbf{M}_N$, where $\mathbf{M}_N$ is an $(N \times N)$ row-normalized regional contiguity matrix. The key feature of an SMA process is that shocks to the unobservables have local rather than global effects. Note that both components of the compound errors $\boldsymbol{\varepsilon}_t$ are

[6] We are grateful to Mark Thiessen, who kindly provided the data. The data can be visualized at http://themasites.pbl.nl/eu-trade/index2.html?vis=net-scores.

[7] They are downloadable from http://cid.econ.ucdavis.edu/data/undata/undata.html, see also Feenstra et al. (2005).

assumed to be subject to this same spatial error dependence processes. The SMA assumption is somewhat distinct from the more usual SAR assumption for the errors, which implies complex simultaneous association involving errors across all regions.

With regard to causation, we make two assumptions. One is that the regressors are strictly exogenous. In this case the moments conditions[8] are given by equations (25) and (29), combined with the moments for the endogenous variable and its spatial lag, as defined in the moments equations (24) and (27). Probably a more realistic assumption is that there will be feedback from employment to the variables $\mathbf{q}_t$ and $\mathbf{k}_t$ but that this will be delayed rather than instantaneous, and thus we estimate the model assuming that the variables are predetermined. In this case estimation is based on the moments (26), (31), (24) and (27).

Columns 1 and 2 of Table 12 gives the resulting parameter estimates[9] for equation (61) assuming either predeterminedness or exogeneity, and this shows that the effects of $\ln \mathbf{q}_t$ and $\ln \mathbf{k}_t$ on $\ln \mathbf{e}_t$ are significant and positive. Note that the negative estimated λ also indicates positive error dependence. If the positive space and time dependence parameter estimates were large, so that $\gamma + \rho_1 > 1$, this would not necessarily imply nonstationarity if $\theta < 0$. On the other hand given a restriction that $\theta = 0$ then large γ would necessarily requires small ρ_1 in order to satisfy stationarity. Therefore under the assumption that $\theta = 0$ the possibility of bias is introduced. However with $\theta \neq 0$, in other words with no restriction imposed on space-time covariance, a large positive γ plus large positive ρ_1 could be offset by a negative covariance term so that collectively the parameters pass the stationarity conditions. In this instance it turns out that with predetermined regressors the maximum absolute eigenvalue of $\mathbf{C}_N\mathbf{B}_N^{-1}$ is $0.6757 < 1$, $\rho_1 + \theta = 0.13472$, $|\gamma| + (\rho_1 + \theta) = 0.7872 < 1$, and $\rho_1 - \theta = 0.9359$, $\gamma - (\rho_1 - \theta) = -0.28337 > -1$, see (6), (7), (8) and (9). Likewise, assuming exogenous regressors, the model parameter estimates indicate stationarity and dynamic stability.

The estimates obtained cast some light on the question of increasing returns, in other words as a region's economy grows, are there productivity

[8]For simplicity, we exclude the additional moments based on $\mathbf{W}_N^2$.

[9]Note that because the estimates are based on differences, no estimate of the constant c is provided. This estimate is subsequently constructed as the difference between $\ln \mathbf{e}$ and the expected $\ln \mathbf{e}$ given by the model without c, using means over time. With the assumption of predetermined regressors this gives $c = 0.4462$, assuming exogeneity gives $c = 0.4076$.

benefits so that employment grows by less than output, due to positive externalities associated with increasing size and diversity of the economy as **q** increases overcoming negative ones such as the effects of congestion. Assuming predetermined regressors, controlling for the effect of **k**, spatial interaction, temporal dependence and space-time covariance, the estimated β_1 suggests that a 1% increase in **q** produces a 0.1272% increase in employment. This would indicate a high level of productivity growth. However, this estimate is misleading. As noted earlier, under model specifications involving autoregressive spatial interdependence of a dependent variable **y**, the partial derivative $\frac{\partial \mathbf{y}}{\partial \mathbf{x}_j}$ is not simply equal to the regression coefficient β_j, as pointed out by LeSage and Pace (2009). The true effect of $\mathbf{x}_j$ differs from β_j because it also includes the consequences of spillovers across regions. Based on the Table 12 estimates, one obtains *via* equations (17) and (19) (the means of) the true derivatives which are given in Table 13, with total effects partitioned into direct and indirect components. As shown in Table 13 the total long-term total effect of **q** equal to 0.4184, assuming predetermined regressors, which remains well below the value of 1.0 which one would associate with constant returns to scale. In order to obtain the standard errors and hence t-ratios given in Table 13, we draw at random from the multivariate normal distribution with mean equal to the parameter estimates given in Table 12, and covariance matrix equal to the estimated covariance matrix for these parameters. Each draw allows us to calculate the corresponding short and long term direct, indirect and total effects. With multiple (500) draws we obtain the distributions of these effects thus giving the standard errors in Table 13. From these draws it is evident that the effects are significantly different to zero under exogeneity and predetermindness. Note that the total long term elasticity with regard to GVA is similar, and clearly less than 1.0, regardless of estimator. It is evident that increasing productivity is an enduring characteristic of the EU regional economy.

The predictive performance of the SMA specification is given by the one- and two-step ahead predictions[10], as measured by the post sample RMSE criterion. Computing the out-of-sample forecast errors for 2011 and for 2012 indicates that the specification with either predetermined regressors or assuming exogeneity, provides relatively accurate predictions compared with the outcome of assuming SAR errors, which do not explicitly focus on lo-

[10] Data limitations mean that for 2012, k in each region is estimated using each region's previous growth rate.

cal spillovers. With SAR errors $\varepsilon_{it} = \rho_2 \sum_{j=1}^{N} m_{ij}\varepsilon_{jt} + u_{it}$ in which m_{ij} denotes cell(i,j) of $\mathbf{M}_N$, and $u_{it} = \mu_i + v_{it}$, in which $\mu_i \sim iid(0,\sigma_\mu^2)$ is a region (i.e. individual)-specific time-invariant effect and the remainder effect $v_{it} \sim iid(0,\sigma_v^2)$. In this case the RMSEs for the SAR errors estimator are calculated from the difference between $\mathbf{y}_t$ and $\widehat{\mathbf{y}}_t$ given by the prediction equation

$$\widehat{\mathbf{y}}_t = \widehat{\mathbf{B}}_N^{-1}\left[\widehat{\mathbf{C}}_N\widehat{\mathbf{y}}_{t-1} + c\boldsymbol{\iota}_t + \mathbf{x}_t\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{H}}_N^{-1}\overline{\boldsymbol{\mu}}\right] \tag{62}$$

in which $\widehat{\mathbf{H}}_N = (\mathbf{I}_N - \widehat{\rho}_2\mathbf{M}_N)$ is a nonsingular matrix and $\overline{\boldsymbol{\mu}}$ is based on

$$\boldsymbol{\mu} = \mathbf{H}_N\left(\mathbf{B}_N\mathbf{y}_t - \mathbf{C}_N\mathbf{y}_{t-1} - \mathbf{x}_t\boldsymbol{\beta} - c\boldsymbol{\iota}_t\right) - \mathbf{v}_t. \tag{63}$$

With exogenous regressors, the the SAR errors estimator is non-stationary and dynamically unstable with out-of-sample RMSE forecast errors for 2011 and for 2012 of 0.8305 and 2.3758. Assuming predetermined regressors, the estimator is stationary and gives RMSE's of 0.1742 and 0.2291.

The final two columns of Table 12 gives estimates for the SMA errors specification with predetermined and exogenous regressors, but with the additional variables $\mathbf{W}_N\mathbf{x}_{1t}$, the spatial lag of $\mathbf{q}_t$, with parameter β_3, and $\mathbf{W}_N\mathbf{x}_{2t}$, the spatial lag of $\mathbf{k}$, with parameter β_4. This is thus a form of spatial Durbin specification with regressors $\mathbf{x}_t = (\mathbf{x}_{1t}, \mathbf{x}_{2t}, \mathbf{W}_N\mathbf{x}_{1t}, \mathbf{W}_N\mathbf{x}_{2t})$ .The additional covariates evidently cause a problem of weak instruments, giving dynamically unstable nonstationary estimates, as reflected by the largest characteristic roots of $\mathbf{B}_N^{-1}\mathbf{C}_N$ equal to 1.0663 and 1.9041 respectively, and the one-step ahead RMSEs are 7.4094 and 3.0746. Assuming the Spatial Durbin with predetermined regressors and with ρ_2 restricted to zero, gives a largest characteristic root equal to 1.1127 and RMSE equal to 3.3007. The same specification but with a spatial autoregressive (SAR) error process gives 2.489 and 23.3138 respectively. Thus omitting the spatial lags $\mathbf{W}_N\mathbf{x}_{1t}$ and $\mathbf{W}_N\mathbf{x}_{2t}$ and hence restricting the direct effect to the total effect ratio to equality across explanatory variables is a necessary simplification. Comparing the RMSEs of the SAR and SMA estimators, the conclusion is that the GM-TS-SMA-RE estimator with predetermined regressors provides more accurate one- and two-step ahead predictions of employment levels across EU regions.

Table 12 - GM-TS-SMA-RE parameter estimates

Parameters	Predetermined	Exogenous	Spatial Durbin (a)	Spatial Durbin (b)
γ	0.6525 (0.003769) (173.1)	0.5634 (0.001425) (395.2)	0.3357 (0.00209) (160.6)	0.5673 (0.002129) (266.4)
ρ_1	0.5353 (0.01045) (51.2)	0.5909 (0.006908) (85.54)	1.5160 (0.01156) (131.2)	1.1990 (0.01358) (88.3)
β_1	0.1272 (0.002116) (60.11)	0.1787 (0.0006129) (291.5)	0.3604 (0.001307) (275.6)	0.1897 (0.001245) (152.4)
β_2	0.02636 (0.0006568) (40.13)	0.01966 (0.0001589) (123.7)	−0.03702 (0.0007463) (−49.61)	0.01581 (0.0002156) (73.34)
β_3	–	–	0.2910 (0.005024) (57.91)	0.3576 (0.006044) (59.18)
β_4	–	–	−0.3930 (0.003494) (−112.5)	−0.3025 (0.002825) (−107.1)
θ	−0.4006 (0.008868) (−45.17)	−0.3807 (0.005913) (−64.39)	−0.8863 (0.008452) (−104.9)	−0.9465 (0.009802) (−96.56)
λ	−0.7975	−0.6545	−0.3079	−0.5852
σ^2_μ	0.0753	0.2786	0.0097	0.4663
σ^2_v	0.0003	0.0003	0.0006	0.0003
Forecasting RMSE				
2011	0.0465	0.1413	7.4094	3.0746
2012	0.0977	0.2168	7.4711	3.1091
Average	0.0721	0.1791	7.4403	3.0918

(a) Predetermined regressors; (b) Exogenous regressors.

Table 13 - GM-TS-SMA-RE short- and long-term effects

	Predetermined			Exogenous		
	Direct	Indirect	Total	Direct	Indirect	Total
GVA($\mathbf{q}$)						
Short-term	0.1278	0.0285	0.1564	0.1798	0.0473	0.2271
	(0.00218)	(0.0007)	(0.0022)	(0.0006)	(0.0009)	(0.0011)
	(60.94)	(38.43)	(70.27)	(288.43)	(51.31)	(213.42)
Long-term	0.3669	0.0515	0.4184	0.4108	0.0779	0.4887
	(0.0068)	(0.0020)	(0.0069)	(0.0008)	(0.0017)	(0.0019)
	(53.624)	(26.08)	(60.90)	(494.39)	(47.03)	(257.59)
GFCF($\mathbf{k}$)						
Short-term	0.0265	0.0059	0.0324	0.0198	0.0052	0.0250
	(0.0007)	(0.0002)	(0.0009)	(0.0002)	(0.0001)	(0.0002)
	(39.67)	(24.20)	(38.07)	(121.60)	(46.80)	(110.39)
Long-term	0.0760	0.0107	0.0867	0.0452	0.0086	0.0538
	(0.0017)	(0.0005)	(0.0020)	(0.0004)	(0.0002)	(0.0005)
	(44.50)	(20.76)	(44.27)	(117.40)	(44.78)	(111.81)

8 Conclusion

In this paper we have developed a new spatial panel data estimator, denoted by GM-TS-SMA-RE, which incorporates dynamic effects and spatial effects including a time-space covariance term, with the spatial effects comprising a spatially autoregressive endogenous spatial lag. This is combined with a spatial moving average compound error process, which we advocate as a means of controlling for local spillovers. The resulting time-space dynamic panel data estimator with spatial moving average errors is shown *via* Monte Carlo simulations to produce estimates which are similar to the true values used in the data generating process. We also show that the estimator is superior to a number of alternatives in terms of its forecasting accuracy, as measured by the mean RMSE. An empirical example examines the efficacy of the estimator in the context of modelling and predicting employment levels across EU NUTS2 regions, which provides evidence for the existence of increasing returns to scale and cumulative agglomeration processes in the EU economy. Estimation based on an assumption of predetermined regressors and SMA errors produces the slightly more accurate one- and two-step ahead forecasts when compared with an assumption of exogeneity, and when compared with forecasts based on an equivalent estimator (GM-TS-SAR-RE) assuming SAR errors.

References

Abdel-Rahman, H. and Fujita, M. (1990). 'Product variety, Marshallian externalities and city size', *Journal of Regional Science*, Vol. 30, pp. 165–83.

Abeysinghe, T. and Lee, C. (1998). 'Best linear unbiased disaggregation of annual GDP to quarterly figures: the case of Malaysia', *Journal of Forecasting*, Vol. 17, pp. 527-537.

Anderson, T.W. and Hsiao, C. (1981). 'Estimation of dynamic models with error components', *Journal of the American Statistical Association*, Vol. 76, pp. 598-606.

Anderson, T.W. and Hsiao, C. (1982). 'Formulation and estimation of dynamic models using panel data', *Journal of Econometrics*, Vol. 18, pp. 47-82.

Anselin, L. (2001). Spatial econometrics. Ch. 14 in B.H. Baltagi, ed., *A Companion to Theoretical Econometrics*, Blackwell Publishers Lte, Massachusetts, pp. 310-330.

Anselin, L., Le Gallo, J. and Jayet, H. (2008). Spatial panel econometrics, Ch. 19 in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Springer-Verlag, Berlin, pp. 625-660.

Arellano, M. and Bond, S. (1991). 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment', *Review of Economic Studies*, Vol. 58, pp. 277-297.

Arellano, M. and Bover, O. (1995). 'Another look at the instrumental variable estimation of error-components models'. *Journal of Econometrics*, Vol. 68, pp. 29-52.

Baltagi, B.H. and Li, D. (2004). Prediction in the panel data model with spatial correlation, Chapter 13 in L. Anselin, R.J.G.M. Florax and S.J. Rey, eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer, Berlin, pp. 283-295.

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

Baltagi, B.H. and Yang, Z. (2013). 'Heteroskedasticity and non-normality robust LM tests for spatial dependence', *Regional Science and Urban Economics*, Vol. 43, pp. 725-739.

Baltagi, B.H., Fingleton, B. and Pirotte, A. (2014). 'Estimating and forecasting with a dynamic spatial panel data model', *Oxford Bulletin of Economics and Statistics*, Vol. 76, pp. 112-136.

Blundell, R. and Bond, S. (1998). 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics*, Vol. 87, pp. 115-143.

Blundell, R., Bond, S. and Windmeijer, F. (2000). 'Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator'. *Advanced in Econometrics*, 15, pp. 53-91.

Bond, S. (2002). 'Dynamic panel data models: a guide to micro data methods and practice', *Portuguese Economic Journal*, Vol. 1, pp. 141-162.

Bouayad-Agha, S. and Védrine, L. (2010). 'Estimation strategies for a spatial dynamic panel using GMM. A new approach to the convergence issue of European regions', *Spatial Economic Analysis*, Vol. 5, pp. 205-227.

Brady, R.R. (2011). 'Measuring the diffusion of housing prices across space and time', *Journal of Applied Econometrics*, Vol. 26, pp. 213-231.

Ciccone, A. and Hall, R. E. (1996). 'Productivity and the density of economic activity', *American Economic Review*, Vol. 86, pp. 54–70.

Chow, G. and Lin, A.-I. (1971). 'Best linear unbiased interpolation, distribution and extrapolation of time series by related series', *Review of Economics and Statistics*, Vol. 53, pp. 372-375.

Debarsy, N., Ertur, C. and LeSage, J.P. (2012). 'Interpreting dynamic space-time panel data models', *Statistical Methodology*, Vol. 9, pp. 158-171.

Doran, J. and Fingleton, B. (2014). 'Economic shocks and growth: spatio-temporal perspectives on Europe's economies in a time of crisis', *Papers in Regional Science*, Vol. 93, pp. S137-S165.

Elhorst, J.P. (2001). 'Dynamic models in space and time', *Geographical Analysis*, Vol. 33, pp. 119-140.

Elhorst, J.P. (2010). 'Dynamic panels with endogenous interaction effects when T is small', *Regional Science and Urban Economics*, Vol. 40, pp. 272-282.

Elhorst, J.P. (2012). 'Dynamic spatial panels: models, methods and inferences', *Journal of Geographical Systems*, Vol. 14, pp. 5-28.

Elhorst, J.P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Heidelberg, New York, Dordrecht, London: Springer-Verlag.

Feenstra, R.C., Lipsey, R.E., Deng, H., Ma, A.C. and Mo, H. (2005). 'World trade flows: 1962-2000', NBER Working Paper No. 11040. Available at: http://www.nber.org/papers/w11040.

Fingleton, B. (2008*a*). 'A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors', *Spatial Economic Analysis*, Vol. 3, pp. 28-44.

Fingleton, B. (2008*b*). 'A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices', *Empirical Economics*, Vol. 34, pp. 35–57.

Fingleton, B. (2009). 'Prediction using panel data regression with spatial random effects', *International Regional Science Review*, Vol. 32, pp. 195-220.

Fingleton, B. and McCombie, J. (1998). 'Increasing returns and economic growth : some evidence for manufacturing from the European Union regions', *Oxford Economic Papers*, Vol. 50, pp. 89-105.

Fingleton, B., Garretsen, H. and Martin, R.L. (2014). 'Shocking aspects of monetary union: the vulnerability of regions in Euroland', *Journal of Economic Geography*, first published online January 8, 2015 doi:10.1093/jeg/lbu055.

Franzese Jr., R.J. and Hays, J.C. (2007). 'Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data', *Political Analysis*, Vol. 15, pp. 140-164.

Fujita, M. and Thisse, J.-F. (2002). *Economics of Agglomeration*, Cambridge University Press, Cambridge.

Gianelle C., Goenaga, X., González, I. and Thissen, M. (2014). 'Smart specialisation in the tangled web of European inter-regional trade', S3 Working Paper Series No. 05/2014, JRC-IPTS, Sevilla.

Girardin, E. and Kholodilin, K. A. (2011). 'How helpful are spatial effects in forecasting the growth of Chinese provinces?', *Journal of Forecasting*, Vol. 30, pp. 622-643.

Goldberger, A.S. (1962). 'Best linear unbiased prediction in the generalized linear regression model', *Journal of the American Statistical Association*, Vol. 57, pp. 369-375.

Hsiao, C. (2003). *Analysis of Panel Data*, Cambridge University Press, New York.

Jacobs, J.P.A.M., Ligthart, J.E. and Vrijburg, H. (2009). 'Dynamic panel data models featuring endogenous interaction and spatially correlated errors', Working Paper n°09-15, International Studies Program, Georgia State University.

Kapoor, M., Kelejian, H.H. and Prucha, I.R. (2007). 'Panel data models with spatially correlated error components', *Journal of Econometrics*, Vol. 140, pp. 97-130.

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

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

Kelejian, H.H. and Prucha, I.R. (2007). 'Relative efficiencies of various predictors in spatial econometric models containing spatial lags', *Regional Science and Urban Economics*, Vol. 37, pp. 363-374.

Kelejian, H.H. (2016). 'Critical issues in spatial models: error term specifications, additional endogenous variables, pre-testing, and Bayesian analysis', *Letters in Spatial and Resource Sciences*, Vol. 9, pp. 113-136.

Kholodilin, K. A., Siliverstovs, B. and Kooths, S. (2008). 'A dynamic panel data approach to the forecasting of the GDP of German Länder', *Spatial Economic Analysis*, Vol. 3, pp. 195-207.

Korniotis, G.M. (2010). 'Estimating panel models with internal and external habit formation', *Journal of Business & Economic Statistics*, Vol. 28, pp. 145-158.

Kukenova, M. and Monteiro, J.-A. (2009). 'Spatial dynamic panel model and system GMM: a Monte Carlo investigation', Munich Personal RePEc Archive paper n°14319. http://mpra.ub.uni-muenchen.de/14319/.

Lee, L.-F. and Yu, J. (2010). 'Some recent developments in spatial panel data models', *Regional Science and Urban Economics*, Vol. 40, pp. 255-271.

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

Lee, L.-F. and Yu, J. (2015). Spatial panel data models, Ch. 12 in B.H. Baltagi, ed., *Oxford Handbook of Panel Data*, Oxford: Oxford University Press.

Lee, L.-F. and Yu, J. (2016). 'Identification of spatial Durbin panel models', *Journal of Applied Econometrics*, Vol. 31, pp. 133-162.

LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*, Chapman & Hall/CRC Press, Boca Raton.

Longhi, S. and Nijkamp, P. (2007). 'Forecasting regional labor market developments under spatial heterogeneity and spatial correlation', *International Regional Science Review*, Vol. 30, pp. 100-119.

Mutl, J. (2006). *Dynamic panel data models with spatially correlated disturbances*, PhD, University of Maryland.

Ord, J.K. (1975). 'Estimation methods for models of spatial interaction', *Journal of the American Statistical Association*, Vol. 70, pp. 120-126.

Pace, R.K., LeSage, J.P. and Zhu, S. (2012). Spatial dependence in regressors, Advances in Econometrics, Vol. 30, T.B. Fomby, R. Carter Hill, I. Jeliazkov, J.C. Escanciano and E. Hillebrand (Series Eds., Volume editors: D. Terrell and D. Millimet), Emerald Group Publishing Limited, pp. 257-295.

Parent, O. and LeSage, J.P. (2010). 'A spatial dynamic panel model with random effects applied to commuting times', *Transportation Research*, Part B, Vol. 44, pp. 633-645.

Parent, O. and LeSage, J.P. (2011). 'A space-time filter for panel data models containing random effects', *Computational Statistics and Data Analysis*, Vol. 55, pp. 475-490.

Parent, O. and LeSage, J.P. (2012). 'Spatial dynamic panel data models with random effects', *Regional Science & Urban Economics*, Vol. 42, pp. 727-738.

Pesaran, M.H. (2015). *Time Series and Panel Data Econometrics*, Oxford: Oxford University Press.

Polasek, W., Verduras, C. and Sellner, R. (2010), 'Bayesian methods for completing data in spatial models', *Review of Economic Analysis*, Vol. 2, pp. 194-214.

Schanne, N., Wapler, R. and Weyh, A. (2010). 'Regional unemployment forecasts with spatial interdependencies', *International Journal of Forecasting*, Vol. 26, pp. 908-926.

Simini F., González M.C., Maritan A. and Barabási A. (2012). A universal model for mobility and migration patterns, *Nature*, 484, 96–100.

Thissen, M., van Oort, F., Diodato, D. and Ruijs, A. (2013*a*). '*Regional Competitiveness and Smart Specialization in Europe: Place-based Development in International Economic Networks*', Cheltenham, UK: Edward Elgar Publishing.

Thissen, M., Diodato, D. and van Oort, F. (2013*b*). 'Integration and Convergence in Regional Europe: European Regional Trade Flows from 2000 to 2010', PBL publication number: 1036, PBL Netherlands Environmental Assessment Agency, The Hague/Bilthoven.

Thissen, M., Diodato, D. and van Oort F. (2013*c*). 'Integrated Regional Europe: European Regional Trade Flows in 2000', PBL publication number: 1035, PBL Netherlands Environmental Assessment Agency, The Hague/Bilthoven.

Vidoli, F. and Mazziotta, C. (2010). 'Spatial composite and disaggregate indicators: Chow-Lin methods and applications', Proceedings of the 45th Scientific Meeting of the Italian Statistical Society, Padua.

Yang, Z.L. (2017). 'Unified M-estimation of fixed-effects spatial dynamic models with short panels', forthcoming in *Journal of Econometrics*.

Yu, J., de Jong, R. and Lee, L.-F. (2008). 'Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large', *Journal of Econometrics*, Vol. 146, pp. 118-134.

Yu, J., de Jong, R. and Lee, L.-F. (2012). 'Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration', *Journal of Econometrics*, Vol. 167, pp. 16-37.

Table 1 – Mean, bias and RMSE of the coefficients $\hat{\gamma}$, $\hat{\rho}_1$, $\hat{\theta}$, $\hat{\beta}$ and $\hat{\lambda}$ for rook contiguity matrix, $(\alpha, \beta, \lambda) = (1, 1, -0.4)$, $(\sigma_\mu^2, \sigma_v^2) = (0.2, 0.8)$, $(N, T) = (100, 12)$, 5,000 replications

		Non-spatial estimators						*Spatial estimators*			
		OLS		*Within*		*GMM*		*GM-TS-RE*		*GM-TS-SMA-RE*	
(γ, ρ_1, θ)		$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
$\hat{\gamma}$	Mean	0.2249	0.8046	0.2019	0.7969	0.2458	0.8114	0.2045	0.7988	0.2018	0.7987
	Bias	0.0249	0.0046	0.0019	-0.0032	0.0457	0.0113	0.0044	-0.0012	0.0019	-0.0014
	RMSE	0.0261	0.0054	0.0075	0.0044	0.0513	0.0123	0.0146	0.0049	0.0147	0.0049
$\hat{\rho}_1$	Mean	0.8474	0.4502	0.8444	0.4556	*na*	*na*	0.8412	0.4516	0.8378	0.4385
	Bias	0.0476	0.0504	0.0446	0.0557	*na*	*na*	0.0413	0.0521	0.0382	0.0385
	RMSE	0.0482	0.0524	0.0452	0.0571	*na*	*na*	0.0430	0.0579	0.0403	0.0470
$\hat{\theta}$	Mean	-0.2546	-0.3488	-0.2300	-0.3521	*na*	*na*	-0.2268	-0.3507	-0.2231	-0.3380
	Bias	-0.0546	-0.0489	-0.0298	-0.0523	*na*	*na*	-0.0269	-0.0510	-0.0232	-0.0384
	RMSE	0.0558	0.0509	0.0318	0.0538	*na*	*na*	0.0333	0.0571	0.0309	0.0469
$\hat{\beta}$	Mean	0.9689	0.9855	0.9883	0.9969	1.2124	1.0376	0.9846	0.9967	0.9888	0.9999
	Bias	-0.0311	-0.0143	-0.0118	-0.0032	0.2128	0.0375	-0.0156	-0.0034	-0.0113	0.0003
	RMSE	0.0324	0.0163	0.0146	0.0078	0.2141	0.0390	0.0230	0.0130	0.0207	0.0128
$\hat{\lambda}$	Mean	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	-0.3210	-0.3358
	Bias	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0795	0.0651
	RMSE	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0914	0.0831
(γ, ρ_1, θ)		$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
$\hat{\gamma}$	Mean	0.7087	0.8056	0.6989	0.7985	0.6853	0.7653	0.7019	0.8010	0.7011	0.8005
	Bias	0.0086	0.0055	-0.0011	-0.0015	-0.0148	-0.0348	0.0020	0.0011	0.0012	0.0006
	RMSE	0.0094	0.0061	0.0038	0.0034	0.0178	0.0357	0.0071	0.0052	0.0070	0.0052
$\hat{\rho}_1$	Mean	0.8400	0.8381	0.8386	0.8379	*na*	*na*	0.8358	0.8353	0.8323	0.8313
	Bias	0.0403	0.0384	0.0387	0.0381	*na*	*na*	0.0358	0.0354	0.0325	0.0316
	RMSE	0.0410	0.0391	0.0394	0.0387	*na*	*na*	0.0376	0.0372	0.0348	0.0339
$\hat{\theta}$	Mean	-0.6421	-0.7376	-0.6333	-0.7329	*na*	*na*	-0.6331	-0.7322	-0.6297	-0.7289
	Bias	-0.0422	-0.0377	-0.0334	-0.0329	*na*	*na*	-0.0331	-0.0324	-0.0299	-0.0291
	RMSE	0.0431	0.0385	0.0344	0.0338	*na*	*na*	0.0359	0.0348	0.0333	0.0320
$\hat{\beta}$	Mean	0.9769	0.9799	0.9902	0.9903	1.2599	1.2583	0.9867	0.9874	0.9911	0.9923
	Bias	-0.0231	-0.0200	-0.0098	-0.0097	0.2599	0.2587	-0.0134	-0.0127	-0.0088	-0.0076
	RMSE	0.0245	0.0215	0.0123	0.0121	0.2605	0.2594	0.0197	0.0186	0.0175	0.0162
$\hat{\lambda}$	Mean	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	-0.3309	-0.3329
	Bias	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0694	0.0675
	RMSE	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0835	0.0820

na = not applicable.

Table 2 – Mean, bias and RMSE of the coefficients $\hat{\gamma}$, $\hat{\rho}_1$, $\hat{\theta}$, $\hat{\beta}$ and $\hat{\lambda}$ for rook contiguity matrix, $(\alpha, \beta, \lambda) = (1, 1, -0.4)$, $(\sigma_\mu^2, \sigma_v^2) = (0.8, 0.2)$, $(N, T) = (100, 12)$, 5,000 replications

		Non-spatial estimators						*Spatial estimators*			
		OLS		*Within*		*GMM*		*GM-TS-RE*		*GM-TS-SMA-RE*	
(γ, ρ_1, θ)		$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
$\hat{\gamma}$	Mean	0.2643	0.8160	0.2007	0.7992	0.2595	0.8123	0.2014	0.7997	0.2005	0.7997
	Bias	0.0637	0.0160	0.0007	-0.0008	0.0594	0.0122	0.0014	-0.0003	0.0006	-0.0003
	RMSE	0.0647	0.0165	0.0038	0.0018	0.0609	0.0126	0.0071	0.0024	0.0073	0.0024
$\hat{\rho}_1$	Mean	0.8334	0.4247	0.8129	0.4148	*na*	*na*	0.8118	0.4135	0.8106	0.4097
	Bias	0.0333	0.0247	0.0129	0.0148	*na*	*na*	0.0119	0.0137	0.0107	0.0098
	RMSE	0.0343	0.0303	0.0137	0.0163	*na*	*na*	0.0139	0.0191	0.0131	0.0168
$\hat{\theta}$	Mean	-0.2680	-0.3246	-0.2093	-0.3139	*na*	*na*	-0.2084	-0.3133	-0.2069	-0.3095
	Bias	-0.0677	-0.0245	-0.0093	-0.0139	*na*	*na*	-0.0087	-0.0137	-0.0071	-0.0095
	RMSE	0.0690	0.0299	0.0110	0.0154	*na*	*na*	0.0137	0.0194	0.0132	0.0169
$\hat{\beta}$	Mean	0.9347	0.9649	0.9967	0.9992	1.2014	0.10371	0.9957	0.9991	0.9970	1.0000
	Bias	-0.0649	-0.0348	-0.0034	-0.0009	0.2014	0.0371	-0.0044	-0.0009	-0.0028	-0.0000
	RMSE	0.0660	0.0363	0.0054	0.0036	0.2018	0.0376	0.0094	0.0064	0.0091	0.0064
$\hat{\lambda}$	Mean	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	-0.3746	-0.3826
	Bias	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0263	0.0184
	RMSE	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.542	0.0530
(γ, ρ_1, θ)		$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
$\hat{\gamma}$	Mean	0.7243	00.8163	0.6998	0.7996	0.6898	0.7690	0.7006	0.8003	0.7003	0.80002
	Bias	0.0242	0.0163	-0.0002	-0.0004	-0.0103	-0.0310	0.0006	0.0003	0.0004	0.0002
	RMSE	0.0248	0.0168	0.0019	0.0016	0.0119	0.0315	0.0034	0.0026	0.0035	0.0026
$\hat{\rho}_1$	Mean	0.8237	0.8215	0.8107	0.8105	*na*	*na*	0.8099	0.8098	0.8088	0.8085
	Bias	0.0237	0.0215	0.0108	0.0106	*na*	*na*	0.0098	0.0098	0.0088	0.0085
	RMSE	0.0250	0.0232	0.0116	0.0114	*na*	*na*	0.0119	0.0119	0.0112	0.0110
$\hat{\theta}$	Mean	-0.6335	-0.7238	-0.6095	-0.7093	*na*	*na*	-0.6093	-0.7091	-0.6082	-0.7079
	Bias	-0.0336	-0.0239	-0.0094	-0.0092	*na*	*na*	-0.0094	-0.0090	-0.0081	-0.0079
	RMSE	0.0350	0.0255	0.0105	0.0102	*na*	*na*	0.0122	0.0116	0.0115	0.0109
$\hat{\beta}$	Mean	0.9533	0.9604	0.9973	0.9973	1.2581	1.2580	0.9963	0.9965	0.9977	0.9980
	Bias	-0.0464	-0.0393	-0.0027	-0.0027	0.2583	0.2582	-0.0037	-0.0035	-0.0023	-0.0020
	RMSE	0.0476	0.0408	0.0046	0.0045	0.2585	0.2584	0.0082	0.0077	0.0078	0.0074

$\hat{\lambda}$	Mean	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	-0.3732	-0.3806
	Bias	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0264	0.0205
	RMSE	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0620	0.0520

na = not applicable.

Table 3 – Mean, bias and RMSE of the average direct, indirect and total short and long-term effects for rook contiguity matrix, $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{1}, \mathbf{1}, -\mathbf{0.4})$, $(\boldsymbol{\sigma_\mu^2}, \boldsymbol{\sigma_v^2}) = (\mathbf{0.2}, \mathbf{0.8})$, $(\boldsymbol{N}, \boldsymbol{T}) = (\mathbf{100}, \mathbf{12})$ and $\mathbf{5,000}$ replications

		GM-TS-RE				*GM-TS-SMA-RE*			
		Short-run		Long-run		Short-run		Long-run	
(γ, ρ_1, θ)		$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
Direct	Mean	1.3964	1.0648	1.6039	5.4020	1.3950	1.0638	1.6017	5.4114
	Bias	0.0595	0.0131	0.0211	-0.0414	0.0580	0.0123	0.0189	-0.0335
	RMSE	0.0664	0.0203	0.0379	0.1799	0.652	0.0196	0.0371	0.1759
Indirect	Mean	4.8371	0.7565	3.8640	4.7070	4.7380	0.7210	3.8205	4.6779
	Bias	1.1500	0.1398	0.4162	-0.0230	1.0520	0.1031	0.3711	-0.0351
	RMSE	1.2285	0.1587	0.5727	1.1094	1.1383	0.1279	0.5393	1.1232
Total	Mean	6.2335	1.8213	5.4679	10.1089	6.1330	1.7848	5.4222	10.0892
	Bias	1.2086	0.1528	0.4345	-0.0803	1.1090	0.1150	0.3894	-0.0777
	RMSE	1.2916	0.1747	0.6035	1.2509	1.1987	0.1431	0.5708	1.2660
		Short-run		Long-run		Short-run		Long-run	
(γ, ρ_1, θ)		$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
Direct	Mean	1.3874	1.3873	3.9611	5.4400	1.3861	1.3858	3.9599	5.4432
	Bias	0.0497	0.0494	0.0121	0.0038	0.0492	0.0488	0.0116	0.0068
	RMSE	0.0566	0.0559	0.1008	0.1337	0.0562	0.0556	0.0989	0.1333
Indirect	Mean	4.6503	4.6355	6.4954	4.9459	4.5535	4.5267	6.4286	4.8696
	Bias	0.9607	0.9495	0.3590	0.3114	0.8706	0.8438	0.2865	0.2387
	RMSE	1.0328	1.0206	1.0285	0.9280	0.9554	0.9287	0.9946	0.8853
Total	Mean	6.0376	6.0228	10.4565	10.3859	5.9397	5.9126	10.3885	10.3128
	Bias	1.0084	1.0010	0.3629	0.3049	0.9175	0.8928	0.2981	0.2447
	RMSE	1.0846	1.0760	1.1147	0.9965	1.0061	0.9835	1.0779	0.9675

Table 4 – Mean, bias and RMSE of the average direct, indirect and total short and long-term effects for rook contiguity matrix, $(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{1}, \mathbf{1}, -\mathbf{0.4})$, $(\boldsymbol{\sigma}_{\boldsymbol{\mu}}^{\mathbf{2}}, \boldsymbol{\sigma}_{\boldsymbol{v}}^{\mathbf{2}}) = (\mathbf{0.8}, \mathbf{0.2})$, $(\boldsymbol{N}, \boldsymbol{T}) = (\mathbf{100}, \mathbf{12})$ and $\mathbf{5,000}$ replications

		GM-TS-RE				*GM-TS-SMA-RE*			
		Short-run		Long-run		Short-run		Long-run	
(γ, ρ_1, θ)		$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
Direct	Mean	1.3517	1.0548	1.5871	5.4242	1.3514	1.0546	1.5866	5.4277
	Bias	0.0148	0.0031	0.0050	-0.0091	0.0147	0.0030	0.0045	-0.0065
	RMSE	0.0207	0.0083	0.0163	0.0910	0.0204	0.0082	0.0163	0.0897
Indirect	Mean	3.9470	0.6497	3.5239	4.6011	3.9208	0.6404	3.5136	4.6003
	Bias	0.2780	0.0345	0.0987	-0.0143	0.2529	0.0246	0.0861	-0.0000
	RMSE	0.3336	0.0491	0.2084	0.5677	0.3147	0.0430	0.2062	0.5795
Total	Mean	5.2987	1.7045	5.1111	10.0253	5.2723	1.6950	5.1002	10.0281
	Bias	0.2931	0.0378	0.1029	-0.0245	0.2669	0.0275	0.0901	-0.0098
	RMSE	0.3522	0.0550	0.2220	0.6438	0.3322	0.0485	0.2204	0.6461
		Short-run		Long-run		Short-run		Long-run	
(γ, ρ_1, θ)		$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
Direct	Mean	1.3492	1.3492	3.9476	5.4335	1.3489	1.3489	3.9478	5.4351
	Bias	0.122	0.0122	0.0034	0.0018	0.0120	0.0119	0.0031	0.0018
	RMSE	0.0177	0.0174	0.0510	0.0681	0.0178	0.0176	0.0496	0.0679
Indirect	Mean	3.8985	3.8951	6.1592	4.6578	3.8748	3.8684	6.1466	4.6427
	Bias	0.2268	0.2265	0.0778	0.0707	0.2059	0.2019	0.0677	0.0579
	RMSE	0.2809	0.2797	0.4856	0.4299	0.2669	0.2631	0.4716	0.4334
Total	Mean	5.2477	5.2443	10.1068	10.0913	5.2237	5.2173	10.0944	10.0778
	Bias	0.2396	0.2388	0.0797	0.0681	0.2176	0.2139	0.0707	0.0598
	RMSE	0.2972	0.2960	0.5291	0.4739	0.2820	0.2791	0.5082	0.4785

Table 5 – Forecasting RMSE for rook contiguity matrix, $(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{1}, \mathbf{1}, -\mathbf{0.4})$, $(\boldsymbol{N}, \boldsymbol{T}) = (\mathbf{100}, \mathbf{12})$, $\mathbf{5,000}$ replications

	Non-spatial estimators						*Spatial estimators*			
	OLS		*Within*		*GMM*		*GM-TS-RE*		*GM-TS-SMA-RE*	
	$(\sigma_\mu^2, \sigma_v^2) = (0.2,0.8)$									
(γ, ρ_1, θ)	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
1st year	1.7685	2.5465	11.3291	2.4040	4.3129	2.3661	1.6872	1.2076	1.6721	1.1953
2sd year	1.9444	2.6361	12.6103	3.3726	4.4696	2.7497	1.8403	1.3967	1.8284	1.3824
3rd year	2.0046	2.6930	13.1446	4.1769	4.5812	3.0752	1.9078	1.5277	1.8956	1.5125
4th year	2.0605	2.7384	13.4392	4. 8622	5.0084	3.4147	1.9818	1.6385	1.9648	1.6213
5th year	2.1281	2.7778	13.6533	5.4516	5.1288	3.7209	2.0263	1.7342	2.0085	1.7158
5-year average	1.9812	2.6784	12.8353	4.0534	4.7002	3.0653	1.8887	1.5009	1.8739	1.4854
(γ, ρ_1, θ)	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
1st year	2.6747	3.0420	10.1648	9.1076	6.1694	5.6732	2.0227	2.0486	2.0010	2.0265
2sd year	2.8263	3.1930	14.1410	13.1759	6.5034	6.0698	2.2441	2.2846	2.2243	2.2648
3rd year	2.9015	3.2727	17.1859	16. 5860	6.7797	6.3661	2.3698	2.4245	2.3500	2.4050
4th year	2.9710	3.3435	19.5698	19.4890	7.3519	6.8989	2.4854	2.5515	2.4614	2.5277
5th year	3.0203	3.3969	21.4966	22.0172	7.6426	7.1959	2.5614	2.6415	2.5371	2.6174
5-year average	2.8788	3.2496	16.5117	16.0752	6.8894	6.4408	2.3367	2.3901	2.3148	2.3683
	$(\sigma_\mu^2, \sigma_v^2) = (0.8,0.2)$									
(γ, ρ_1, θ)	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
1st year	1.8157	4.1948	9.2422	1.9864	4.1539	2.1895	0.8105	0.5818	0.8092	0.5800
2sd year	1.8968	4.2613	10.2323	2.8592	4.2452	2.5449	0.8960	0.6825	0.8951	0.6804
3rd year	1.9269	4.3171	10.6665	3.5919	4.3312	2.8574	0.9305	0.7533	0.9295	0.7511
4th year	1.9581	4.3673	10.8986	4.2161	4.7867	3.1969	0.9561	0.8121	0.9544	0.8095
5th year	1.9766	4.4131	11.0466	4.7510	4.9068	3.5038	0.9736	0.8630	0.9719	0.8602
5-year average	1.9148	4.3107	10.4172	3.4809	4.4848	2.8585	0.9133	0.7385	0.9120	0.7362
(γ, ρ_1, θ)	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
1st year	3.8344	4.6037	8.4496	7.5269	6.0764	5.5896	0.9567	0.9688	0.9552	0.9681
2sd year	3.9016	4.6766	11.7747	10.9352	6.3501	5.9249	1.0809	1.1005	1.0792	1.0995
3rd year	3.9402	4.7248	14.3434	13.8143	6.6010	6.1934	1.1522	1.1798	1.1499	1.1785
4th year	3.9769	4.7700	16.3596	16.2718	7.1923	6.7487	1.2067	1.2420	1.2037	1.2400
5th year	4.0063	4.8078	17.9697	18.3914	7.4830	7.0483	1.2476	1.2903	1.2444	1.2882
5-year average	3.9319	4.7166	13.7794	13.3879	6.7406	6.3010	1.1288	1.1563	1.1265	1.1548

Table 6 – Mean, bias and RMSE of the coefficients $\hat{\gamma}$, $\hat{\rho}_1$, $\hat{\theta}$, $\hat{\beta}$ and $\hat{\lambda}$ for rook contiguity matrix, $(a, \beta, \lambda) = (1, 1, -0.4)$, $(\sigma_\mu^2, \sigma_v^2) = (0.2, 0.8)$, $(N, T) = (200, 12)$, 5,000 replications

		Non-spatial estimators						*Spatial estimators*			
		OLS		*Within*		*GMM*		*GM-TS-RE*		*GM-TS-SMA-RE*	
(γ, ρ_1, θ)		$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
$\hat{\gamma}$	Mean	0.2264	0.8054	0.2025	0.7981	0.2209	0.8116	0.2037	0.7998	0.2005	0.7994
	Bias	0.264	0.0055	0.0025	-0.0020	0.0210	0.0117	0.0038	-0.0002	0.0006	-0.0006
	RMSE	0.0270	0.0057	0.0058	0.0029	0.0262	0.0120	0.0075	0.0027	0.0064	0.0027
$\hat{\rho}_1$	Mean	0.8445	0.4468	0.8418	0.4520	*na*	*na*	0.8263	0.4259	0.8223	0.4162
	Bias	0.0445	0.0468	0.0419	0.0521	*na*	*na*	0.0265	0.0261	0.0225	0.0166
	RMSE	0.0449	0.0479	0.0422	0.0529	*na*	*na*	0.0274	0.0288	0.0238	0.0205
$\hat{\theta}$	Mean	-0.2538	-0.3454	-0.2293	-0.3489	*na*	*na*	-0.2202	-0.3250	-0.2155	-0.3159
	Bias	-0.0537	-0.0454	-0.0293	-0.0488	*na*	*na*	-0.0201	-0.0253	-0.0155	-0.0159
	RMSE	0.0543	0.0466	0.0302	0.0497	*na*	*na*	0.0225	0.0281	0.0186	0.0201
$\hat{\beta}$	Mean	0.9674	0.9825	0.9861	0.9932	1.3224	1.0670	0.9895	0.9961	0.9963	1.0002
	Bias	-0.0326	-0.0175	-0.0139	-0.0069	0.3221	0.0669	-0.0106	-0.0038	-0.0037	0.0002
	RMSE	0.0333	0.0183	0.0152	0.0085	0.3226	0.0673	0.0131	0.0072	0.0085	0.0061
$\hat{\lambda}$	Mean	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	-0.3431	-0.3635
	Bias	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0572	0.0370
	RMSE	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0657	0.0504
(γ, ρ_1, θ)		$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
$\hat{\gamma}$	Mean	0.7093	0.8061	0.6995	0.7989	0.6820	0.7553	0.7013	0.8006	0.7003	0.8001
	Bias	0.0094	0.0062	-0.0005	-0.0012	-0.0181	-0.0448	0.0012	0.0006	0.0003	0.0001
	RMSE	0.0097	0.0064	0.0028	0.0025	0.0192	0.0451	0.0037	0.0029	0.0035	0.0028
$\hat{\rho}_1$	Mean	0.8367	0.8348	0.8373	0.8365	*na*	*na*	0.8238	0.8230	0.8199	0.8191
	Bias	0.0368	0.0349	0.0374	0.0365	*na*	*na*	0.0239	0.0231	0.0201	0.0192
	RMSE	0.0373	0.0354	0.0377	0.0369	*na*	*na*	0.0249	0.0241	0.0214	0.0205
$\hat{\theta}$	Mean	-0.6396	-0.7351	-0.6323	-0.7314	*na*	*na*	-0.6222	-0.7211	-0.6184	-0.7177
	Bias	-0.0397	-0.0351	-0.0323	-0.0313	*na*	*na*	-0.0223	-0.0213	-0.0186	-0.0178
	RMSE	0.0402	0.0357	0.0328	0.0319	*na*	*na*	0.0238	0.0227	0.0204	0.0194
$\hat{\beta}$	Mean	0.9742	0.9782	0.9875	0.9877	1.3396	1.3287	0.9906	0.9911	0.9965	0.9968
	Bias	-0.0252	-0.0217	-0.0125	-0.0123	0.3395	0.3285	-0.0095	-0.0089	-0.0036	-0.0032

	RMSE	0.0259	0.0225	0.0136	0.0134	0.3397	0.3287	0.0117	0.0111	0.0076	0.0072
$\hat{\lambda}$	Mean	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	-0.3476	-0.3492
	Bias	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0529	0.0512
	RMSE	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0619	0.0608

na = not applicable.

Table 7 – Mean, bias and RMSE of the coefficients $\hat{\gamma}$, $\hat{\rho}_1$, $\hat{\theta}$, $\hat{\beta}$ and $\hat{\lambda}$ for rook contiguity matrix, $(a, \beta, \lambda) = (1, 1, -0.4)$, $(\sigma_\mu^2, \sigma_v^2) = (0.8, 0.2)$, $(N, T) = (200, 12)$, 5,000 replications

		Non-spatial estimators						*Spatial estimators*			
		OLS		*Within*		*GMM*		*GM-TS-RE*		*GM-TS-SMA-RE*	
(γ, ρ_1, θ)		$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
$\hat{\gamma}$	Mean	0.2645	0.8160	0.2008	0.7995	0.2248	0.8110	0.2010	0.7999	0.2001	0.7998
	Bias	0.0645	0.0160	0.0008	-0.0005	0.0245	0.0110	0.0011	-0.0001	0.0001	-0.0002
	RMSE	0.0650	0.0162	0.0028	0.0012	0.0265	0.0111	0.0035	0.0013	0.0032	0.0013
$\hat{\rho}_1$	Mean	0.8348	0.4324	0.8121	0.4139	*na*	*na*	0.8072	0.4066	0.8059	0.4039
	Bias	0.0347	0.0322	0.0121	0.0139	*na*	*na*	0.0073	0.0067	0.0060	0.0041
	RMSE	0.0353	0.0349	0.0125	0.0147	*na*	*na*	0.0083	0.0091	0.0072	0.0073
$\hat{\theta}$	Mean	-0.2706	-0.3301	-0.2089	-0.3131	*na*	*na*	-0.2057	-0.3064	-0.2042	-0.3039
	Bias	-0.0705	-0.0301	-0.0089	-0.0131	*na*	*na*	-0.0057	-0.0064	-0.0042	-0.0039
	RMSE	0.0711	0.0328	0.0098	0.0139	*na*	*na*	0.0079	0.0090	0.0068	0.0073
$\hat{\beta}$	Mean	0.9316	0.9605	0.9960	0.9982	1.3192	1.0671	0.9972	0.9990	0.9992	1.0001
	Bias	-0.0681	-0.0394	-0.0040	-0.0019	0.3193	0.0672	-0.0028	-0.0010	-0.0008	0.0001
	RMSE	0.0687	0.0402	0.0050	0.0031	0.3195	0.0673	0.0048	0.0032	0.0039	0.0030
$\hat{\lambda}$	Mean	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	-0.3819	-0.3885
	Bias	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0185	0.0116
	RMSE	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0384	0.0355
(γ, ρ_1, θ)		$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
$\hat{\gamma}$	Mean	0.7237	0.8162	0.6999	00.7997	0.6812	0.7543	0.7003	0.80002	0.70001	0.8000
	Bias	0.0238	0.0162	-0.0001	-0.0003	-0.0190	-0.0457	0.0003	0.0002	0.0000	0.0000
	RMSE	0.0240	0.0164	0.0014	0.0012	0.0194	0.0459	0.0018	0.0014	0.0017	0.0013
$\hat{\rho}_1$	Mean	0.8259	0.8236	0.8104	0.8101	*na*	*na*	0.8064	0.8061	0.8051	0.8049
	Bias	0.0258	0.0234	0.0104	0.0101	*na*	*na*	0.0064	0.0061	0.0052	0.0050
	RMSE	0.0268	0.0246	0.0108	0.0105	*na*	*na*	0.0075	0.0072	0.0065	0.0062
$\hat{\theta}$	Mean	-0.6355	-0.7264	-0.6091	-0.7088	*na*	*na*	-0.6059	-0.7056	-0.6047	-0.7046
	Bias	-0.0356	-0.0265	-0.0091	-0.0089	*na*	*na*	-0.0059	-0.0057	-0.0049	-0.0047

	RMSE	0.0364	0.0275	0.0097	0.0094	*na*	*na*	0.0073	0.0070	0.0065	0.0062
$\hat{\beta}$	Mean	0.9489	0.9561	0.9965	0.9966	1.3408	1.3295	0.9975	0.9976	0.9992	0.99993
	Bias	-0.0510	-0.0438	-0.0035	-0.0034	0.3409	0.3296	-0.0025	-0.0024	-0.0008	-0.0008
	RMSE	0.0516	0.0445	0.0044	0.0043	0.3409	0.3296	0.0043	0.0041	0.0034	0.0033
$\hat{\lambda}$	Mean	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	-0.3836	-0.3841
	Bias	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0169	0.0164
	RMSE	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*	0.0374	0.0371

na = not applicable.

Table 8 – Mean, bias and RMSE of the average direct, indirect and total short and long-term effects for rook contiguity matrix, $(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{1}, \mathbf{1}, -\mathbf{0.4})$, $(\boldsymbol{\sigma_\mu^2}, \boldsymbol{\sigma_v^2}) = (\mathbf{0.2}, \mathbf{0.8})$, $(\boldsymbol{N}, \boldsymbol{T}) = (\mathbf{200}, \mathbf{12})$ and $\mathbf{5,000}$ replications

		GM-TS-RE				*GM-TS-SMA-RE*			
		Short-run		Long-run		Short-run		Long-run	
(γ, ρ_1, θ)		$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
Direct	Mean	1.3607	1.0543	1.5824	5.4077	1.3625	1.0557	1.5835	5.4118
	Bias	0.0322	0.0036	0.0078	-0.0180	0.0341	0.0050	0.0091	-0.0139
	RMSE	0.0353	0.0079	0.0160	0.0778	0.0373	0.0087	0.0168	0.0777
Indirect	Mean	4.3449	0.6813	3.6277	4.6377	4.2538	0.6582	3.5940	4.5850
	Bias	0.6698	0.0651	0.2005	0.0484	0.5786	0.0425	0.1639	-0.0052
	RMSE	0.7034	0.0728	0.2534	0.4076	0.6206	0.0531	0.2279	0.3976
Total	Mean	5.7056	1.7356	5.2101	10.0454	5.6163	1.7140	5.1772	9.9968
	Bias	0.7014	0.0684	0.2080	0.0252	0.6117	0.0476	0.1742	-0.0247
	RMSE	0.7374	0.0776	0.2648	0.4551	0.6561	0.0598	0.2424	0.4458
		Short-run		Long-run		Short-run		Long-run	
(γ, ρ_1, θ)		$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
Direct	Mean	1.3574	1.3567	3.9337	5.4178	1.3581	1.3571	3.9376	5.4268
	Bias	0.0291	0.0283	0.0017	-0.0072	0.0301	0.0288	0.0056	0.0012
	RMSE	0.0322	0.0314	0.0416	0.0650	0.0333	0.0320	0.0432	0.0654
Indirect	Mean	4.2727	4.2516	6.2806	4.7631	4.1827	4.1606	6.2271	4.7056
	Bias	0.5960	0.5742	0.2031	0.1832	0.5067	0.4836	0.1479	0.1250
	RMSE	0.6302	0.6084	0.3955	0.3693	0.5483	0.5238	0.3713	0.3370

Total	Mean	5.6301	5.6082	10.2143	10.1809	5.5409	5.5177	10.1647	10.1324
	Bias	0.6248	0.6028	0.2023	0.1737	0.5368	0.5132	0.1542	0.1246
	RMSE	0.6606	0.6395	0.4180	0.3787	0.5808	0.5560	0.3972	0.3611

Table 9 – Mean, bias and RMSE of the average direct, indirect and total short and long-term effects for rook contiguity matrix, $(\boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{1}, \mathbf{1}, -\mathbf{0.4})$, $(\boldsymbol{\sigma}_{\mu}^{2}, \boldsymbol{\sigma}_{v}^{2}) = (\mathbf{0.8}, \mathbf{0.2})$, $(\boldsymbol{N}, \boldsymbol{T}) = (\mathbf{200}, \mathbf{12})$ and $\mathbf{5,000}$ replications

		GM-TS-RE				*GM-TS-SMA-RE*			
		Short-run		Long-run		Short-run		Long-run	
(γ, ρ_1, θ)		$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
Direct	Mean	1.3365	1.0516	1.5762	5.4195	1.3370	1.0520	1.5766	5.4207
	Bias	0.0082	0.0008	0.0018	-0.0049	0.0088	0.0013	0.0022	-0.0037
	RMSE	0.0110	0.0036	0.0072	0.0386	0.0115	0.0038	0.0073	0.0388
Indirect	Mean	3.8378	0.6321	3.4749	4.5910	3.8127	0.6261	3.4660	4.5779
	Bias	0.1651	0.0162	0.0493	0.0119	0.1408	0.0104	0.0395	-0.0021
	RMSE	0.1925	0.0224	0.0905	0.2061	0.1729	0.0184	0.0857	0.2004
Total	Mean	5.1743	1.6837	5.0511	10.0105	5.1497	1.6780	5.0426	9.9985
	Bias	0.1739	0.0170	0.0516	0.0068	0.1505	0.0116	0.0414	-0.0047
	RMSE	0.2037	0.0244	0.0957	0.2309	0.1845	0.0209	0.0914	0.2240
		Short-run		Long-run		Short-run		Long-run	
(γ, ρ_1, θ)		$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
Direct	Mean	1.3355	1.3353	3.9316	5.4219	1.3358	1.3355	3.9328	5.4247
	Bias	0.0072	0.0070	0.0003	-0.0020	0.0076	0.0073	0.0014	0.0001
	RMSE	0.0099	0.0097	0.0212	0.0329	0.0103	0.0100	0.0217	0.0331
Indirect	Mean	3.8178	3.8127	6.1234	4.6245	3.7942	3.7889	6.1093	4.6092
	Bias	0.1451	0.1400	0.0529	0.0486	0.1224	0.1171	0.0369	0.0307
	RMSE	0.1721	0.1670	0.1767	0.1661	0.1543	0.1486	0.1720	0.1598
Total	Mean	5.1533	5.1480	10.0550	10.0465	5.1299	5.1244	10.0421	10.0339
	Bias	0.1524	0.1472	0.0537	0.0466	0.1294	0.1247	0.0395	0.0303
	RMSE	0.1817	0.1763	0.1891	0.1739	0.1638	0.1581	0.1849	0.1698

Table 10 – Forecasting RMSE for rook contiguity matrix, $(\alpha, \beta, \lambda) = (1, 1, -0.4)$, $(N, T) = (200, 12)$, $5,000$ replications

	Non-spatial estimators						*Spatial estimators*			
	OLS		*Within*		*GMM*		*GM-TS-RE*		*GM-TS-SMA-RE*	
	$(\sigma_\mu^2, \sigma_v^2) = (0.2,0.8)$									
(γ, ρ_1, θ)	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
1st year	14.1073	2.5969	1.8673	2.9520	4.4517	2.6146	1.5887	1.1320	1.5758	1.1276
2sd year	15.7153	2.6845	2.0028	4.2041	4.6497	3.0051	1.7510	1.3102	1.7405	1.3052
3rd year	16.4150	2.7403	2.0417	5.2471	4.6637	3.3068	1.8100	1.4311	1.8011	1.4257
4th year	16.8072	2.7825	2.0881	6.1360	4.9916	3.5781	1.8573	1.5277	1.8461	1.5214
5th year	17.0370	2.8167	2.1035	6.8979	5.2407	3.8155	1.8819	1.6085	1.8704	1.6012
5-year average	16.0163	2.7242	2.0207	5.0874	4.7995	3.2640	1.7778	1.4019	1.7668	1.3962
(γ, ρ_1, θ)	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
1st year	2.7290	3.0913	12.7410	11.1849	5.9901	5.5729	1.8144	1.8472	1.8033	1.8361
2sd year	2.8873	3.2467	17.7832	16.2703	6.5245	6.2241	2.0468	2.0950	2.0342	2.0825
3rd year	2.9429	3.3095	21.6797	20.5747	6.8178	6.6154	2.1647	2.2303	2.1531	2.2188
4th year	3.0006	3.3711	24.7688	24.2849	7.3056	7.2078	2.2628	2.3435	2.2474	2.3283
5th year	3.0503	3.4239	27.2291	27.4773	7.8872	7.9035	2.3432	2.4372	2.3233	2.4176
5-year average	2.9220	3.2885	20.8404	19.9584	6.9050	6.7048	2.1263	2.1907	2.1123	2.1766
	$(\sigma_\mu^2, \sigma_v^2) = (0.8,0.2)$									
(γ, ρ_1, θ)	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$	$(0.2; 0.8; -0.2)$	$(0.8; 0.4; -0.3)$
1st year	1.8516	4.2376	11.7437	2.5583	4.2589	2.4465	0.7775	0.5550	0.7759	0.5548
2sd year	1.9324	4.3072	13.0199	3.6976	4.4024	2.7991	0.8636	0.6510	0.8623	0.6507
3rd year	1.9562	4.3662	13.5889	4.6542	4. 3928	3.0879	0.8966	0.7170	0.8954	0.7165
4th year	1.9805	4.4188	13.8956	5.4682	4.7390	3.3505	0.9170	0.7694	0.9154	0.7687
5th year	1.9946	4.4665	14.0802	6.1649	5.0017	3.5837	0.9294	0.8130	0.9277	0.8120
5-year average	1.9431	4.3592	13.2657	4.5086	4.5590	3.0536	0.8768	0.7011	0.8754	0.7006
(γ, ρ_1, θ)	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$	$(0.7; 0.8; -0.6)$	$(0.8; 0.8; -0.7)$
1st year	3.9067	4.6695	10.7659	9.4032	5.7851	5.3301	0.8889	0.9031	0.8882	0.9024
2sd year	3.9724	4.7426	15.0300	13.7074	6.2863	5.9621	1.0083	1.0315	1.0073	1.0306
3rd year	4.0064	4.7875	18.3457	17.3695	6.5754	6.3688	1.0731	1.1059	1.0720	1.1048
4th year	4.0330	4.8243	20.9653	20.5170	7.0843	7.0024	1.1199	1.1616	1.1182	1.1599
5th year	4.0578	4.8580	23.0548	23.2318	7.6913	7.7305	1.1559	1.2057	1.1534	1.2032
5-year average	3.9953	4.7764	17.6323	16.8458	6.6845	6.4788	1.0492	1.0816	1.0478	1.0802