




Spatial panel count data: modeling and forecasting of urban crimes

Stephanie Glaser¹ · Robert C. Jung¹  · Karsten Schweikert²

Received: 22 March 2021 / Accepted: 22 December 2021 / Published online: 21 February 2022
© The Author(s) 2022

Abstract

The steadily growing access to high-quality spatio-temporal crime count data with a high level of spatial detail allows to uncover interesting relationships between crime types within and between small regional units. Data coherent forecasting of such counts has to take the integer and non-negative nature of the data into account. Spatial panel data models that meet the criterion of coherency are relatively sparse. This paper proposes a new spatial panel regression framework with fixed effects to overcome these shortcomings. Depending on whether time dynamic effects are included in the model specification, estimation and inference are based either on a pseudo maximum likelihood method or on quasi-differenced generalized methods of moments. The models' usefulness is demonstrated in a forecasting exercise of monthly crime counts at census tract level from Pittsburgh, Pennsylvania.

Keywords Count data · Spatial panel models · Fixed effects · Predictive modeling

JEL Classification C33 · C53 · K42

✉ Robert C. Jung
robert.jung@uni-hohenheim.de

Stephanie Glaser
ssg.glaser@gmail.com

Karsten Schweikert
karsten.schweikert@uni-hohenheim.de

¹ Institute of Economics and Computational Science Lab (CSL), University of Hohenheim, Schloss Hohenheim 1 C, 70593 Stuttgart, Germany

² Core Facility Hohenheim & Institute of Economics, University of Hohenheim, Schloss Hohenheim 1 C, 70593 Stuttgart, Germany

1 Introduction

Modeling and forecasting criminal activities in urban areas is an important field of study in a range of social science disciplines like, e.g., economics, criminology, and geography. The results of which are increasingly used for public policy purposes. The steadily growing access to high-quality spatio-temporal data sets with a high level of spatial detail allows uncovering interesting relationships within and between small regional units, like, e.g., census tracts. Variables of interest often studied in this context are in the form of counts (Taddy 2010; Aldor-Noiman et al. 2016; Weinborn et al. 2017). In particular, data coherent forecasting of such crime counts has to take the integer and non-negative nature of the data into account. Most commonly, the observed counts are transformed into continuous data and well-established linear spatial panel models are employed (see e.g. Jin et al. 2020, for a recent contribution in this respect). Spatial panel data models that meet the criterion of coherency are relatively sparse, in particular in the econometrics literature.

A prominent and recent contribution in this area is Liesenfeld et al. (2017), who propose a Poisson random effects panel framework with a latent Gaussian spatio-temporal state process for the number of severe crimes. Inference for the model is based on a (high dimensional) likelihood function, which is not available in closed form and thus requires a Monte Carlo integration technique, like the efficient importance sampling (EIS) proposed by Liesenfeld et al. (2016). Their model is applied to monthly data for the 138 census tracts of Pittsburgh, PA, available from January 2008 to December 2013.

Complementary to this work, we propose a static and a dynamic version of a fixed effects spatial panel data model for counts and demonstrate how these models can be used for predictive purposes. The two model variants are fitted using pseudo maximum likelihood (PMLE) and generalized methods of moments (GMM), respectively. In comparing and contrasting our approach to that of Liesenfeld et al. (2017), we illustrate our approaches using the same data set on severe crime counts in Pittsburgh, PA.

Our modeling framework extends the standard Poisson fixed effects panel count data model, which is the workhorse model for longitudinal counts (see, for example, Cameron and Trivedi 2013, Ch. 9). As in the popular spatial autoregressive (SAR) model for continuous data, we include the spatially lagged counts into the Poisson conditional mean, which allows us to estimate spatial spillover effects along with other parameters of interest. We thereby provide a class of spatio-temporal panel count models for which the regressors and entity fixed effects do not necessarily have to be uncorrelated. Moreover, as pointed out by Elhorst (2014, p. 55) and Cameron and Trivedi (2013, Ch. 9) fixed effect specifications in spatial panel models have the conceptual advantage over random effects models in confining the analysis to explain the sample at hand, which is precisely the aim in our application.

Forecasting criminal activity in neighborhoods within a given specific city limit based on crime counts by using the spatial and temporal correlation of the data constitutes a relevant and critical focus in the literature. An excellent and up-to-date survey of the recent work done in this area is provided by Kounadi et al. (2020).

Also, the paper by Liesenfeld et al. (2017) is a rich source of highly relevant papers in this context.

To foreshadow our main findings, we qualitatively confirm the results reported in Liesenfeld et al. (2017). We obtain out-of-sample forecasting results and find that our fixed effects model performs competitively to theirs. This is a striking result considering the fact that the random effects model can incorporate socio-demographic indicators in yearly frequency, which are often good predictors for future crimes. Instead, these effects are absorbed by our entity and time fixed effects.

The paper's remainder is organized as follows: Section 2 introduces the proposed models and their respective estimation frameworks. Section 3 presents and discussed the results of our Monte Carlo simulation experiments. Our results of the empirical crime forecasting application are provided in Sect. 4 while Sect. 5 concludes.

2 Spatial panel count data models

The econometric framework proposed here is quite general, but takes into account the main features of the crime counts to be found in the empirical application below as documented in Chapter 2 of Liesenfeld et al. (2017).

We start with the static Poisson spatial panel model, which is a straightforward extension of the spatial autoregressive model proposed in Glaser and Jung (2021) to accommodate longitudinal data. It uses multiplicative fixed effects to account for individual heterogeneity. This model allows for incorporating both a contemporaneous and a lagged spatial term in the conditional mean function. We derive an approximate loglikelihood function using the concept of composite likelihood theory and obtain estimates for all relevant parameters of interest using standard numerical optimization procedures. Moreover, a predictive distribution based on the underlying Poisson distribution is readily available to form density forecasts of the counts.

Since crime counts often exhibit substantial autocorrelation (Li et al. 2014; Liesenfeld et al. 2017), we attempt to capture this dependence structure by including time dynamic effects into the spatial panel model and hope to improve the forecasting performance of the model. More precisely, we include the lagged outcome in the conditional mean equation. This approach extends the well-known linear feedback model of Blundell et al. (2002) to the spatial dimension. Since the spatial panel linear feedback model exhibits a dynamic panel structure and assumes entity fixed effects, we have to rely on quasi-differenced GMM to estimate the coefficients consistently (Blundell et al. 2002; Windmeijer 2005, 2008). The price to be paid for the richer dynamic structure is that only point forecasts based on this model are available.

2.1 A Poisson spatial panel model

2.1.1 Model specification, estimation, and inference

We start with a static predictive model for spatial count data which allows us to set up an approximate likelihood function for inferential purposes and obtain probabilistic forecasts. To achieve this goal, we choose a specification for the conditional mean function and the individual or entity-specific effect, which ultimately leads to the well-known form of the likelihood function in non-spatial panel models for counts as proposed in Hausman et al. (1984).

The general model specification is as follows: denote the discrete count variable to be explained with y_{it} , $i = 1, \dots, N$ and $t = 1, \dots, T$ and assume it follows a (conditional) Poisson distribution as it is standard in count data regression models. For the regression structure of the Poisson spatial panel model with multiplicative entity-specific effects v_i , we propose to use

$$E[y_{it} | \mathcal{F}_{it}, v_i] = \mu_{it} v_i, \quad (1)$$

where the conditioning information set \mathcal{F}_{it} contains appropriately defined vectors of the counts and additional predetermined regressors to be defined below in more detail.

To close the model, μ_{it} needs to be specified. The literature on count data regression and time series models for counts has developed a range of specifications to balance out modelling issues like avoiding explosive behaviour of the counts over time, problems with logarithmic transformations of zero counts and useful interpretations for the model parameters while preserving the non-negativity of the entire regression function. All of these have advantages and disadvantages while none of them has emerged as a standard (see the discussion in Glaser and Jung 2021). The specification proposed here is a straightforward extension of the regression framework used in Glaser and Jung (2021) for cross-section spatial counts and is also influenced by the so-called linear feedback model of Blundell et al. (2002) for non-spatial count data panels. We experimented with alternative specifications but found them to be inferior in our simulation experiments and empirical application.

The base variant for μ_{it} takes the form

$$\mu_{it} = \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{jt} + \exp(X_{it} \beta), \quad (2)$$

where $X_{it} = [x_{it1}, \dots, x_{itK}]$ is the row vector of explanatory variables and β is the corresponding column vector of regression parameters. w_{ij} is an element of the $N \times N$ time-invariant spatial weights matrix W , which is assumed to be row-standardized. In the spirit of the spatial autoregressive (SAR) model for continuous data, the first term on the right-hand side of Equation (2) captures possible spatial autocorrelations in the observed counts and therefore, ρ is called a (global) spatial

autocorrelation parameter. The regressors enter the conditional mean function in log-linear form as it is custom in count data regression models. Accordingly, the conditioning information set for this model specification takes the form $\mathcal{F}_{it} = \{X_{it}, y_{-it}\}$, where y_{-it} is the vector of observations of all neighbors to unit i in period t . With this specification, we implicitly define the nonlinear regression model

$$y_{it} = \mu_{it}v_i + u_{it} \tag{3}$$

for the observed counts, where the error term u_{it} is assumed to be uncorrelated with y_{jt} for $i \neq j$. Consequently, our model belongs to the broad class of conditional spatial regression models as discussed e.g. in Cressie (1993, p. 408), which enables us to use a conditional likelihood function to estimate the model parameters consistently. Note further, that (3) does not allow for direct interactions as in a linear SAR model for continuous data.

Still, within the static modeling framework, we consider an extension to the base variant specification, which allows for dependence between the observation y_{it} and the neighboring observations in the previous time period $t - 1$. This dependence is in addition to spatial effects in the cross-section of the data and is captured in the parameter λ as follows

$$\mu_{it} = \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij}y_{jt} + \lambda \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij}y_{jt-1} + \exp(X_{it}.\beta), \tag{4}$$

where the conditioning information set now is of the form $\mathcal{F}_{it} = \{X_{it}, y_{-it}, y_{-it-1}\}$.

To preserve the non-negativity of the conditional mean (1), restrictions on the parameter spaces of ρ and, if included, λ have to be imposed. Naturally, these restrictions depend on the exact values of W, X , and β and would have to be checked individually for each application. In principle, ρ and λ are allowed to be negative as long as the estimated conditional expectation is positive for each unit. Alternatively, the conditions $\rho, \lambda \geq 0$ can be set independently of the data, although the parameter space might then be restricted more strictly than necessary.¹

As discussed in Glaser and Jung (2021), the derivation of a full likelihood function based on the multiplicative structure (1) with conditional mean specification (2) or (4) is not possible due to their nonlinear form. Moreover, an operational high-dimensional multivariate count distribution is also not readily available. Therefore, as in Glaser and Jung (2021), we propose to approximate the likelihood using a pseudo-likelihood function as introduced by Besag (1975) and further developed into the composite likelihood approach of Varin et al. (2011). Within this framework, it can be shown that under suitable regularity conditions maximizing the conditional pseudo-likelihood yields consistent estimators for the parameters of interest.

As already outlined above, to cope with the unobserved entity-specific effect v_i , we follow the proposal of Palmgren (1981) and Hausman et al. (1984) and condition

¹ Throughout our simulation experiments and the empirical application, we follow the second strategy and require that $\rho, \lambda \geq 0$.

the pseudo-loglikelihood on v_i . Based on (4) and ignoring terms that do not depend on the model parameters, it is straightforward to arrive at the well-known function²

$$\log L_p(\rho, \lambda, \beta) = \sum_{i=1}^N \sum_{t=1}^T \left[y_{it} \log(\mu_{it}) - y_{it} \log \left(\sum_{s=1}^T \mu_{is} \right) \right]. \quad (5)$$

Note the special form of the (conditional) pseudo loglikelihood function which is related to that of the multinomial logit model as discussed in Hausman et al. (1984). The function (5) can be maximized using standard numerical methods and it is straightforward to obtain parameter estimates and their associated standard errors, if desired. Moreover, an estimate of the entity-specific effect v_i can easily be obtained as follows:

$$\hat{v}_i = \frac{\sum_{t=1}^T y_{it}}{\sum_{t=1}^T \hat{\mu}_{it}}, \quad (6)$$

where μ_{it} is either of the form (2) or (4).

2.1.2 Forecasting using the Poisson spatial panel model

As the focus of this paper is on forecasting, we now derive the minimum mean squared error (MMSE) predictor for the Poisson spatial panel model. We focus on one-step-ahead forecasts, which involves an evaluation of the conditional mean function (1) at $t = T + 1$ to start with. Due to its SAR structure which involves a contemporaneous spatial lag of the observed counts, we make use of the predictors suggested in the literature on SAR models for continuous data (see e.g. Kelejian and Piras 2017, Ch. 4). To simplify the exposition, we introduce a matrix notation for the one-step-ahead MMSE predictor as follows

$$E[y_{T+1}|y_T, X_{T+1}, v] \equiv f_{T+1|T} = \mu_{T+1|T} \circ v, \quad (7)$$

where $\mu_{T+1|T}$ is an $N \times 1$ vector of conditional means, $v = (v_1, \dots, v_N)'$, $y_{T+1} = (y_{1T+1}, \dots, y_{NT+1})'$ and y_T accordingly, X_{T+1} is a $N \times k$ matrix, with k being the number of regressors and \circ denotes the Hadamard product. $\mu_{T+1|T}$ can in principle be either of the form (2) or (4). For the exposition here, we employ the latter and present a forecasting approach within a model with contemporaneous and lagged spatial effects.

The conditional mean component of the MMSE predictor for the Poisson spatial panel model can then be written in matrix notation as

$$f_{T+1|T} = \left[\rho W f_{T+1|T} + \lambda W y_T + \exp(X_{T+1} \beta) \right] \circ v, \quad (8)$$

where the exponential function is applied element-wise to the vector $X_{T+1} \beta$. This

² Cameron and Trivedi (2013) provide a derivation of the (conditional) pseudo-loglikelihood function for Poisson panel models with fixed effects. They prefer to call it a 'concentrated' loglikelihood function after conditioning on v_i .

exponential can also be expressed via a power series as $\exp(A) = \sum_{k=0}^{\infty} 1/k!A^{ok}$, where A^{ok} is a Hadamard power. Solving (8) for the MMSE predictor, we then have

$$f_{T+1|T} = (I_N - \rho WD_v)^{-1} \{ [\lambda W y_T + \exp(X_{T+1} \beta)] \circ v \}, \quad (9)$$

where I_N is an $N \times N$ identity matrix and D_v is the diagonal matrix of v . A one-step-ahead prediction of counts can be obtained by replacing the parameter (vectors) in (9) with their estimated counterparts and element-wise multiplication by the estimated entity-specific effect. As the prediction is not necessarily an integer, an appropriate rounding procedure can be applied to produce point forecasts.

Data coherent forecasts can directly be obtained from the one-step-ahead density forecast

$$\hat{p}(y_{T+1|T}) = \text{Pois}(\hat{f}_{T+1|T}), \quad (10)$$

and data coherent point forecasts by taking, for example, its mode. The predictive models can now be compared across different model specifications using (proper) scoring rules, like, e.g., the logarithmic, the quadratic, and the ranked probability score, and evaluated using a number of tools based on the probability integral transform (PIT) method (see e.g. Czado et al. 2009; Jung et al. 2016). Specifically, it is straightforward to obtain the (nonrandomized) PIT for the probabilistic forecasts based on the Poisson spatial panel model to assess the distributional assumption. In the application below, we discuss the results of χ^2 goodness-of-fit tests of the null hypotheses of uniform distribution of the nonrandomized PITs (see e.g. Jung et al. 2016).

2.2 The spatial panel linear feedback model

2.2.1 Model specification, estimation, and inference

In many applications, the observed counts exhibit serial correlation. One way to accommodate this is to extend the static model specification above with time dynamic components to improve its forecasting performance. A parsimonious time dynamic specification introduces a lagged dependent variable into the regression equation (1). In the log-linear framework of count data regression models, this is, however, not straightforward to accomplish because the inclusion of lagged outcomes in the exponential function could lead to explosive series or to problems with transforming zero values (see e.g. Cameron and Trivedi 2013, Ch. 7). An elegant solution to the problem has been proposed by Blundell et al. (2002) in the context of a dynamic non-spatial panel model for counts. Their specification is called linear feedback model (LFM) and is rooted in the linear regression properties of certain integer valued autoregressive processes (INAR) for time series data (see e.g. Jung and Tremayne 2011, for a survey).

Upon this basis, we propose to specify the conditional mean of the so-called spatial panel linear feedback model with fixed effects to take the following form

$$E[y_{it}|y_{-it}, y_{t-1}, X_{it}, v_i] = \gamma y_{it-1} + \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{jt} + \exp(X_{it}\beta) v_i. \quad (11)$$

As in the LFM, the entity-specific effects are only multiplied with $\exp(X_{it}\beta)$ and can therefore be understood as a permanent scaling factor for the individual specific mean. Note that this specification represents a slight deviation from the static Poisson spatial panel model in Sect. 2.1. However, it has the conceptual advantage that the spatial panel linear feedback model reduces to the well-known LFM if $\rho = 0$.

As in the INAR framework, the autocorrelation parameter γ needs to fulfill $\gamma \geq 0$ to ensure non-negativity of the conditional expectation (11). In addition, $\rho \geq 0$ must be assumed as in the Poisson spatial panel model specification above. Note that unlike in the static specification, we do not include a lagged spatial term in our spatio-temporal model because the inclusion of such terms leads to identification problems (see, e.g., Anselin et al. 2008; Elhorst 2010, for a more detailed discussion).

Aside from the restrictions on the autocorrelation and spatial correlation parameters that guarantee a positive conditional expectation, the parameters also need to fulfill conditions to warrant stationarity. Stationarity conditions for miscellaneous dynamic spatial (panel) models are summarized in Elhorst (2012). Relevant for this model are the following conditions, derived in Parent and LeSage (2011) and Elhorst (2012), which imply that there is a trade-off between the autocorrelation and spatial correlation to maintain a stable model (Elhorst 2008):

$$\begin{aligned} \rho \omega_{max} - 1 < \gamma < 1 - \rho \omega_{max} & \quad \text{if } \rho \geq 0 \\ \rho \omega_{min} - 1 < \gamma < 1 - \rho \omega_{min} & \quad \text{if } \rho < 0, \end{aligned} \quad (12)$$

where ω_{min} and ω_{max} denote the smallest and largest characteristic root of the spatial weights matrix W . The largest characteristic root of row-standardized spatial weights matrices is unity by definition.

Estimation of the spatial panel linear feedback model can be based on quasi-differenced GMM (qdGMM). In the non-spatial framework, such a two-step GMM estimator is asymptotically efficient and consistent (Blundell et al. 2002; Windmeijer 2005). In order to apply the GMM method, the fixed effects have to be eliminated from the conditional mean (11) for which in principle both the Chamberlain (1992) and Wooldridge (1997) quasi-differencing transformations of the regression errors, defined to be $u_{it} = y_{it} - E[y_{it}|y_{-it}, y_{t-1}, X_{it}, v_i]$, are available.

As long as the regressor matrix X does not contain endogenous variables, both transformations can be applied. If the regressor matrix contains endogenous variables, the multiplicative specification of the fixed effects together with additive error terms cannot be estimated using the Chamberlain moment conditions because they would not be valid anymore (Windmeijer 2008). In contrast, the Wooldridge moment conditions are also valid for endogenous explanatory variables (Windmeijer 2000). Consequently, we base our discussion on the more general Wooldridge quasi-differenced errors (q_{it}), which are given by

$$\begin{aligned}
 q_{it} &= \frac{u_{it}}{\exp(X_{it}.\beta)} - \frac{u_{it-1}}{\exp(X_{it-1}.\beta)} \\
 &= \frac{y_{it} - \gamma y_{it-1} - \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{jt}}{\exp(X_{it}.\beta)} - \frac{y_{it-1} - \gamma y_{it-2} - \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{jt-1}}{\exp(X_{it-1}.\beta)}. \tag{13}
 \end{aligned}$$

For predetermined regressors ($E[X_{it}u_{it+j}] = 0, j \geq 0$ and $E[X_{it}u_{it-s}] \neq 0, s \geq 1$) it holds that

$$E[q_{it}|y^{t-2}, X_i^{t-1}] = 0, \tag{14}$$

with $y^{t-2} = [y_1, \dots, y_{t-2}]$ and $X_i^{t-1} = [X_{i1}, \dots, X_{i,t-1}]$. In contrast to the linear feedback model discussed in Windmeijer (2008), we note that it is necessary to condition on the past of all entities (y^{t-2}) instead of the individual past (y_i^{t-2}), since the spatial panel linear feedback model uses information from all neighbours.

These moment conditions guide the choice of instruments for the GMM estimation, which are gathered in matrix Z below. For the model considered here, two lags of the dependent variable, y_{t-2} and y_{t-3} , two lags of the simultaneous spatial term, Wy_{t-2} and Wy_{t-3} , and two lags of the predetermined regressor X_t, X_{t-1} and X_{t-2} , form the instrument matrix. Additional exogenous variables added to the model can also be included in Z .³

To reduce the instrument count and to limit the overfitting of the endogenous variables due to too many instruments, Roodman (2009) proposes the use of collapsed instruments. That is, moment conditions are summarized over t so that the estimator minimizes, for example, the magnitude of the empirical moment $\sum_t \Delta y_{it-2} q_{it}$ instead of the $T - 2$ empirical moments $\Delta y_{it-2} q_{it}, t = 3, \dots, T$. Using collapsed instruments leads to a more precise estimation of the optimal weight matrix in the second step and reduces bias (Roodman 2009). The instrument matrix then takes the form:

$$Z_i = \begin{bmatrix} y_{i1} & 0 & [Wy_1]_i & 0 & X_{i2} & X_{i1} \\ y_{i2} & y_{i1} & [Wy_2]_i & [Wy_1]_i & X_{i3} & X_{i2} \\ y_{i3} & y_{i2} & [Wy_3]_i & [Wy_2]_i & X_{i4} & X_{i3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{iT-2} & y_{iT-3} & [Wy_{T-2}]_i & [Wy_{T-3}]_i & X_{iT-1} & X_{iT-2} \end{bmatrix},$$

where $[Wy_t]_i$ denotes the i th row of the product Wy_t for $t = 1, \dots, T - 2$.

The quasi-differenced error terms and the instrument matrix enter the GMM function which is given by

³ While we use only linear moment conditions in this paper, in principle, it is possible to use nonlinear moment conditions as well to improve the efficiency of the GMM estimator for spatial models (Lee 2007). However, this is not straightforward without the guidance of further theoretical results, which, to the best of our knowledge, are not yet available in the literature.

$$Q(\theta) = \left(\frac{1}{N} \sum_{i=1}^N q_i(\theta)' Z_i \right) H^{-1} \left(\frac{1}{N} \sum_{i=1}^N Z_i' q_i(\theta) \right), \tag{15}$$

where $q_i(\theta) = (q_{i3}, q_{i4}, \dots, q_{iT})'$, θ is the vector of parameters to be estimated, and H is a weight matrix. This function is minimized with respect to the parameters θ in two steps to obtain consistent and efficient estimates. The weight matrix is set to be equal to the identity matrix for the first step and the weight matrix for the second step is calculated using the results from the first step such that

$$H(\hat{\theta}_1) = \frac{1}{N} \sum_{i=1}^N Z_i' q_i(\hat{\theta}_1) q_i(\hat{\theta}_1)' Z_i. \tag{16}$$

The asymptotic variance of the resulting efficient two-step GMM estimator ($\hat{\theta}_2$) is

$$\widehat{var}(\hat{\theta}_2) = \frac{1}{N} \left(C(\hat{\theta}_2)' H^{-1}(\hat{\theta}_1) C(\hat{\theta}_2) \right)^{-1}, \tag{17}$$

with

$$C(\hat{\theta}_2) = \frac{1}{N} \sum_{i=1}^N \left. \frac{\partial Z_i' q_i(\theta)}{\partial \theta} \right|_{\hat{\theta}_2}.$$

2.2.2 Forecasting using the spatial panel linear feedback model

To derive a one-step-ahead predictor, it should be noticed that it cannot be obtained straightforwardly. Unlike in the case of difference GMM for continuous data, the quasi-differenced dependent variable forecast cannot simply be added to the level of the previous period to obtain a level forecast. Instead, another equation has to be found which provides a forecast \hat{y}_{T+1} based only on values known at time T . Especially, the function is not allowed to depend on the multiplicative fixed effects v_i . Therefore, we employ the following equation which is based on the Wooldridge transformation for the model in $t = T + 1$,

$$\frac{y_{iT+1}}{\exp(X_{iT+1}.\beta)} - \frac{y_{iT}}{\exp(X_{iT}.\beta)} = \frac{\gamma y_{iT} + \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{jT+1} + u_{iT+1}}{\exp(X_{iT+1}.\beta)} - \frac{\gamma y_{iT-1} + \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{jT} + u_{iT}}{\exp(X_{iT}.\beta)}.$$

Rearranging yields an expression for y_{T+1} :

$$\begin{aligned}
 y_{iT+1} = & \left(\frac{y_{iT}}{\exp(X_{iT}.\beta)} - A_i \left(\gamma \left(\frac{y_{T-1}}{\exp(X_{iT}.\beta)} - \frac{y_T}{\exp(X_{iT+1}.\beta)} \right) \right. \right. \\
 & \left. \left. + \frac{\rho W y_T}{\exp(X_{iT+1}.\beta)} \right) \right) \exp(X_{iT+1}.\beta) \\
 & - A_i \left(\frac{u_{iT}}{\exp(X_{iT}.\beta)} - \frac{u_{iT+1}}{\exp(X_{iT+1}.\beta)} \right) \exp(X_{iT+1}.\beta),
 \end{aligned}
 \tag{18}$$

with A_i being the i^{th} row of the Leontief inverse $A = (I - \rho W)^{-1}$. Because of the assumed predetermination of X and Equations (13) and (14), the expected value of Equation (18) is given by

$$\begin{aligned}
 E[y_{iT+1}|y^T, X_i^{T+1}] = & \left(\frac{y_{iT}}{\exp(X_{iT}.\beta)} - A_i \left(\gamma \left(\frac{y_{T-1}}{\exp(X_{iT}.\beta)} - \frac{y_T}{\exp(X_{iT+1}.\beta)} \right) \right. \right. \\
 & \left. \left. + \frac{\rho W y_T}{\exp(X_{iT+1}.\beta)} \right) \right) \exp(X_{iT+1}.\beta),
 \end{aligned}
 \tag{19}$$

which can serve as a one-step-ahead predictor to obtain point predictions. Note that no straightforward density forecast is available within this framework.

3 Monte Carlo simulations

Before presenting an empirical application of both models proposed in the previous section, we provide some simulation evidence about key properties of parameter estimates in empirically relevant situations.⁴ According to the data-generating process implied by our proposed models, obtaining adequate simulated data sets requires the specification of a suitable spatial weights matrix W and draws of random numbers from the joint distribution of the counts which include the spatial effects.

As already discussed above, in the context of deriving a suitable likelihood function for the models under study here, such a discrete-valued joint distribution is not readily available. Following Glaser and Jung (2021), we therefore propose to produce the joint distribution to draw from by means of the Gibbs sampling algorithm of Geman and Geman (1984). This algorithm iteratively draws from the conditional distributions of y_{it} , $i = 1, \dots, N$ given all other y_j , $j \neq i$ at a given time t to obtain the joint distribution of y_1, \dots, y_n asymptotically. Thus, in the k^{th} iteration of this algorithm at time t , the data-generating process for the static model draws from the following distributions:

⁴ The subsequent analysis is conducted using R software. The programs can be obtained from the authors.

$$\begin{aligned}
 y_{1t}^{(k)} &\sim \text{Pois}(y_{1t}|y_{2t}^{(k-1)}, \dots, y_{Nt}^{(k-1)}) \\
 y_{2t}^{(k)} &\sim \text{Pois}(y_{2t}|y_{1t}^{(k)}, y_{3t}^{(k-1)}, \dots, y_{Nt}^{(k-1)}) \\
 &\vdots \\
 &\vdots \\
 y_{Nt}^{(k)} &\sim \text{Pois}(y_{Nt}|y_{1t}^{(k)}, \dots, y_{N-1t}^{(k)}).
 \end{aligned}$$

The starting values $y_{it}^{(0)}$ are drawn from a distribution with a non-spatial model specification.

For the generation of the spatial weights matrix W , we use a nearest neighbors inverse distance matrix.⁵ To compute it, N random coordinates on the two-dimensional Cartesian plane are generated from a $U(0, 100) \times U(0, 100)$ distribution. Then, we compute the Euclidean inverse distance matrix for these points, which represent the N spatial units of the simulated dataset. Furthermore, we select a threshold value to determine the nearest neighbors of each unit, all other entries of the matrix are set to zero and the resulting matrix is finally row-standardized. In our simulations, we consider two choices of the threshold value, $\pi \in \{25, 50\}$, to generate a sparse and a dense spatial weight matrix.

3.1 The Poisson spatial panel model

First, we study the properties of the static Poisson spatial panel model from Sect. 2.1. Parameter estimates for this model are obtained using the pseudo maximum likelihood method. The following data-generating process is employed

$$\begin{aligned}
 y_{i,t} &\sim \text{Pois} \left(\left[\begin{array}{c} \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{j,t} + \lambda \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{j,t-1} + \exp(\beta x_{i,t}) \end{array} \right] v_i \right), \\
 x_{i,t} &\sim i.i.d. N(0, \sigma_x^2),
 \end{aligned} \tag{20}$$

where $v_i = \exp \eta_i$ and $\eta_i \sim i.i.d. N(0, \sigma_\eta^2)$, $\sigma_x^2 = 0.5$, and $\sigma_\eta^2 = 0.5$. We consider panels with $T = \{4, 8, 16\}$ time periods and $N = \{100, 200, 400\}$ entities.

The results are reported in Table 1. Overall, we find that the results for both specifications of the spatial weight matrix are very similar with slightly better results for the less dense matrix. We find that the pseudo maximum likelihood estimator of both spatial effects is biased in small samples but the bias vanishes for larger samples. However, the estimator’s properties for contemporaneous and lagged spatial effects are affected differently by the increase of N and T . It generally seems to be more difficult to estimate the contemporaneous effect. Specifically, we report a smaller RMSE for the lagged spatial effect in the first specification with equal coefficient values. β can be estimated without problems, even in situations

⁵ We considered other specifications of the spatial weight matrix in our simulations, for example the 8-nearest neighbors inverse distance matrix, but this did not change our results substantially.

where the other coefficients are slightly biased. Similarly, the distribution of our entity-specific effects can be recovered using Equation (6). To verify this, we compute the mean entity-specific effect for each replication and compare it to its true value of $E(v_i) = \exp(\sigma_\eta^2/2) \approx 1.284$. We find that the mean of the estimated fixed effects converges to this number for increasing N and T .⁶ We conclude that spatial effects can be estimated consistently but not very precisely in small- N /small- T settings.

3.2 The spatial panel linear feedback model

The remainder of our simulation study provides some insights into the quasi-differenced GMM estimates' behavior for specifications of the spatial panel linear feedback model. To obtain a detailed picture of the model's properties, the values of the spatio-temporal parameters γ , and ρ are varied, and the resulting combinations are used in samples of different size, obtained by varying N and T . Four exemplary configurations of the parameters are chosen for this purpose, representing different combinations of autocorrelation and spatial correlation. The autocorrelation and spatial correlation parameters take the values: $(\gamma, \rho) := \{(0.5, 0.2), (0.5, 0.4), (0.7, 0.1), (0.2, 0.3)\}$. All specifications are chosen such that the stationarity conditions for spatio-temporal models are satisfied. Since we include a third lag in the instrument matrix, the minimum number of periods to estimate the model is four. Consequently, we consider the sample sizes $(T, N) := \{8, 16\} \times \{50, 100, 200, 400\}$ for all combinations of parameters. The simulation set-up closely follows Blundell et al. (2002), where the model contains one additional regressor x_{it} whose coefficient β is set to 0.5 in all configurations. We generate 1000 replications for each Monte Carlo experiment. The dependent variable outcomes are generated using the iterative Gibbs sampling algorithm of Geman and Geman (1984) which first iterates 50 times over the N units for time t and repeats this for all time periods discarding a burn-in of 50 time periods. After B iterations over all units $i = 1 \dots N$, the last vector is kept as y_t and is then used as a regressor for generating the outcomes for $t + 1$. The data-generating process is given by

$$\begin{aligned}
 y_{i,t} &\sim Pois(\gamma y_{j,t-1} + \rho \sum_{\substack{j=1 \\ j \neq i}}^N w_{ij} y_{j,t} + \exp(\beta x_{i,t} + \eta_i)), \\
 x_{i,t} &= \varrho_x x_{i,t-1} + \tau \eta_i + \varepsilon_{i,t}, \\
 x_{i,0} &= \frac{\tau}{1 - \varrho_x} \eta_i + \xi_i, \\
 y_{i,0} &\sim Pois(\exp(\beta x_{i,0} + \eta_i)),
 \end{aligned}
 \tag{21}$$

where the error term $\varepsilon_{i,t} \sim i.i.d. N(0, \sigma_\varepsilon^2)$, the entity effects $\eta_i \sim i.i.d. N(0, \sigma_\eta^2)$, and $\xi_i \sim i.i.d. N(0, \sigma_\varepsilon^2/(1 - \varrho_x^2))$. The spatial weight matrix W is again generated for each replication using randomly generated coordinates. Our results are based on the parameters $\beta = 0.5$, $\varrho_x = 0.5$, $\tau = 0.1$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.5$.

⁶ More detailed results are available from the authors upon request.

Table 1 Simulation results for the Poisson spatial panel model

N	T	$\pi = 25$						$\pi = 50$					
		RMSE			Bias			RMSE			Bias		
		$\hat{\rho}$	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\beta}$
$\rho = 0.2, \lambda = 0.2$													
100	4	0.220	0.146	0.159	0.025	-0.005	-0.003	0.275	0.185	0.183	0.041	0.028	0.019
200	4	0.177	0.107	0.124	0.019	-0.008	-0.001	0.228	0.110	0.137	0.031	0.004	0.007
400	4	0.147	0.071	0.096	0.011	-0.009	-0.003	0.178	0.080	0.106	0.010	0.003	-0.001
100	8	0.113	0.076	0.084	0.007	-0.035	-0.017	0.159	0.092	0.107	0.017	-0.025	-0.011
200	8	0.096	0.056	0.066	0.012	-0.025	-0.008	0.122	0.068	0.076	0.014	-0.014	-0.004
400	8	0.073	0.044	0.047	0.016	-0.019	-0.001	0.094	0.047	0.055	0.012	-0.009	-0.001
100	16	0.038	0.038	0.032	0.001	-0.013	-0.007	0.060	0.042	0.039	0.002	-0.012	-0.007
200	16	0.021	0.020	0.017	0.001	-0.004	-0.002	0.029	0.019	0.016	0.001	-0.002	0.000
400	16	0.008	0.005	0.006	0.000	0.000	0.000	0.008	0.006	0.004	0.000	0.000	0.000
$\rho = 0.2, \lambda = 0.5$													
100	4	0.253	0.170	0.159	0.084	-0.007	0.020	0.314	0.188	0.176	0.091	0.024	0.021
200	4	0.219	0.125	0.125	0.080	-0.003	0.027	0.273	0.134	0.147	0.076	0.011	0.026
400	4	0.190	0.088	0.101	0.079	-0.002	0.031	0.221	0.090	0.117	0.063	0.004	0.023
100	8	0.080	0.042	0.036	0.018	-0.009	0.002	0.110	0.052	0.056	0.025	-0.005	0.006
200	8	0.049	0.016	0.023	0.009	-0.001	0.003	0.071	0.022	0.030	0.012	-0.001	0.004
400	8	0.015	0.005	0.005	0.001	0.000	0.000	0.017	0.008	0.007	0.001	0.000	0.000
100	16	0.001	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
200	16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
400	16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 1 continued

$\rho = 0, \lambda = 0.4$													
100	4	0.166	0.217	0.142	0.085	-0.018	0.025	0.227	0.247	0.166	0.109	0.013	0.050
200	4	0.137	0.170	0.114	0.074	-0.014	0.025	0.181	0.185	0.128	0.092	0.008	0.047
400	4	0.112	0.118	0.082	0.068	-0.023	0.024	0.133	0.129	0.092	0.072	-0.009	0.032
100	8	0.097	0.132	0.078	0.057	-0.085	-0.015	0.144	0.136	0.092	0.085	-0.062	0.008
200	8	0.091	0.101	0.057	0.058	-0.069	-0.005	0.116	0.100	0.067	0.070	-0.040	0.014
400	8	0.080	0.072	0.042	0.051	-0.044	0.005	0.092	0.072	0.050	0.055	-0.027	0.014
100	16	0.041	0.077	0.035	0.015	-0.034	-0.009	0.060	0.073	0.036	0.021	-0.029	-0.004
200	16	0.023	0.036	0.015	0.005	-0.009	-0.002	0.033	0.036	0.017	0.008	-0.008	0.000
400	16	0.008	0.013	0.004	0.001	-0.001	0.000	0.010	0.011	0.005	0.001	-0.001	0.000
$\rho = 0.4, \lambda = 0$													
100	4	0.263	0.067	0.148	0.050	0.030	0.025	0.319	0.086	0.185	0.035	0.040	0.018
200	4	0.232	0.041	0.128	0.035	0.021	0.016	0.267	0.053	0.145	0.017	0.026	0.002
400	4	0.189	0.028	0.099	0.013	0.015	0.004	0.213	0.031	0.116	0.001	0.017	-0.003
100	8	0.143	0.033	0.079	0.007	0.015	0.005	0.193	0.052	0.103	0.015	0.025	0.012
200	8	0.108	0.026	0.061	-0.004	0.012	0.001	0.140	0.031	0.076	0.003	0.015	0.005
400	8	0.090	0.018	0.047	0.001	0.008	0.003	0.112	0.020	0.060	-0.005	0.010	0.000
100	16	0.041	0.014	0.023	0.000	0.004	0.001	0.071	0.021	0.038	0.003	0.006	0.003
200	16	0.018	0.006	0.011	0.000	0.001	0.000	0.031	0.008	0.016	0.001	0.001	0.001
400	16	0.005	0.001	0.003	0.000	0.000	0.000	0.005	0.002	0.003	0.000	0.000	0.000

Monte Carlo simulation results for the static spatial panel model with multiplicative fixed effects using pseudo maximum likelihood estimation. We draw 1,000 replications from the DGP outlined in Equation (20). π denotes the threshold values for the construction of the spatial weight matrix. The true parameter value of β is 0.5 for all simulations

Table 2 Simulation results for the spatial panel linear feedback model

N	T	$\pi = 25$						$\pi = 50$					
		RMSE			Bias			RMSE			Bias		
		$\hat{\gamma}$	$\hat{\rho}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{\beta}$
$\gamma = 0.5, \rho = 0.2$													
50	8	0.210	0.317	0.254	-0.155	0.214	-0.157	0.231	0.357	0.271	-0.176	0.259	-0.175
100	8	0.158	0.237	0.186	-0.096	0.127	-0.093	0.163	0.241	0.179	-0.101	0.133	-0.085
200	8	0.106	0.158	0.124	-0.050	0.063	-0.035	0.106	0.156	0.129	-0.050	0.060	-0.034
400	8	0.073	0.113	0.095	-0.026	0.030	-0.004	0.074	0.114	0.100	-0.024	0.027	-0.015
50	16	0.124	0.191	0.179	-0.092	0.132	-0.121	0.134	0.216	0.182	-0.102	0.157	-0.117
100	16	0.076	0.104	0.102	-0.044	0.051	-0.059	0.078	0.112	0.107	-0.047	0.059	-0.054
200	16	0.046	0.069	0.078	-0.018	0.017	-0.032	0.046	0.071	0.080	-0.019	0.021	-0.022
400	16	0.032	0.054	0.066	-0.009	0.005	-0.016	0.031	0.054	0.066	-0.010	0.008	-0.011
$\gamma = 0.5, \rho = 0.4$													
50	8	0.213	0.256	0.314	-0.153	0.177	-0.247	0.221	0.265	0.304	-0.157	0.181	-0.237
100	8	0.148	0.176	0.224	-0.088	0.097	-0.157	0.156	0.186	0.234	-0.091	0.102	-0.160
200	8	0.095	0.114	0.156	-0.037	0.040	-0.088	0.099	0.118	0.153	-0.038	0.039	-0.090
400	8	0.065	0.078	0.112	-0.013	0.013	-0.048	0.065	0.078	0.116	-0.017	0.017	-0.045
50	16	0.114	0.141	0.267	-0.079	0.095	-0.232	0.118	0.148	0.257	-0.089	0.109	-0.218
100	16	0.063	0.074	0.180	-0.030	0.030	-0.148	0.066	0.077	0.173	-0.038	0.040	-0.140
200	16	0.039	0.046	0.122	-0.010	0.006	-0.090	0.040	0.047	0.122	-0.015	0.012	-0.087
400	16	0.027	0.032	0.087	-0.004	0.000	-0.052	0.028	0.034	0.088	-0.004	0.001	-0.053
$\gamma = 0.7, \rho = 0.1$													
50	8	0.425	0.531	0.340	-0.369	0.453	-0.270	0.446	0.567	0.341	-0.401	0.501	-0.268

Table 2 continued

$\gamma = 0.7, \rho = 0.1$														
100	8	0.381	0.480	0.290	-0.324	0.400	-0.211	0.397	0.503	0.293	-0.341	0.424	-0.209	
200	8	0.322	0.403	0.232	-0.255	0.311	-0.144	0.340	0.426	0.235	-0.273	0.336	-0.138	
400	8	0.274	0.343	0.180	-0.204	0.250	-0.083	0.278	0.344	0.166	-0.213	0.258	-0.076	
50	16	0.247	0.304	0.250	-0.202	0.237	-0.198	0.272	0.346	0.256	-0.227	0.282	-0.204	
100	16	0.162	0.194	0.157	-0.117	0.129	-0.112	0.171	0.205	0.162	-0.122	0.137	-0.113	
200	16	0.101	0.119	0.102	-0.061	0.063	-0.059	0.102	0.118	0.098	-0.062	0.065	-0.051	
400	16	0.058	0.071	0.068	-0.028	0.029	-0.026	0.063	0.076	0.069	-0.031	0.033	-0.027	
$\gamma = 0.2, \rho = 0.3$														
50	8	0.106	0.253	0.198	-0.060	0.151	-0.078	0.104	0.269	0.198	-0.060	0.171	-0.082	
100	8	0.076	0.169	0.129	-0.034	0.076	-0.022	0.080	0.169	0.124	-0.041	0.090	-0.016	
200	8	0.055	0.127	0.107	-0.019	0.040	0.005	0.055	0.128	0.108	-0.022	0.041	0.001	
400	8	0.039	0.105	0.103	-0.010	0.021	0.017	0.040	0.109	0.111	-0.008	0.018	0.014	
50	16	0.078	0.193	0.133	-0.051	0.141	-0.042	0.075	0.199	0.128	-0.048	0.152	-0.036	
100	16	0.048	0.112	0.079	-0.025	0.068	0.004	0.049	0.114	0.082	-0.023	0.073	0.009	
200	16	0.034	0.087	0.079	-0.014	0.046	0.023	0.033	0.081	0.074	-0.014	0.048	0.024	
400	16	0.024	0.070	0.073	-0.008	0.031	0.025	0.024	0.069	0.073	-0.008	0.031	0.026	

Monte Carlo results for the dynamic spatial panel model with multiplicative fixed effects using the qdGMM estimator. We draw 1,000 replications from the DGP outlined in Equation (21). The instrument matrix includes time dummies. π denotes the threshold values for the construction of the spatial weights matrix

The results for our four main specifications are displayed in Table 2. Again, the difference between both specifications of the spatial weights matrix are marginal. The most striking result is that quasi-difference GMM estimation generally requires much larger samples to produce accurate coefficient estimates than the MLE for the static model. It appears difficult to estimate spatio-temporal effects in small samples. Blundell et al. (2002) also remark that small sample bias and imprecision are common problems with the quasi-differenced GMM estimator. Still, the quasi-differenced GMM estimator appears to be asymptotically unbiased and the RMSE decreases with the sample size if substantial spatio-temporal effects are present. While γ is on average overestimated, λ is underestimated for small to moderate sample sizes. As expected, γ converges faster by increasing the time dimension of the panel than by increasing the number of entities. We need a sufficient number of time periods to estimate γ precisely for more persistent model specifications. Unfortunately, wrongly estimated spatio-temporal effects using quasi-differenced GMM can have negative effects on the precision of β estimates (see, for example, the third panel of Table 2). It seems that the relative bias of β is positively related to the spatio-temporal persistence of the panel model.

To investigate how the degree of spatial autocorrelation affects the properties of the quasi-difference GMM estimator, we increase only the spatial autocorrelation going from the first to the second specification. The results show that the estimator still converges for all parameters if we move closer to the bounds of the stationarity region. More specifically, the relative bias of $\hat{\rho}$ is substantially lower in the second specification. The third specification is characterized by high temporal autocorrelation and low spatial autocorrelation. Here, we observe that $T = 8$ time periods are not enough to estimate the spatio-temporal effects properly. Not surprisingly, we find that increasing the time dimension to $T = 16$ leads to greater performance gains than doubling the number of entities. Finally, our fourth specification has temporal autocorrelation coefficient that is slightly lower than the spatial autocorrelation coefficient.⁷ In this case, we find that moving from $T = 8$ to $T = 16$ does not improve the estimates by much. However, using a combination of moderate N and T yields relatively precise results.

4 Application

The proposed models will now be applied in a predictive exercise using the same data set as in Liesenfeld et al. (2017) on severe crime counts in Pittsburgh, PA. We briefly describe the data set and then present empirical results from fitting the Poisson spatial panel model and the spatial panel linear feedback model. For further details about the data and the predictors to be included in the regression framework, we refer to Chapter 2 of that paper.

⁷ Since spatial autocorrelation is measured in a geographical scale and temporal autocorrelation is measured in a time scale, it is not straightforward to compare the speed of both dimensions of the dynamic space-time process.

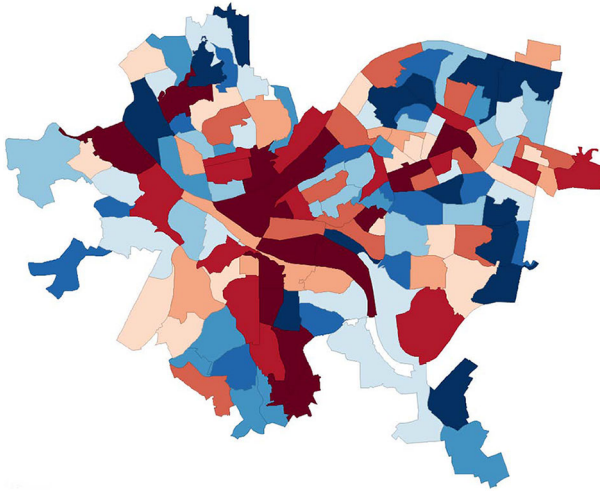


Fig. 1 Map of time averages of Part I crimes in Pittsburgh, grouped in deciles. Dark red (dark blue) areas indicate highest (lowest) deciles

Table 3 Descriptive statistics of the Pittsburgh crime dataset

	Mean	Std. Dev.	Min.	Max.
Number of Part I crimes	8.436	8.897	0	106
Number of Part II crimes	11.495	10.980	0	120
Number of Part I crimes				
2008	9.626	9.719	0	99
2009	9.059	9.474	0	106
2010	8.455	8.848	0	97
2011	7.722	7.885	0	84
2012	8.143	8.723	0	89
2013	7.612	8.232	0	102
Number of Part II crimes				
2008	13.283	12.463	0	120
2009	12.137	11.089	0	91
2010	11.201	10.681	0	91
2011	10.993	10.312	0	89
2012	10.992	10.457	0	83
2013	10.361	10.234	0	90

4.1 The data

The data set consists of crime counts for the 138 census tracts of Pittsburgh, Pennsylvania (in the borders for the 2010 census) sampled at a monthly frequency from January 2008 to December 2013. It contains 33 different crime types which are

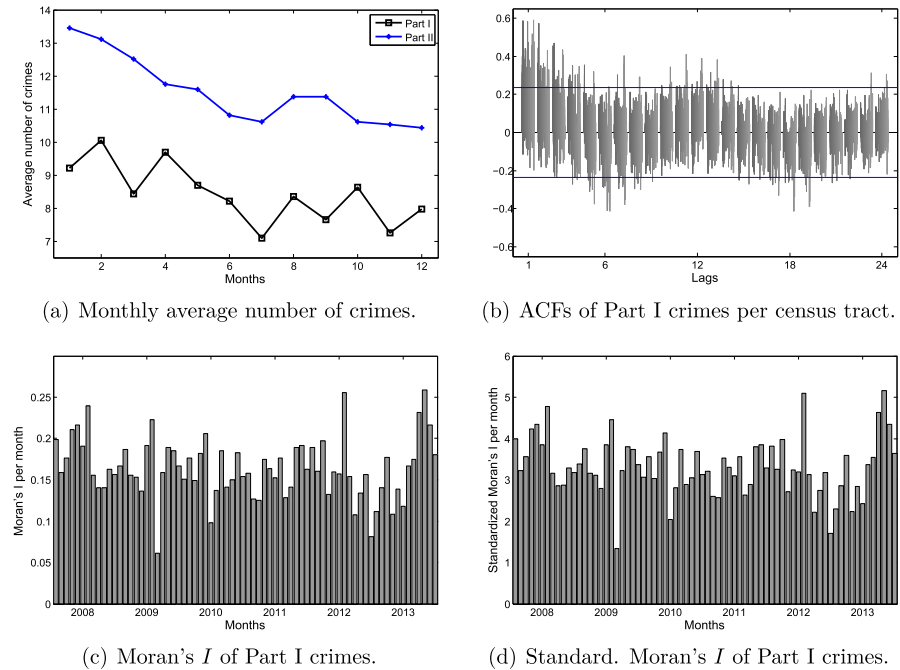


Fig. 2 Descriptives for Pittsburgh crime data. The blue horizontal lines in (b) indicate the 95 % confidence interval

reported separately and summarized into two categories according to the U.S. Department of Justice and Federal Bureau of Investigation (2004). Part I crimes encompass major crimes such as aggravated assault, burglary, larceny, motor vehicle theft, murder/manslaughter, negligent manslaughter, rape, and robbery. Part II crimes are the remaining minor offences like drunken driving, fraud, and vandalism, etc. (see U.S. Department of Justice and Federal Bureau of Investigation 2004, p. 8 for the full list). The logarithm of the lagged Part II crimes is used here as an explanatory variable for the Part I crimes models according to the so-called ‘broken-windows’ hypothesis (Wilson and Kelling 1982; Kelling and Coles 1996). It suggests that the intensity of less severe crimes in a neighborhood often precedes the occurrence of more severe crimes.

Figure 1 displays the geographical distribution of the average number of Part I crimes in Pittsburgh’s census tracts. Colors indicate the deciles of the data with dark red tracts belonging to the highest decile of Part I crimes and dark blue ones to the lowest deciles. The highest number of Part I crimes with an average of 76 per month is observed for census tract 201 (“Downtown”) followed by census tract 1702 (“Southside”), which lies to its southwest and has an average number of Part I crimes of 42. Small values are observed for tracts at the city border (dark blue areas in Fig. 1) and those with a small population density. Also, the dataset contains several cross-sectional variables measuring the socio-demographic characteristics of the inhabitants of each census tract, more specifically, their median income, the

fraction of the population age 18 or older, the fraction of the population which has a bachelor's degree or higher, and the fraction of female-led households. These variables are known from the criminology literature to indicate the criminal activity in a neighborhood and are used for forecasting in the random effects model proposed by Liesenfeld et al. (2017). Unfortunately, these variables are only available in yearly sampling frequency and would be picked off in our model by the inclusion of time fixed effects.

Table 3 reports selected descriptive statistics of the variables in the dataset for the whole sample period and for each year separately. On average, more Part II crimes than Part I crimes are observed per month and census tract, although the difference is quite small considering the large difference in the number of crime types in both classes. The highest amount of crimes was reported in 2008. A downward tendency is recognizable over time in this sample, although the differences between the years are small. 2013 has the lowest average crime rates.

Figure 2(a) gives a visual impression of the seasonal pattern of the crime counts. Generally, there are more crimes in the first half of the year and a hump in the Part II crimes during August and September. The autocorrelation functions (ACF) of the Part I crimes for each of the 138 census tracts are plotted in Fig. 2(b). They indicate significant but low autocorrelation at the first lag for most census tracts.

The spatial weights matrix used for this application is a queen contiguity matrix. Its entries are defined in the following: $w_{ij} = \frac{a_{ij}}{\#neighbours}$, $i, j = 1, \dots, N$, where $a_{ij} = 1$, if i is a neighbour of j (sharing a common border or a common vertex) and $a_{ij} = 0$, if i is not a neighbour of j . Each census tract has on average 5.6 neighbors. Finally, the spatial weights matrix is row-standardized.

To measure spatial association, we use Moran's I which was introduced by Moran (1950) for binary weights and generalised for arbitrary weight matrices by Cliff and Ord (1981, p. 17). It is the most prevalent measure of spatial association, quantifying the dependence across the entire dataset by summarising cross-products of deviations from the mean as follows

$$I_t = \frac{N}{W_0} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (y_{it} - \bar{y}_t)(y_{jt} - \bar{y}_t)}{\sum_{i=1}^N (y_{it} - \bar{y}_t)^2}, \quad t = 1, \dots, T, \quad (22)$$

where W is the chosen spatial weights matrix and $W_0 = \sum_{i=1}^N \sum_{j \neq i}^N w_{ij}$, which equals N for row-standardized weight matrices. The measure was developed for a cross-section of data, so we have to compute it repeatedly for each month, indicated by the subscript t . Unfortunately, there is no distributional theory available for Moran's I calculated from count data, but we can employ a bootstrap procedure to obtain inferential statements (Lin et al. 2011; Ren et al. 2014; Jin and Lee 2015). To do so, we draw elements of y_t , place it randomly on our map, and recompute I_t . Using 400 bootstrap draws, we can compare the original I_t statistic to the bootstrap distribution and determine the p -value.

Figure 2(c) and (d) show the values of Moran's I and standardized Moran's I for each month in the dataset. The lowest values are obtained for August 2009, June 2010, and December 2012. Using our bootstrap test, we are able to reject the null

Table 4 Estimation results for different specifications of the Poisson spatial panel model

Model Estimator	(1) PMLE	(2) PMLE	(3) PMLE	(4) qdGMM
Estimates for Jan. 2008 to Dec. 2012				
$PartI_{t-1}$				0.057*** (0.002)
$WPartI_t$	0.116*** (0.018)		0.124*** (0.016)	0.499*** (0.004)
$WPartI_{t-1}$		0.058*** (0.020)	0.051*** (0.012)	
$\log(PartIII_{t-1})$	0.123*** (0.012)	0.102*** (0.018)	0.128*** (0.008)	0.321*** (0.012)
<i>Jan</i>	-0.050 (0.231)	0.036 (0.217)	0.020 (0.160)	-0.341*** (0.037)
<i>Feb</i>	-0.006 (0.218)	0.048 (0.206)	0.025 (0.147)	0.001 (0.036)
<i>Mar</i>	0.055 (0.222)	0.089 (0.219)	0.072 (0.144)	-0.039 (0.036)
<i>Apr</i>	0.093 (0.289)	0.121 (0.242)	0.110 (0.193)	0.013 (0.036)
<i>May</i>	0.129 (0.243)	0.170 (0.215)	0.146 (0.165)	0.018 (0.036)
<i>Jun</i>	0.170 (0.248)	0.176 (0.231)	0.186 (0.166)	0.085*** (0.035)
<i>Jul</i>	0.058 (0.237)	0.025 (0.240)	0.036 (0.159)	0.094*** (0.036)
<i>Aug</i>	0.026 (0.229)	0.040 (0.217)	0.028 (0.153)	-0.006 (0.036)
<i>Sep</i>	-0.036 (0.219)	-0.038 (0.213)	-0.048 (0.146)	0.022 (0.036)
<i>Oct</i>	-0.080 (0.207)	-0.093 (0.203)	-0.092 (0.143)	-0.084*** (0.035)
<i>Nov</i>	-0.095 (0.186)	-0.108 (0.187)	-0.105 (0.127)	-0.089*** (0.036)
<i>Dec</i>	-0.341* (0.203)	-0.440* (0.253)	-0.414*** (0.157)	-0.068*** (0.034)

Table 4 continued

Model Estimator	(1) PMLE	(2) PMLE	(3) PMLE	(4) qdGMM
Log L	-279008.7	-279174.5	-278973.1	-
Sargan	-	-	-	0.050 [0.997]
Av. Forecast Results for Jan. to Dec. 2013				
RMSFE	4.296	3.851	4.187	4.978
MAFE	2.772	2.622	2.733	3.571

Results for the Poisson spatial panel model applied to Pittsburgh crime data using maximum likelihood estimation. Cluster-robust standard errors are given in parentheses. The *p*-value of the Sargan test is given in brackets. *W* is a queen contiguity spatial weights matrix. *, **, and *** denote 10%, 5%, and 1% statistical significance, respectively

Table 5 Point forecasting evaluation for the Poisson spatial panel model

Model	(1)		(2)		(3)		(4)	
	RMSFE	MAFE	RMSFE	MAFE	RMSFE	MAFE	RMSFE	MAFE
Jan 13	4.897	3.083	4.362	2.752+	4.734	2.973	4.020+	3.020
Feb 13	2.855	2.158+	2.936	2.278	2.825+	2.168	7.149	4.79
Mar 13	3.620	2.499	3.486+	2.481+	3.755	2.575	4.597	3.520
Apr 13	4.073	2.743	3.895+	2.733+	4.143	2.735	5.173	3.759
May 13	3.686	2.730	3.518+	2.666+	3.660	2.689	5.355	3.736
Jun 13	3.754+	2.669+	3.761	2.695	3.935	2.737	5.338	3.916
Jul 13	3.560	2.421	3.366+	2.372+	3.670	2.434	4.224	2.950
Aug 13	5.686	3.269	5.157	3.030+	5.655	3.274	4.874+	3.472
Sep 13	5.421	3.099	4.375+	2.713+	5.022	2.955	5.515	3.651
Oct 13	4.715	2.801	3.616+	2.554+	4.206	2.684	4.181	3.046
Nov 13	4.625	2.845	3.634+	2.521+	4.185	2.731	4.277	3.220
Dec 13	4.662	2.941	4.106+	2.669+	4.457	2.840	5.028	3.759
Average	4.296	2.772	3.851+	2.622+	4.187	2.733	4.978	3.571

Point forecast evaluation for the Pittsburgh crime data using static Poisson panel models. RMSFE stands for Root Mean Squared Forecast Error and MAFE stands for Mean Absolute Forecast Error, respectively[†] Indicates the smallest value of RMSFE and MAFE

hypothesis of no spatial dependence at the 5% level for all months except August 2009. Hence, these initial tests based on Moran’s *I* clearly point to spatial correlation in the data.

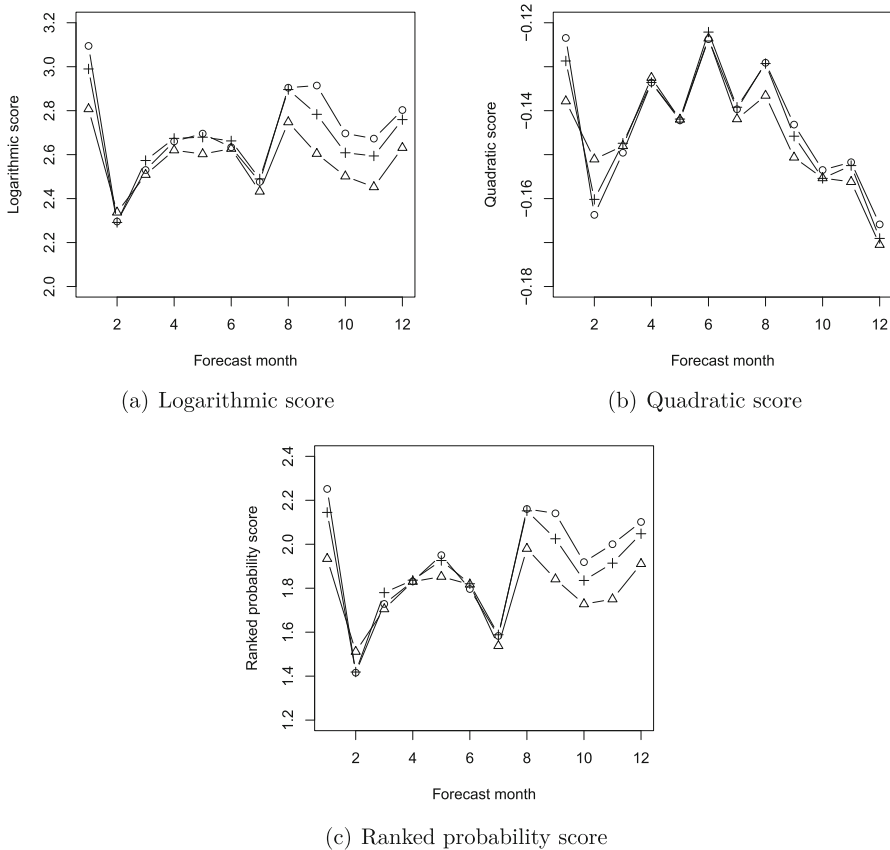


Fig. 3 Scoring rules of density forecasts for the Poisson spatial panel model with contemporaneous spatial term (\circ), with lagged spatial term (Δ), and with both spatial terms ($+$)

4.2 Forecasting Pittsburgh's crime counts using spatial panel count data models

In this section, we begin by applying static Poisson spatial panel models to Pittsburgh's crime counts. To forecast the number of Part I crimes in the census tracts of Pittsburgh for each month of 2013, a one-step-ahead expanding window forecast is employed. For this matter, we choose the estimation sample ranging from January 2008 to December 2012. Resulting parameter estimates of this first estimation sample are reported in Table 4. We employ different Poisson spatial panel model specifications with additional time fixed effects dummies for each month to capture seasonal effects. While Model (1) only contains the contemporaneous spatial term, Model (2) only contains the lagged spatial term, and Models (3) has the full spatial specification. The coefficient of lagged Part II crimes is highly significant in all specifications, and spatial terms are highly significant for each configuration. We find that the lagged spatial correlation parameter λ is

relatively small compared to the contemporaneous spatial term ρ . Its size remains stable, whether or not Part II crimes are included as a regressor, which entails information about the same period as the lagged spatial term.⁸ Time fixed effects are insignificant for all specifications of the static model. The summer months have the highest coefficient values, which seem to support the hypothesis that criminal activity increases with the temperature because a bad temper is evoked faster in hotter conditions (Gorr et al. 2003).

To compare and contrast our results with those provided in the literature, we first obtain point forecasts for the Part I crimes. The lower panel of Table 4 gives the root mean squared forecast error (RMSFE) and mean absolute forecast error (MAFE) averaged over the 12 months of the forecasting period (January to December 2013). Table 5 displays the forecast results for each month separately. In general, both measures show a very similar pattern across models. Still, we find very different predictions across model specifications for some months, for example, in February and June. It turns out that the model specification, which only includes the lagged spatial term, performs best in terms of RMSE and MAFE. This is surprising considering that the contemporaneous spatial effect was larger than the lagged spatial effect during the estimation period. Additional estimates (not reported) reveal that specifications with entity fixed effects clearly outperform pooled and time fixed effects models in terms of point forecast accuracy. Comparing these results to the one-step predictions obtained for the random effects model in Liesenfeld et al. (2017), which has an average RMSE (MAFE) over 2013 of 3.466 (2.535), we find that our fixed effects model performs competitively in terms of both measures. This is a striking result considering that the random effects model can incorporate socio-demographic indicators in yearly frequency, which are often good predictors for future crimes. Instead, these effects are absorbed by our entity and time fixed effects.

As pointed out in Sect. 2.1, density forecasts can also be obtained for the Poisson spatial panel models based on (10). Figure 3 displays various scoring rules of the density forecasts for all static model specifications over the different forecast months. All model specifications perform very similarly across all three scoring rules, and it is difficult to pick a preferred model specification based on this criterion. Using the logarithmic and ranked probability scores, we find that Model (2) outperforms Model (1) and Model (3) during the latter half of 2013.

Finally, we can use the nonrandomized PITs to check the distributional assumption's adequacy for the Poisson spatial panel model. The χ^2 goodness-of-fit tests of the null hypotheses of uniformity of the distribution of the nonrandomized PITs cannot be rejected for all forecasting months and all three model specifications.⁹ We conclude from these results, that the data do not reject the conditional Poisson assumption.

Now, we present the results based on the spatial panel linear feedback model (Model (4)) and compare our results to those for the static model (Models (1)–(3)). The relevant stationarity conditions from Equation (12) are fulfilled for the spatio-

⁸ Estimation results are not included but can be obtained from the authors upon request

⁹ The detailed results are not displayed here but are available upon request.

temporal model specification ($\omega_{min} = -0.333$, $\omega_{max} = 1$) and the Sargan test does not reject instrument exogeneity. The spatial regressor and the lagged Part II crimes are highly significant which is in line with our results for the static model where spatial autocorrelation coefficients are statistically significant over all specifications. We observe some substantial differences in the estimated coefficient values. Both the spatial autocorrelation coefficient and the coefficient of lagged Part II crimes are much higher for the spatial panel linear feedback model. However, to properly compare direct and indirect effects of lagged Part II crimes would involve the estimation of marginal effects (LeSage and Pace 2009), something that is not attempted at this stage. Strikingly, the coefficient of the dynamic effect, although significant, is quite small numerically.

We evaluate the model specifications' predictive performance by comparing them to one-step-ahead predictions for the static panel models in Table 5. Here, it becomes obvious that the inclusion of a dynamic effect leads to a worse forecasting performance for this specific dataset. Our dynamic models therefore also perform worse than the random effects model of Liesenfeld et al. (2017) in terms of RMSFE and MAFE. Together with the results from the Monte Carlo illustration these outcomes indicate that the quasi-differenced GMM estimation approach requires larger sample sizes and may not be useful for this specific application. Although the total sample size is moderate, we cannot benefit from the large- T panel setting if the spatial effects drives forecasting performance and the dynamic effect only contributes marginally. Blundell et al. (2002) also report similar shortcomings of the quasi-differenced GMM estimation for the non-spatial linear feedback model and propose to alternatively use a pre-sample mean estimator which has better small sample properties. However, for this estimator, a long time series of the dependent variable must be available prior to the actual estimation sample. Therefore, its applicability critically depends on the data availability, which is not given here.

5 Conclusion

This paper proposes spatial panel models for counts with fixed effects and apply them to forecast crime counts in an urban environment. While the static Poisson spatial panel model proved to be reliable in simulations with small sample sizes, the dynamic spatial panel linear feedback model requires rather large sample sizes to yield accurate results. The former model can be estimated using pseudo maximum likelihood estimation, while the latter is estimated using quasi-differenced GMM. The models are applied to forecast crime counts for the census tracts of Pittsburgh, PA in 2013. For this, the static model outperforms the dynamic model. Modeling the dynamic effect is outweighed in terms of forecasting performance by the less precise quasi-differenced GMM estimation. Our preferred model specification (Poisson spatial panel model with lagged spatial effects) performs comparably to the random effects model suggested in Liesenfeld et al. (2017).

The proposed spatial models for count data incorporate spatially lagged dependent variables to estimate a global spatial effect. Estimation of these models is relatively straightforward and only imposes moderate computational costs. These

aspects should foster the applicability of these models by applied researchers. Due to the inclusion of lagged observed counts, rather than lagged intensities, the interpretation of the measured spatial correlation is closer to the one usually encountered when working with linear spatial models for continuous data, making this class of count data models more accessible.

Acknowledgements We thank Manuel Huth, two anonymous referees, the editor, Lung-Fei Lee, and participants of the 8th International conference of the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics) in London for valuable comments and suggestions. Further, we thank Roman Liesenfeld, Jean-François Richard, and Jan Vogler for sharing their data.

Funding Open Access funding enabled and organized by Projekt DEAL. S. Glaser and R. Jung acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) under the Grant JU 2276/3-1.

Availability of data and material Data are publicly available as an online supplement to Liesenfeld et al. (2017).

Code availability The programs used to compute the simulation and empirical results can be obtained from the authors upon reasonable request.

Declarations

Conflict of interest Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aldor-Noiman S, Brown LD, Fox EB, Stine RA (2016) Spatio-temporal low count processes with application to violent crime events. *Stat Sin* 26(4):1587–1610
- Anselin L, Le Gallo J, Jayet H (2008) Spatial panel econometrics. In: Matyas, L., Sevestre, P. (Eds.), *The econometrics of panel data: fundamentals and recent developments in theory and practice*. Kluwer, Dordrecht, Ch. 19, pp 627–662
- Besag J (1975) Statistical analysis of non-lattice data. *Statistician* 24(3):179–195
- Blundell R, Griffith R, Windmeijer F (2002) Individual effects and dynamics in count data models. *J Economet* 108:113–131
- Cameron AC, Trivedi PK (2013) *Regression analysis of count data*, 2nd edn. Cambridge University Press, Cambridge
- Chamberlain G (1992) Comment: sequential moment restrictions in panel data. *J Business Econ Stat* 10(1):20–26
- Cliff A, Ord J (1981) *Spatial processes: models and applications*. Pion, London
- Cressie NAC (1993) *Statistics for spatial data*. Wiley, New York
- Czado C, Gneiting T, Held L (2009) Predictive model assessment for count data. *Biometrics* 65(4):1254–1261
- Elhorst JP (2008) Serial and spatial error correlation. *Econ Lett* 100(3):422–424

- Elhorst JP (2010) Applied spatial econometrics: raising the bar. *Spat Econ Anal* 5(1):9–28
- Elhorst JP (2012) Dynamic spatial panels: models, methods, and inferences. *J Geogr Syst* 14:5–28
- Elhorst JP (2014) Spatial panel data models. In: *Spatial econometrics*. pp 37–93
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 6:721–741
- Glaser S, Jung R. (2021) Modelling and diagnostics of spatially autocorrelated counts, Working Paper
- Gorr W, Olligschlaeger A, Thompson Y (2003) Short-term forecasting of crime. *Int J Forecast* 19:579–594
- Hausman J, Hall BH, Griliches Z (1984) Econometric models for count data with an application to the patents-R & D relationship. *Econometrica* 52(4):909–938
- Jin B, Wu Y, Rao CR, Hou L (2020) Estimation and model selection in general spatial dynamic panel data models. *Proc Natl Acad Sci* 117(10):5235–5241
- Jin F, Lee LF (2015) On the bootstrap for Moran’s I test for spatial dependence. *J Economet* 184(2):295–314
- Jung RC, McCabe BPM, Tremayne AR. (2016) Model validation and diagnostics. In: Davis, R. A., Holan, S. H., Lund, R., Ravishanker, N. (Eds.), *Handbook of Discrete-Valued Time Series*. Chapman and Hall/CRC, pp 189–218
- Jung RC, Tremayne AR (2011) Useful models for time series of counts or simply wrong ones? *ASTA Adv Stat Anal* 95(1):59–91
- Kelejian H, Piras G (2017) *Spatial econometrics*. Academic Press, Cambridge
- Kelling GL, Coles CM (1996) *Fixing broken windows: restoring order and reducing crime in our communities*. Free Press, New York
- Kounadi O, Ristea A, Araujo A, Leitner M (2020) A systematic review on spatial crime forecasting. *Crime Sci* 9:1–22
- Lee L-F (2007) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J Economet* 137(2):489–514
- LeSage JP, Pace RK (2009) *Introduction to spatial econometrics*. Chapman and Hall, London
- Li G, Haining R, Richardson S, Best N (2014) Space-time variability in burglary risk: a Bayesian spatio-temporal modelling approach. *Spatial Stat* 9:180–191
- Liesenfeld R, Richard J-F, Vogler J (2016) Likelihood evaluation of high-dimensional spatial latent Gaussian models with on-Gaussian response variables. In: Baltagi BH, LeSage J, Pace R (eds) *Spatial and spatiotemporal econometrics of advances in econometrics*, vol 37. Emerald, Bingley, UK, pp 35–77
- Liesenfeld R, Richard J-F, Vogler J (2017) Likelihood-based inference and prediction in spatio-temporal panel count models for urban crimes. *J Appl Economet* 32:600–620
- Lin KP, Long ZH, Ou B (2011) The size and power of bootstrap tests for spatial dependence in a linear regression model. *Comput Economet* 38(2):153–171
- Moran PAP (1950) Notes on continuous stochastic phenomena. *Biometrika* 37(1):17–23
- Palmgren J (1981) The fisher information matrix for log-linear models arguing conditionally in the observed explanatory variables. *Biometrika* 68:563–566
- Parent O, LeSage JP (2011) A space-time filter for panel data models containing random effects. *Comput Stat Data Anal* 55(1):475–490
- Ren T, Long Z, Zhang R, Chen Q (2014) Moran’s I test of spatial panel data model—based on bootstrap method. *Econ Model* 41:9–14
- Roodman D (2009) Practitioners’ corner: a note on the theme of too many instruments. *Oxford Bull Econ Stat* 71(1):135–158
- Taddy MA (2010) Autoregressive mixture models for dynamic spatial Poisson processes: application to tracking intensity of violent crime. *J Am Stat Assoc* 105(492):1403–1417
- U.S. Department of Justice, Federal Bureau of Investigation, (2004) *Uniform Crime Reporting Handbook*. Tech. rep. <http://www2.fbi.gov/ucr/killed/2009/aboutucr.html>
- Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Stat Sin* 21:5–42
- Weinborn C, Ariel B, Sherman LW, Dwyer EO (2017) Hotspots vs. harmspots: Shifting the focus from counts to harm in the criminology of place. *Appl Geogr* 86:226–244
- Wilson JQ, Kelling GL (1982) The police and neighborhood safety: broken windows. *Atlantic* 249:29–38
- Windmeijer F (2000) Moment conditions for fixed effects count data models with endogenous regressors. *Econ Lett* 68:21–24
- Windmeijer F (2005) A finite sample correction for the variance of linear efficient two-step GMM estimators. *J Economet* 126(1):25–51

- Windmeijer F (2008) GMM for panel count data models. In: Mátyás L, Sevestre P (eds) *The econometrics of panel data*. Springer, Berlin Heidelberg, pp 603–624
- Wooldridge JM (1997) Multiplicative panel data models without the strict exogeneity assumption. *Economet Theor* 13(5):667–678

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.