Faculty of Business, Economics and Social Sciences
University of Hohenheim

Modelling of Spatial Effects in Count Data

A thesis submitted by

Stephanie Glaser

to attain the degree of

Doctor oeconomiae (Dr. oec.)

Faculty of Business, Economics and Social Science
Institute of Economics
Chair of Econometrics and Statistics

Stuttgart-Hohenheim
2017

Die vorliegende Arbeit wurde im Jahr 2017 von der Fakultät Wirtschafts- und Sozialwissenschaften der Universität Hohenheim als Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften (Dr. oec.) angenommen.

Datum der Disputation: 7. März 2017

Dekan: Prof. Dr. Dirk Hachmeister
Prüfungsvorsitz: Prof. Dr. Katja Schimmelpfeng
Erstgutachter: Prof. Dr. Robert Jung
Zweitgutachter: Prof. Dr. Aderonke Osikominu

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Methods of spatial statistics have been widely applied in fields like biometrics and geostatistics after Whittle (1954) introduced the first spatial models. Spatial econometrics, however, has only been studied for the past 40 years. Paelinck and Klaassen (1979) published the first work which deals solely with this sub-discipline of econometrics. The characteristic property of spatial data in contrast to non-spatial data is that it links attributes to a geographic location (Fischer and Wang, 2011).[1] In case of a space-time data set information about time is included, too. Due to this additional information, it is possible to model dependencies between different observations which rely on geographical proximity. A similar concept in non-spatial modelling can only be found with regard to the time dimension. Hence, many concepts of spatial models have been inspired by the time series literature. Unfortunately, an essential property of time series does not hold in spatial modelling: Whereas time series have a natural ordering along the time line – from the oldest to the most recent observation – spatial data form a network which does not have a defined starting and end point. Concepts like predetermination, which often facilitates estimation in the time series context, generally do not exist in spatial modelling. Tobler's first law of geography summarizes this circumstance nicely: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). It also highlights another key assumption of spatial modelling: While it is anticipated that everything, for example every county or census tract, is connected to the other units through the spatial process, it is also assumed that this connection depends on the proximity of units, i.e. is diminishing with increasing distance.

The use of spatial models can either be motivated from a theoretical or practical viewpoint. In practice, data may show peculiar properties, e.g. spatial heterogeneity, that make the use of a spatial model specification advisable. From a theoretical perspective, spatial models can help to formalize the relations between agents which interact in such a way that aggregate patterns are observable (Anselin, 2002). Different reasons can be thought of why a spatial autoregressive structure might exist in the data at hand and why it should be modelled: First of all, one might originally be interested in modelling the interactions of agents or their reactions on previous decisions of neighbors which depend on the proximity (of geographic or other nature) to each other, e.g. trade flows. This can be accomplished by using panel data (often called space-time data) but also cross-sectional data can contain such dependencies. The observed state in a cross-section represents an equilibrium which has been constituted by iterative actions and reactions of neighboring units. But even if the actual research question does not suggest a spatial dependency in the data, it can be introduced by missing covariates. For example, if the outcome of these missing covariates is identical for several units, which are part of the same larger scale region, neglecting them

[1] This definition reflects the basic understanding of spatial data. In econometric modelling the geographic information is sometimes replaced by a different proximity measure, e.g. by technological proximity of industry sectors (Abdelmoula and Bresson, 2005, 2007).

will impose a positive spatial dependency on the dependent variable for which the actually included regressors do not account for. An empirical example for this is provided by Bhat et al. (2014) who suppose that the unobserved overall perceptions regarding the profitability of potential location decisions are similar for neighboring units, if they are defined on a small scale, since the perceptions apply to a larger region. LeSage and Pace (2009, pp. 25) offer a comprehensive discussion of these and other motivational aspects for spatial modelling.

Spatial econometrics as an own sub-discipline has borrowed much from spatial statistics, but still it focusses on different aspects of spatial processes. Spatial statistics (which is employed in fields like Biostatistics, Ecology, or Geography) is mostly interested in visualizing the spatial structure within the data (Kauermann et al., 2012). Therefore, formalized models containing non-spatial covariates are rarely used and estimation methods are strongly driven by their application. Spatial econometrics in contrast addresses two features of spatial data – spatial dependence between observations and spatial heterogeneity, i.e. "place-to-place" non-constant variance (Griffith and Paelinck, 2007). The focus often lies on estimating the size of spillover effects (Kauermann et al., 2012) usually in combination with effects of explanatory variables. This coincides with the scope of this thesis: The central purpose is to develop original count data models for spatial data which excludes models with log transformed counts or rates of counts as the dependent variable. These models need to estimate a (global) spatial effect, which allows that changes in the observation of one geographical unit potentially affect the outcomes of all other units. Additionally, the model must allow for explanatory variables aside the spatial terms which is a standard in econometric modelling.

A large variety of techniques have evolved in spatial econometrics, most of them for the modelling of continuous spatial data, but also other data types are considered. But still, the modelling of spatial count data, i.e. data which consists of non-negative integers, is in its infancy, with some propositions but few established methods. While spatial heterogeneity is included into count data models on a regular basis, spatial autoregression of the dependent variable is rarely addressed in count data applications so far. Seemingly, the only well established modelling strategy, which has been transferred from spatial statistics, is the modelling of a spatially correlated error term using a conditional autoregressive (CAR) scheme. There, the errors conditional on their neighbors are assumed to be normally distributed. However, this only induces a spatial structure in the error term, not in the observations, regarding it as a nuisance. Aside from that, the special structure of count data models has hindered the direct transfer of the model structure for continuous spatial data. Instead of spatially lagged dependent variables, spatially lagged explanatory variables are more often included in count data models. These are only able to represent local spatial effects and the explained part of spatial dependency, which will further clarified in Chapter 1.2.

The interest of this thesis lies in explicitly modelling a spatial structure in discrete valued count observations. More precisely, the aim is to estimate a global spatial autocorrelation parameter in the framework of a count data regression model. For this purpose, count data models are developed which incorporate spatial autocorrelation and are straightforwardly applicable for practitioners who benefit from a computationally hassle-free estimation of such a model. Previous proposals are often cumbersome to implement and ready-to-use packages for statistical software are not available for spatial count data models yet. To achieve this goal, cross-sectional and panel data models are presented and applied to a cross-sectional start-up firm births data set from the U.S. and a panel data set about crime in Pittsburgh.

The thesis is organized as follows: In the remainder of this chapter, a short introduction to spatial econometric modelling is given to explain the most important concepts of this subdiscipline. The related literature on econometric modelling of spatial counts is discussed in Chapter 2 with the focus on spatial autoregressive models. Chapter 3 contains an analysis of the Poisson spatial autoregressive model of Lambert et al. (2010), which serves as a starting point for the development of further so-called observation-driven spatial models. The cross-sectional spatial linear feedback model is introduced in Chapter 4 and applied to a start-up firm births data set, followed by spatial panel models in Chapter 5 which are employed to forecast crime counts for Chicago. Finally, Chapter 6 concludes.

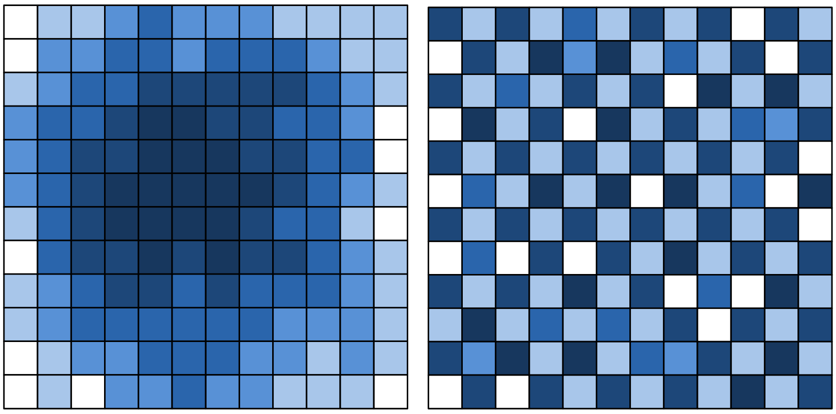
If not stated otherwise in the respective chapter the computations are executed in MATLAB[2] using code written by myself. For optimization the function `fminunc` with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method is used.

[2]The MathWorks, Inc., Natick, Massachusetts, United States.

1.2 Introduction to Spatial Econometric Modelling

Before starting with the investigation of spatial count data models, this section is intended to give an introduction to the most important concepts of spatial econometrics. These concepts are introduced for the standard case which are cross-sectional models for continuous data. In the last decade, the theoretical basics of spatial econometrics have been well documented in a number of textbooks from which the following summary is compiled (Arbia, 2014; Elhorst, 2014; Fischer and Wang, 2011; LeSage and Pace, 2009; Ward and Gleditsch, 2008).

Spatial dependence or spatial autocorrelation "reflects a situation where values observed at one location or region [...] depend on the values of neighboring observations at nearby locations" (LeSage and Pace, 2009, p. 2). This means that the observations either form clusters of similar values, which conforms to positive spatial correlation or, at the opposite extreme, a checkerboard pattern indicating negative spatial correlation, i.e. low values in one region foster high values in its neighboring regions. Figure 1 displays the stylized patterns of positive and negative spatial correlation.



(a) Positive spatial correlation (b) Negative spatial correlation

Figure 1: Stylized patterns of spatial correlation. Adapted from Fischer and Wang (2011, p. 24).

As mentioned before spatial data assigns geographical information to some attribute information. The space described by this information can be viewed as a continuous surface ('field view of space') or as being filled with discrete objects ('entity view of space'). In the field view, data can (theoretically) be measured at any point of a given surface having changing values across the surface. In the entity view, data is matched to one-dimensional objects (e.g. rivers or roads) or two-dimensional ones (e.g. counties or grid cells). Looking closer, four types of spatial data can be defined. First, geostatistical data conforms to the field view of space and is inherently continuous, but measured at a set of predefined points. A typical example is temperature which is measured at certain points but changes continuously over the earth surface. Second, point pattern data also consists of a set of point observations but those locations are not predefined. Instead, the points indicate the locations at which events of interest occur. Examples are locations of a certain tree species or of pedestrian casuality incidents. Third, area data consists of observations which are assigned to a fixed set of units. These units can either form a regular lattice, like the rectangles of an agricultural test field, or an irregular one, e.g. districts of a city. The unemployment rate of counties is a typical example. Fourth, spatial interaction or origin-destination flow data arises from measurements of interactions between two units, e.g. trade flows between countries. The last two types are commonly found in spatial econometric modelling. This thesis concentrates on area data, as does the rest of this section.

Before being able to model spatial effects using area data, the relationships between the n units of the lattice, i.e. the neighbors of each unit, need to be summarized in an $n \times n$ spatial weight matrix W. The elements of matrix W are positive if the corresponding units are neighbors ($w_{ij} > 0$ if $i \sim j, i \neq j$) and zero otherwise. The entries on the main diagonal are zero by convention ($w_{ii} = 0 \ \forall i$). If the matrix is symmetric, it means that each unit is a neighbor of its neighbors ($w_{ij} = w_{ji}$). Although this is a very intuitive concept of neighborhood, we will see later that depending on how the neighbors of a unit are determined, asymmetric matrices are employed as well. Spatial weight matrices used in spatial econometrics are usually transformed to be "row-stochastic" (also called "row-normalized"), meaning that each row sums up to one ($\sum_{j=1}^{n} w_{ij} = 1 \ \forall i$).

Broadly three types of spatial weight matrices can be distinguished according to the concept of neighborhood employed. The entries of a contiguity matrix equal one if the respective units share a common border (rook contiguity), a common vertex (bishop contiguity), or either one of them (queen contiguity). This is a very intuitive way of defining neighborhood leading to a symmetric matrix (before row-normalization). The reasoning behind it is that there must be a point of contact between units to enable them to affect each other. A k-nearest neighbor matrix follows a different idea and contains ones for the k closest neighbors to each unit (usually measured at the centroid). Here, the units being denoted as neighbors do not need to have a common border or vertex. On the one hand, depending on the structure of the lattice, this can lead to neighbors which actually lie far away from each other. On the other hand, this specification ensures that each unit has the same number of neighbors. The resulting weight matrices are in general not symmetric. For the third class of weight matrices the geographical or otherwise defined distance is used to compute the weights. Typically, the inverse of the distance between two units, or a function of it, serves as the weight. Giving higher weights to closer units and smaller weights to units further away this complies with Tobler's much cited first law of geography (Tobler, 1970, p. 236). Again, for this concept of neighborhood it is irrelevant whether the units share boundary points or not, but it leads to a symmetric weight matrix (before row-normalization). A full inverse distance matrix defines relationships between all units on the lattice. This is not plausible for all applications and may also result in computational difficulties for large data sets. An alternative to the full inverse distance matrix is to set a distance threshold, up to which units are considered neighbors. In the remainder of this thesis the spatial weight matrix is assumed to be predetermined and not part of the unknown parameter set.

There are different parts of a linear regression model where a spatial component can be incorporated using the previously described weight matrices. If the spatial dependence in the data is assumed to be a nuisance resulting in spatial correlation in the error terms, spatial error models (SEM), also called spatial heterogeneity models, can be used to account for this data property and to achieve more efficient estimation, especially in small samples. The SEM in matrix notation is

$$y = X\beta + \epsilon \tag{1}$$

$$\epsilon = \rho W \epsilon + u \Leftrightarrow \epsilon = (I - \rho W)^{-1} u \tag{2}$$

where y is the vector of the dependent variables for the n units, the vector u contains i.i.d. error terms, W is exogenous and row-standardized, and X is a matrix of explanatory variables with parameter vector β. To ensure the existence of the inverse, the spatial autoregressive parameter must fulfil $|\rho| < 1$. The model equations can be reduced to

$$y = X\beta + (I - \rho W)^{-1} u. \tag{3}$$

And the resulting covariance matrix is given by

$$E[uu'] = \sigma_u^2 (I - \rho W)^{-1} (I - \rho W')^{-1} = \sigma_u^2 \Sigma \tag{4}$$

Equation (3) visualizes that there is only a spatial structure in the unexplained part of the dependent variable in a SEM, the explanatory variables are supposed to be spatially uncorrelated. Assuming that the error terms u_i are i.i.d. normally distributed, the parameter estimates can be obtained using maximum likelihood estimation. Alternatively, a feasible generalized least squares procedure derived by Kelejian and Prucha (1998) can be employed.

The next two models explicitly model spatial effects in the explained part of the model. The spatially lagged covariates (SLX) model incorporates the regressors of the neighbors into the model equation:

$$y = X\beta + W\bar{X}\boldsymbol{\gamma} + \epsilon \tag{5}$$

where $\bar{X}$ are the explanatory variables excluding the constant and $\boldsymbol{\gamma}$ is a parameter vector instead of a scalar like ρ in the SEM. ϵ is a vector of i.i.d. error terms. If the spatial weight matrix W is row-stochastic then $W\bar{X}_k$ is a weighted average of the neighboring observations of the k^{th} explanatory variable with γ_k being the corresponding parameter. Because the spatial effect is only introduced through spatially lagged regressors, neither endogeneity nor heterogeneity are induced and estimation of the model need not to be adapted. In the case of the classical linear regression model this means estimation can be conducted with ordinary least squares.

Finally, the spatial autoregressive (SAR) model includes spatial effects of the dependent variable:

$$
\begin{aligned}
y &= \lambda W y + X\beta + \epsilon \Leftrightarrow \\
y &= (I - \lambda W)^{-1} X\beta + (I - \lambda W)^{-1}\epsilon
\end{aligned}
\tag{6}
$$

where λ is the parameter of spatial autocorrelation in the dependent variable and ϵ_i is i.i.d. Equation (6) gives the reduced form of this model. The endogenous regressor Wy is usually named spatially lagged dependent variable and, in the case of a row-stochastic weight matrix, equals a weighted average of the neighboring observations of the dependent variable. All other entries of W including the elements of the main diagonal are zero, meaning that only observations of neighbors enter the average. The domain of the spatial autocorrelation parameter λ depends on the minimum and maximum eigenvalues of W, $/\omega_{min}$ and $/\omega_{max}$. To ensure stationarity it should lie in the interval $(1/\omega_{min}, 1/\omega_{max})$. If W is row-stochastic, $-1 \leq \omega_{min}^{-1} < 0$ and $\omega_{max}^{-1} = 1$ hold and the spatial autocorrelation parameter ranges from negative values to unity, making interpretation and comparison more comfortable. Like the SEM this model can be estimated via maximum likelihood assuming normality for the error terms. A two-step least squares procedure with $[X, WX, W^2X, W^3X, \ldots, W^pX]$ as instruments has also been developed.

Whether the spatial dependence is supposed to be part of the error process and regarded as a nuisance or if it is of explicit interest and modelled as spatial dependence of X or y depends on the specific application and the interests of the researcher. The choice between a SLX and a SAR model also entails whether a local or global spatial effect is modelled. The SLX model contains the spatial term WX which causes a change in $X_{i.}$ to potentially affect all neighbors of i but not the rest of the units on the lattice. This is denoted as a local spatial effect (Anselin, 2003, pp. 156). The reduced form of the SAR model contains $(I - \lambda W)^{-1}$, called spatial multiplier or Leontief inverse, multiplied with the explanatory variables X. Through this inverse, a change in a regressor of one of the units does not only affect the direct neighbors of that unit but potentially all units on the lattice. To clarify this, the infinite series expansion of the inverse is helpful:

$$
(I - \lambda W)^{-1} = I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 \ldots \tag{7}
$$

From this, it can be seen that this inverse does not only contain weights for the direct neighbors, which are λW, but also for the neighbors of the neighbors, $\lambda^2 W^2$, and so on. This is called a global spatial effect (Anselin, 2003, pp. 155) since a change in one unit leads to changes in potentially all other units. Whether or not all units will be affected and the strength of these effects depend on the position on the map of the changing unit, the amount of connectivity between the units (specified in W), and the size of λ and β.

To visualize the different effects of a change in regressor matrix X in a SAR model, Figure 2 displays a regular 7×7 square grid. The observed effect is split up into different unobservable intermediate steps, of which three are explained here, denoted by y^I, y^{II} and y^{III}. Only the final outcome of the dependent variable in an equilibrium state is observed. Figure 2(a) displays the initial change of y_i caused by a change in x_{il}:

$$y^I = \lambda W y + (X + \Delta X)\beta + \epsilon \qquad \Delta x_{il} \neq 0,\ \Delta x_{ik} = 0 \quad \forall k, l = 1, \ldots, K,\ l \neq k \tag{8}$$

This is the effect of a change in x_{il}, which would also be observed in a non-spatial linear model, and equals β_l if $\Delta x_{il} = 1$. Due to the spatial term Wy a change in y_i affects the outcomes of all neighbors of unit i, i.e. a spillover effect occurs (Figure 2(b)):

$$y^{II} = \lambda W(y + \Delta y^I) + (X + \Delta X)\beta + \epsilon \qquad \Delta y_i^I \neq 0,\ \Delta y_j^I = 0 \quad \forall i, j = 1, \ldots, N,\ j \neq i \tag{9}$$

This part of the spillover effect is a local effect like it is obtained in SLX models because only the immediate neighbors of unit i are affected. Looking at one particular neighbor of unit i (Figure 2(c)), the change in the outcome variable of the neighbors now itself leads to spillover effects to the neighbors' neighbors (Figure 2(d)) which also leads to an additional change in the outcome of unit i. The latter is called a feedback loop:

$$y^{III} = \lambda W(y + \Delta y^I + \Delta y^{II}) + (X + \Delta X)\beta + \epsilon \qquad \Delta y_j^{II} \neq 0 \text{ if } j \sim i, 0 \text{ otherwise} \tag{10}$$

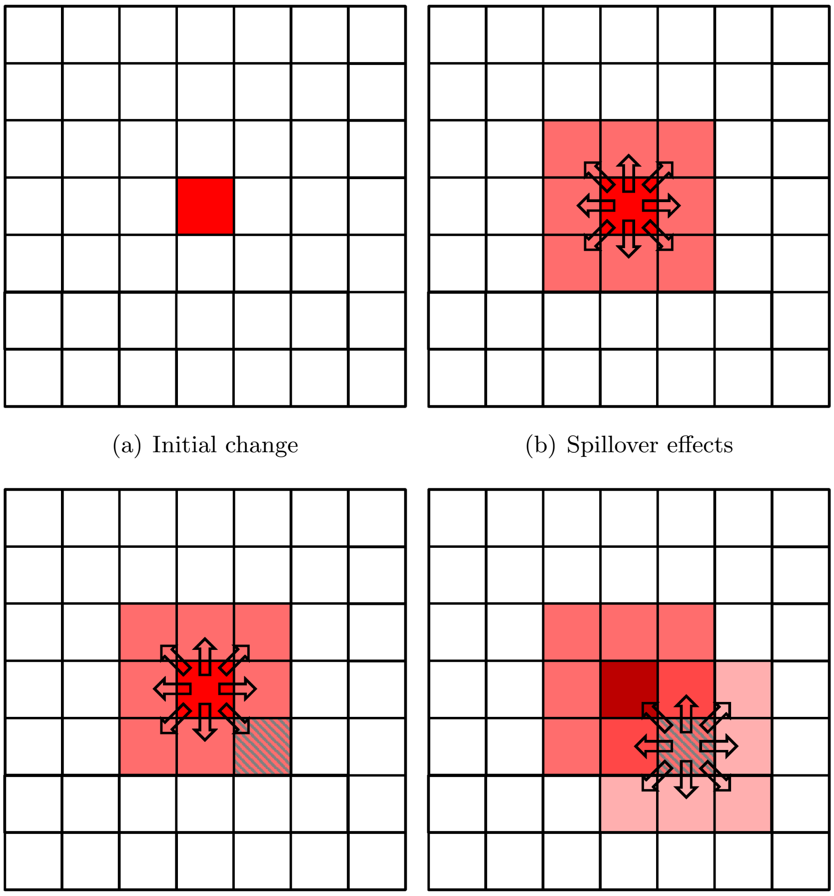
These repeating spillover effects spread over the whole map and form a global spatial effect. Also, feedback loops with longer paths, e.g. from observation i to j to k and back to i, are part of the entire effect. It is important to note that the models introduced here assume a simultaneous dependence system, i.e. simultaneity of all these effects. The resulting (observed) value of the dependent variable can be seen as an equilibrium outcome or steady state.

Finally, some remarks about the interpretation of the effect of covariates are necessary. Since the SEM model does not allow for spillover effects of a change in X, the interpretation of β is the one known from the classical linear model. Similarly, the parameters β of the SLX model measure the direct effects, whereas the vector γ equals the size of the spillover effects of a change in a neighbor unit. To evaluate the size of the effect of a change in X in the SAR model, marginal effects, usually called spatial impacts in the spatial literature, have to be calculated. They are obtained by deriving Equation (6). The direct marginal effect, i.e. the effect of a change in x_{ik} on the outcome of the same unit, y_i, is given by

$$\frac{\partial y_i}{\partial x_{ik}} = a_{ii}\beta_k \tag{11}$$

where a_{ii} is the according element of the Leontief inverse $A = (I - \lambda W)^{-1}$. Correspondingly, the indirect marginal effect of a change in the regressor x_{ik} on the outcome of a different unit j, is

$$\frac{\partial y_j}{\partial x_{ik}} = a_{ji}\beta_k \qquad i \neq j \tag{12}$$



(a) Initial change (b) Spillover effects

(c) Choosing one neighbor (d) Spillover effects and feedback loop

Figure 2: Components of the spatial autoregressive effect.

LeSage and Pace (2009) introduced several impact measures based on marginal effects. To obtain a measure comparable to the β in a classical linear regression model, averages of the effects are taken. The "average direct impact" gives the average change in y if regressor x_k of the same unit changes.

$$\bar{M}(k)_{direkt} = \frac{1}{n} \sum_{i=1}^{n} a_{ii} \beta_k \tag{13}$$

The "average total impact" summarizes either the effects of a change in x_{ik} on the dependent variables of all units ("impact from an observation") or the effects of changes (of the same size) in the regressors $X_{.k}$ of all units on y_i ("impact to an observation"). Both averages are numerically equal and only reflect two ways of interpreting the average total impact.

$$\bar{M}(k)_{total} = \frac{1}{n} \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ij} \beta_k \right) = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} \beta_k \right) \tag{14}$$

Eventually, the "average indirect effect" is given as the difference of average total and average direct impact.

$$\bar{M}(k)_{indirekt} = \bar{M}(k)_{total} - \bar{M}(k)_{direkt} \tag{15}$$

2 Literature on Spatial Econometric Models for Count Data

2.1 Introduction

Spatial models, at least for continuous dependent variables, have found broad application in econometrics during the last 30 to 40 years (for a survey see e.g. Anselin (2010) or Lee and Yu (2009)). With regard to count data analysis the most widely used approach is the modelling of spatial heterogeneity. Spatial autocorrelation (SAR) models, in contrast, are not studied extensively and the propositions of such a model did only seldom find application by others than the authors themselves. The obvious reason for a lack of SAR models for count data is that unlike in classical models for continuous data, there is no direct functional relationship between the dependent variable y and the regressors X. To illustrate this, the general specification of a count data model is given:

$$y|\mu, \theta \sim D(\mu, \theta), \quad \mu = \exp(X\beta) \tag{16}$$

with X being a matrix of exogenous variables and β the corresponding parameter vector. D stands for an arbitrary distribution suitable for count data with intensity μ and optional further parameters θ. The most common special cases of this class are the Poisson regression model with $y|\mu \sim Po(\mu)$ and the negative binomial regression model with $y|\mu, \alpha \sim NB(\mu, \alpha)$. The negative binomial model deals with a restriction of the Poisson model namely its equidispersion.

From Equation (16) we see that instead of y the intensity parameter μ, which equals the conditional expectation $E[y|X]$, is a function of the regressors. Because of this peculiarity of count data modelling, a direct transfer of the spatial model types for continuous data, introduced in the previous section, is not possible. In the following of this chapter, several ways to handle this are reported. Aside from spatial error models, spatially lagged covariate models (SLX) can be used to consider spatial structures without dealing with the problems created by including endogenous spatial terms into the functional form Equation (16). Nevertheless, the focus of this literature review are approaches introducing a SAR-like structure into count data models since the main focus of this thesis lies in modelling global spatial effects.

Spatial count data is very common in other disciplines including ecological statistics, biostatistics, and epidemiology for example. Articles from these areas have also been considered in the following if they meet the conditions set for a spatial *econometric* model. First, spatial econometric data is usually given on a (irregular) lattice (see Figure 2 in Chapter 1.2 for an example of a regular lattice and Figure 9 in Chapter 3.6.1 for an irregular one). Point processes, which are for example common in ecological statistics (plant counts), are therefore excluded from the survey. Second, spatial econometric models usually aim at estimating a parameter of spatial autocorrelation from the data and identifying spatial spillover effects. On the contrary, in spatial statistics the focus often lies on visualizing a spatial process (Kauermann et al., 2012, p. 437), for example in disease mapping which is a very common application for spatial count data modelling (a survey can be found e.g. in Best et al. (2005)). The examined SAR models therefore all include such a parameter. Third and last, econometric modelling is almost always concerned with the effect of covariates on the dependent variable. Because of this, the following models must allow the analysis of the influence of non-spatial covariates as well. Having said that, a natural condition for all models in this thesis is that they model original count data and do not use linear approximations like log transformed counts or rates of counts.

The following shall give an overview of the literature on spatial modelling of count data and the applications for which such models are employed. For SLX and spatial heterogeneity (SEM) modelling only examples of models and applications are given. In contrast, the survey of spatial autoregressive models gives to my knowledge a full picture of the approaches documented in the literature. Models are presented with a focus on the approach of introducing a spatial structure into the model. For all other information regarding more details on model specification, distributional theory, and details on the pursued estimation strategy the reader is referred to the cited articles.

2.2 Spatially Lagged Covariates Models

The easiest way of incorporating a spatial structure into a model is via its covariates. This way, spatially lagged or otherwise spatial regressors can be computed before the actual regression is performed and be treated the same way as the non-spatial ones. In the following, two examples of the use of spatially lagged covariates in a count data setting are described without going into detail regarding the actual models employed.

Buczkowska and de Lapparent (2014) use an SLX model for the location choices of new establishments in the Paris metropolitan area. They investigate different industry sectors and check several count data models. The results of a Poisson hurdle model with spatial spillover effects are reported in the article. The spillover effects are calculated prior to the estimation as a regressor (p. 76):

$$X_{l,s} = \log(\sum_{j=1}^{L} e^{-d_{l,j}} z_{j,s}) \tag{17}$$

where $j = 1, \dots, L$ are the spatial units in the data set, $d_{l,j}$ is the distance between the centroid of unit l and j and $z_{j,s}$ is an attribute of unit j that applies to industry sector s, e.g. the number of pre-existing establishments. The inclusion of $X_{l,s}$ into the intensity equation of the model therefore introduces a spatial effect. But due to its predetermined nature, it does not have any consequences on the estimation of the model, which is still done using conventional estimation strategies for non-spatial models.

A different approach of using spatially lagged regressors for counts is employed by Abdelmoula and Bresson (2005, 2007). They use a panel linear feedback model for count data (introduced by Blundell et al. (1995)) to model spillover effects of R&D expenditures on patent activity. In their linear model equation, which is estimated with quasi-differenced generalized method of moments (GMM) (Blundell et al., 2002), the number of patents is a function of the R&D expenditures of the other regions. The R&D expenditures of the other regions are summarized into K geographical distance classes, each with its own elasticity parameter λ_k. The resulting spatial term is

$$\sum_{k=1}^{K} \lambda_k \log R_{t-1,k} \tag{18}$$

where $R_{t-1,k}$ denotes the R&D expenditure in period $t-1$ and geographical distance class k. In a second application they transfer this approach to classes of technological instead of geographical proximity.

Other applications of spatially lagged covariates models for firm location and firm births, respectively, can be found in Alañón Pardo et al. (2007), Arauzo-Carod and Manjón-Antolín (2012), Arzaghi and Henderson (2008), Bonaccorsi et al. (2013), Buczkowska et al. (2014), Martinez Ibañez et al. (2013), Liviano and Arauzo-Carod (2013), and Stuart and Sorenson (2003). Patent data and SLX models are also used by Acosta et al. (2012) and Corsatea and Jayet (2014). Other economic applications include U.S. crime data (Bhati (2005) and Payton et al. (2015)), foreign direct investment (Castellani et al., 2016), terrorist attacks in countries eligible for foreign aid (Savun and Hays, 2011), and traffic accidents (Chiou et al. (2014), and Cai et al. (2016)).[3]

On the one hand, SLX models are very compelling because of the straightforward implementation especially in the context of count data, but on the other hand they only allow for spatial dependence in the covariates, i.e. only local spillovers are obtained (Anselin, 2003, p. 161). Also, they do not consider any spatial structure in the unexplained part of the dependent variable, which might not be plausible in applications, for which not all relevant factors can be observed. The next spatial model class employs the opposite approach and accounts solemnly for spatial correlation in the error terms, i.e. spatial heterogeneity. This solves the limitations just outlined but also means that the spatial structure is a mere nuisance and not of interest by itself.

2.3 Spatial Error Models

Spatial error or spatial heterogeneity models as introduced in Section 1.2 include spatial correlation into the error term of a regression model. Other than in the SLX model, where local spillover effects of a change in X are present, and in the SAR models, where global spillover effects of a change in X are considered, the expectation of y in a SEM model remains unchanged compared to the one in a non-spatial model. Besides the simultaneous autoregressive scheme of the linear SEM described in Section 1.2 a widely used approach

[3]Different approaches, in which not the outcomes of the regressors vary depending on the neighbors and the spatial location but the coefficients, are geographical weighted regressions, applied for example to industrial investments in Indiana by Lambert et al. (2006) and car ownership in Florida by Nowrouzian and Srinivasan (2014), or the smooth transition count model of Brown and Lambert (2014, 2016) applied to location decisions in the U.S. natural gas industry.

in count data modelling is the conditional autoregressive (CAR) scheme introduced by Besag (1974). The standard CAR scheme assumes that the spatial errors in Equation (2) conditional on the neighboring errors are independent and normally distributed i.e.

$$\epsilon_i|\epsilon_{(-i)} \sim N(\rho \sum_{j=1}^{n} w_{ij}\epsilon_j, \sigma_i^2) \tag{19}$$

where $\epsilon_{(-i)}$ denotes the errors of all neighbors of unit i, ρ the spatial correlation parameter of the errors, and σ_i^2 their conditional variance. This leads to the joint distribution (see Besag (1974), for a summary of the derivation see also Cressie and Chan (1989, pp. 396))

$$\epsilon \sim N(0, (I_n - \rho W)^{-1}\Sigma) \tag{20}$$

with $\epsilon = [\epsilon_1, \ldots, \epsilon_n]'$ and $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. This means the error terms follow an auto-Gaussian process. An intrinsic variant (ICAR) has been introduced by Besag and Kooperberg (1995) and an extension to the multivariate case (MCAR) can be found in e.g. Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003). Banerjee et al. (2004) and more recently Czado et al. (2014) give an overview of the different CAR models.

Spatial errors following the CAR scheme are included in count data models which are typically estimated using Bayesian Markov chain Monte Carlo (MCMC) and applied to a wide range of data, e.g. traffic crash data (Aguero-Valverde and Jovanis, 2006; Buddhavarapu et al., 2016; Li et al., 2007; Miaou et al., 2003; Quddus, 2008; Truong et al., 2016), pedestrian casuality counts (Graham et al., 2013; Wang and Kockelman, 2013), crime counts (Jones-Webb et al., 2008; Haining et al., 2009), emergency department visits (Neelon et al., 2013), commuting patterns (Chakraborty et al., 2013), claim numbers on insurances (Czado et al., 2014; Dimakos and Rattalma, 2002; Gschlößl and Czado, 2007, 2008), and firm births (Liviano and Arauzo-Carod, 2014). The CAR approach for modelling spatial heterogeneity is also very popular in biometrics, e.g. for cancer counts (Bernardinelli and Montomoli, 1992; Torabi, 2016; Waller et al., 1997; Xia et al., 1997; Xia and Carlin, 1998; Wakefield, 2007), diabetes mellitus cases (Bernardinelli and Clayton, 1995; Bernardinelli et al., 1997), or Malaria counts (Briet, 2009; Villalta et al., 2012). Various other specifications of spatial error models for count data are applied in the literature as well: LeSage et al. (2007) use a simultaneous autoregressive scheme to model European patent data, Jiang et al. (2013) multiply two different spatial random effects in their Poisson temporal-spatial random effect model for traffic crashes in Florida, and Basile et al. (2013) employ a geoadditive negative binomial model for greenfield investments in the European Union, which includes a bivariate smooth term of latitude and longitude, to name a few.

As mentioned earlier, this way of dealing with spatial association in the data lays emphasis on efficiency but not on explicitly modelling the spatial autocorrelation of the observations. This is the concern of the approaches presented in the next section.

2.4 Spatial Autocorrelation Models

For continuous data an intuitive approach to incorporate a spatial effect into a model is to include the spatially lagged dependent variable, i.e. the weighted observations of the neighbors. While there are plenty of econometric applications for linear spatial models with spatially lagged dependent variables (a review can be found in Anselin (2010), for example), only few authors use spatial models for count data which include a global spatial autocorrelation parameter. One reason for the lack of a widely applied SAR count model is that there is no direct functional relationship between dependent variable y and regressors X in the classical count data models (see for example Equation (16)). A direct transfer of the spatial structure from continuous SAR models is therefore not possible. While the SAR model goes back to Whittle (1954), its adaption to count data modelling took another 20 years until Besag (1974) introduced his auto-Poisson models among others like the auto-Gaussian and auto-binomial models (without giving an example of their estimation).

In the auto-Poisson model the spatially lagged dependent variable is included in the intensity equation of a regression model in which the dependent variable conditional on its neighbors follows a Poisson distribution: $Y(i)|\{Y(j)\}, j \in N(i) \sim Po(\mu(i))$ where $N(i)$ is the set of all neighbors of i and

$$\mu(i) = \exp\left(\alpha(i) + \sum_{j \in N(i)} \beta_{i,j} y(j)\right) \tag{21}$$

which introduces the spatial effect as a weighted sum of neighboring observations with weights $\beta_{i,j}$. Translated to the nowadays common notation, Besag's weights can be divided into a spatial autocorrelation parameter λ and the element of a spatial weights matrix $w_{i,j}$, i.e. $\beta_{i,j} = \lambda w_{i,j}$. The weights satisfy $\beta_{j,i} = 0$ if i and j are not neighbors and $\beta_{i,j} = \beta_{j,i}$, i.e. the relationships are symmetric and no row-standardization of the weight matrix takes place. The remaining, non-spatial regressors are introduced through $\alpha(i)$ (Besag, 1974, p. 202). For estimation Besag (1974) proposes a coding technique for which the set of spatial units is divided into mutually independent subsets. For each subset the model is estimated conditional on the other subsets and the results are combined. In a later article Besag (1975) also proposes a pseudo-likelihood estimation for the auto-models which uses the product of the conditional probability functions instead of a full likelihood function.

Besag's auto-Poisson model suffers from a severe limitation. The inclusion of neighboring observations, whose range is infinite, into the exponential function might cause the process to be explosive if $\beta_{i,j} > 0$. This means that only negative spatial dependence can be modelled. This restriction on the spatial correlation is derived from the necessity that the normalizing constant of the joint probability function derived from the conditional model given above is finite (Besag (1974, p. 202). For a summary of the derivation see also Cressie and Chan (1989, pp. 396)).

Nevertheless, Mears and Bhati (2006) use specification (21) in their negative binomial model of the relationship between homicides and resource deprivation in Chicago. The spatially lagged dependent variable is only considered as a control variable and maximum likelihood estimation is carried out as usual. An auto-model specification is also chosen by Andersson et al. (2009), who estimate, among various spatial and non-spatial specifications, the effect of university decentralization on the number of patents by using a spatial panel Poisson and a spatial panel negative binomial model, respectively, with intensity

$$\mu_{it} = \exp\left(\lambda \sum_{i \neq j} w_{ij} y_{jt} + \beta X_{it} + \sum_{j=1}^{n} \alpha_j \mathbb{I}_j + \sum_{t=1}^{T} \gamma_t \mathbb{I}_t\right) \tag{22}$$

where X_{it} is a set of regressors, $\alpha_j, \quad j = 1, \ldots n$, represent entity fixed effects, $\gamma_t, \quad t = 1, \ldots T$, time fixed effects and $\mathbb{I}$ dummy variables for entity and year. The model is estimated using the not amplified Bayesian methods of "Geobugs". Both papers do not consider any restrictions to ensure the non-positiveness of the spatial autocorrelation parameter.

Several suggestions have been made on how to overcome the shortcomings of the auto-Poisson model, but none of them have found broad, if any application in the empirical analysis of count data: Cressie and Chan (1989) use auto-Gaussian models as an approximation for modelling transformed sudden infant death syndrome (SIDS) counts from North Carolina. Griffith (2006, p. 163) and Kaiser and Cressie (1997, p. 423) point out that the auto-Poisson model can be approximated with an auto-binomial model, which is able to capture positive spatial autocorrelation, by choosing an artificially large n for the binomial distribution. Ferrandiz et al. (1995) model cancer mortality data from Valencia, Spain, by restricting their dependent variable to a finite range so that the auto-Poisson model can also model positive spatial correlation and propose maximum pseudo-likelihood or Monte Carlo scoring for estimation. Kaiser and Cressie (1997) use Winsorization $(Z = Y\ \mathbb{I}(Y \leq R) + R\ \mathbb{I}(Y > R))$ where the largest values are replaced by the truncation value R and therefore the range of the dependent variable is no longer infinite. In their paper, Kaiser and Cressie provide a simulated example with $n = 6$ which they estimate via maximum likelihood. Due to the form of the normalizing constant of the joint winsorized distribution, the maximum likelihood estimation of this model becomes infeasible for large n (Augustin et al., 2006). Augustin et al. (2006) employ a truncated auto-Poisson model as a practical alternative to the winsorized Poisson model to investigate the spatial correlation in leaf and seed counts, respectively. They also run a small simulation study to compare the results from coding, maximum pseudo-likelihood and Monte Carlo maximum likelihood finding that the maximum pseudo-likelihood estimation leads in their setting on average to the smallest bias in parameter estimation but also to asymptotic standard errors that are too small (Augustin et al., 2006, pp. 13).

Analogous to the time series literature for counts, the classification of Cox (1981) can be adopted for spatial autoregressive models as well. It distinguishes between 'parameter-driven' models in which the (spatial) correlation stems from a random process and 'observation-driven' models in which the correlation is driven by actual observations. Therefore, the auto-models and their variants described above all count to the observation-driven models because the included observable spatially lagged dependent variable drives the spatial correlation.

For the sake of completeness, the spatial autocorrelation filtering for count data is mentioned, even though this approach does not fulfill the requirements described in Section 2.1. It has been proposed by Griffith (2002, 2003) as an alternative to the auto-Poisson model. He runs a Poisson regression on eigenvectors of the matrix $(I - \mathbf{11}^T/n)W(I - \mathbf{11}^T/n)$, where I is the identity matrix, $\mathbf{1}$ denotes a vector of ones, and W is a spatial connectivity matrix. Doing this, he obtains data without spatial autocorrelation which can then be analysed with standard models. Empirical examples are given using several plant count data sets. In an empirical comparison of the Winsorized auto-Poisson model and their spatial filtering model using Irish drumlin counts, Griffith (2006) points out the higher flexibility of his spatial modelling structure which allows for several spatial autocorrelation parameters and gives a more detailed picture of the underlying spatial dependence than a model with one spatial parameter. Other applications of spatial filtering can be found in Haining et al. (2009) for offend counts in Sheffield, England, in Chun (2014) for vehicle burglary incidents in Plano, Texas, and in Tevie et al. (2014) for human West Nile virus counts in California and Colorado.

The auto-models and the mentioned variants thereof all try to model spatial dependence by including the spatially lagged dependent variable in the intensity equation of a Poisson regression or other standard count data distributions. This approach bears the problem that a reduced form of that model cannot be obtained. Specifically, it is not possible to use a Leontief inverse $(I - \lambda W)^{-1}$ to obtain a reduced form, like in the linear SAR model (see Section 1.2), which can be estimated by full maximum likelihood. Accordingly, different models have been proposed which promise a more comfortable handling than the previously discussed approaches. Two new count data models which include a spatial autocorrelation parameter have been introduced in recent years, the spatial autoregressive Poisson model (P-SAR) of Lambert et al. (2010) and the spatial autoregressive lagged dependent variable (SAL) Poisson model of Liesenfeld et al. (2016b). By introducing the spatially lagged conditional expectation μ into the intensity equation – instead of the spatially lagged dependent variable – the Leontief inverse can be used to obtain a reduced form. Also, these models do not suffer from the limitation to negative spatial dependence which applies to the auto-Poisson model.

The P-SAR model in its reduced form is given by

$$y|\mu \sim Po(\mu) \tag{23}$$

$$\log \mu = \lambda W \log \mu + X\beta$$

$$\Leftrightarrow \log \mu = (I - \lambda W)^{-1} X\beta \tag{24}$$

where W is a $(n \times n)$ row-standardized spatial weight matrix and λ the spatial autocorrelation parameter. y denotes the observed counts, X is a matrix of exogenous variables, and β denotes the corresponding parameter vector. The reduced form of the P-SAR model makes it obvious that this way of introducing spatial dependence only allows for spatial dependence in the regressors, not in the unexplained part of the observations, since only X enters Equation (24). This is a severe limitation, as it implies that all spatial dependency in the data must be covered by the observed covariates. Obviously, it would be preferable to capture also the unexplained part of spatial correlation in many applications. However, this model does not count to the SLX models in which only local spillover effects (i.e. a change in unit i only affects the proximate neighbors of unit i) are modelled. Here, a change in the regressors of one unit affects all other units via the Leontief inverse which relates all units to each other (Anselin, 2003, p. 156). Therefore, the model entails global spatial effects. For estimation Lambert et al. (2010) suggest a two-step limited information maximum likelihood approach which is described in detail in Section 3.2. A full information maximum likelihood approach is also derived but reported to be numerically infeasible. Although the spatial correlation is introduced by the spatially lagged intensity μ, the reduced form of the P-SAR model clarifies that μ itself is a function of the observed explanatory variables X and does not contain any other random processes. Hence, the model can be classified as observation-driven.

An earlier approach to include spatial correlation by Bhati (2008) also belongs to the class of observation-driven models. He uses the relationship in Equation (24) to obtain a spatial generalized cross-entropy model by replacing the original independent variables in the model with $\tilde{X} = (I - \lambda W)^{-1} X$. By inserting the Leontief inverse into his model, Bhati allows for global spillover effects as it is the case in the P-SAR model. This cross-sectional model has been applied to homicide counts for Chicago.

In a working paper, Hays and Franzese (2009) introduce their observation-driven "S-Poisson" model, which is similar to Lambert's P-SAR model but assumes an additive structure:

$$y \quad = \quad \mu + u, \quad \text{with } \log(\mu) = \lambda W \log(\mu) + X\beta \tag{25}$$

where μ is a vector of the conditional means of $y = [y_1, \ldots, y_n]'$, and the errors $u_i,\ i = 1 \ldots n$ are independently and heteroskedastically distributed. For estimating this model they propose two estimators, a nonlinear least-squares and a generalized method-of-moments estimator, and illustrate this with simulated data.

Two other implementations of an observation-driven spatial count data model have been published: Beger (2012) uses a negative binomial regression model to estimate counts of civilian deaths in the Bosnian war. To account for spatial dependence he includes the spatially lagged dependent variable with an exponentiated coefficient into the intensity equation:

$$\mu_i = (y_{s,i})^{\lambda} \exp(x_i \beta) p_i \tag{26}$$

with $y_{s,i}$ being the average number of counts in the neighbor units of unit i, λ a parameter measuring the strength of the spatial diffusion, and p_i the population of unit i used as an offset variable. By including the parameter of the spatial lag as an exponent the author aims at allowing for positive and negative spatial diffusion while ensuring the positiveness of the intensity at the same time (Beger, 2012, pp. 36). The model is estimated using MCMC methods.

Held et al. (2005) propose to use the sum of the observed counts in neighboring units of unit i ($j \sim i$) in the intensity equation of their space-time model. The intensity of their Poisson or negative binomial model is given by

$$\mu_{it} = \lambda y_{i,t-1} + \phi \sum_{j \sim i} y_{j,t-1} + \eta_{it} \nu_{it} \tag{27}$$

where η_{it} are population counts of unit i and ν_{it} is an exponential function of all remaining regressors, including a trend. They estimate their model using maximum likelihood and apply it to measles case counts for Lower Saxony.

Liesenfeld et al. (2016b) turn away from observation-driven modelling of spatial counts and adopt the parameter-driven models for time series of counts by Zeger (1988) with their SAL-Poisson model. Their resulting spatial parameter-driven model for the i-th observed count is given as

$$y_i|\mu_i \sim Po(\mu_i) \quad \text{with} \quad E[y_i|\mu_i] = \exp(\mu_i) \tag{28}$$

Collecting all the μ_i's in the latent state vector μ, the structure of the model can compactly be written as

$$\mu = \lambda W \mu + X\beta + \epsilon \tag{29}$$

$$\Rightarrow \mu = (I - \lambda W)^{-1} X\beta + (I - \lambda W)^{-1} \epsilon \tag{30}$$

Due to the error term $\epsilon \sim N(0, \sigma^2 I)$ the model allows for spatial dependence in the unexplained part of the variation in the data, too. In that sense it is more flexible and closer to the continuous SAR model specification than the P-SAR model. The SAL-model cannot be estimated via standard maximum likelihood methods as the likelihood contains an n-dimensional integral. Liesenfeld et al. (2016b) propose an efficient importance sampling (EIS) procedure to evaluate the integral and obtain the likelihood function.

A panel data version of the SAL model is proposed in Liesenfeld et al. (2016a) by generalizing the model and the EIS procedure to allow for temporal dependency and unobserved heterogeneity (by including random effects). Equation (29) then becomes:

$$\mu_t = \kappa \mu_{t-1} + \lambda W \mu_t + X_t \beta + \epsilon_t \tag{31}$$

where μ_t denotes the $(n-1) \times 1$ vector of latent state variables in period t and the error term follows a Gaussian random-effect specification:

$$\epsilon_t = \tau + e_t, \quad \text{with } e_t|X_t \sim N_N(0, \sigma_e^2 I_N), \quad \tau|X_t \sim N_N(o, \sigma_\tau^2 I_N) \tag{32}$$

The model is used to estimate and forecast crime counts for the U.S. cities Pittsburgh and Rochester.

Besides the model of Liesenfeld et al. (2016a), two other parameter-driven specifications are available. In the framework of generalized ordered-response probit (GORP) models Castro et al. (2012) implement a Poisson model as a special case. It contains spatial dependence of the underlying latent continuous variable y_{it}^*:

$$y_{it}^* = \delta \sum_{j=1}^{n} w_{ij} y_{jt}^* + \beta_i x_{it} + \epsilon_{it} \tag{33}$$
$$y_{it} = m_{it} \;\text{ if } \psi_{i,m_{it}-1,t} < y_{it}^* < \psi_{i,m_{it},t}$$

The error term ϵ_{it} is supposed to be standard normally distributed and uncorrelated across observation unit i but to have a temporal first-order autoregressive structure. The latent variable y_{it}^* is mapped to the observed counts by the thresholds $\psi_{i,m_{it},t}$ (for details on their form see p. 258). The model is applied to crash frequencies at urban intersections in Arlington, Texas, and is estimated using pairwise composite marginal likelihood.

A variation of the model has been introduced by Bhat et al. (2014), who model the number of new businesses in the counties of Texas for 11 different sectors in a multivariate setting. They allow the error terms ϵ_{is} to be correlated over the sectors $s = 1, \dots, S$. Additionally, they add spatial lags of the K explanatory variables to the model, leading to the following latent process

$$y_{is}^* = \delta_s \sum_{j=1}^{n} w_{ij} y_{js}^* + \beta_s x_i + \sum_{k=1}^{K} \pi_{sk} \sum_{j=1}^{n} w_{ij} x_{jk} + \epsilon_{is} \tag{34}$$

Estimation is again carried out using composite marginal likelihood.

In the framework of generalized linear modelling Melo et al. (2015) introduce a generalized linear space-time autoregressive model with space-time autoregressive disturbances (GLSTARAR) for discrete and binary data. The model is applied to a count data set on armed actions of guerillas in Columbia.

$$\eta_{it} = \log E[y_{it}|x_{it,\epsilon_{it}}] = \beta_0 + x_{it}'\beta_t + \pi_t \sum_{j=1}^{n} w_{ij}^{(1)} \eta_{jt} + \epsilon_{it} \tag{35}$$
$$\epsilon_{it} = \psi_t \sum_{j=1}^{n} w_{ij}^{(2)} \epsilon_{jt} + e_{it}$$

where the coefficients of the explanatory variables β_t as well as the spatial autocorrelation parameter π_t and the spatial autocorrelation parameter of the error term ψ_t are allowed to vary over time. e_{it} is assumed to be i.i.d. normally distributed with zero mean, $E(e_{it}, e_{is}) = \sigma_{ts}\ \forall i, t, s$ and $E(e_{it}, e_{jt}) = 0\ \forall i, j, t$. The number of armed actions y_{it} is supposed to be independently Poisson distributed given the explanatory variables and the unobserved space-time process ϵ_{it}, which is a spatial error term. Additionally, the model can contain a second vector of explanatory variables which are time-invariant. For estimation they propose space-time generalised estimation equations.

At the end of this chapter a class of models is described which has been developed from an entirely different viewpoint. While all previous models try to incorporate the SAR component of continuous models into count models, the following models start from the perspective of the observations-driven integer-valued autoregressive (INAR) model (McKenzie, 1985) and extend its structure to model spatial dependency. Ghodsi et al. (2012) propose a first-order spatial integer-valued autoregressive (SINAR(1,1)) model on a two-dimensional regular lattice. In a regular lattice each observation is characterized by its position on the lattice denoted by i, j and neighbors of unit (i, j) are for example $y_{i,j-1}$, $y_{i+1,j}$ or $y_{i-1,j-1}$, i.e. all eight rectangles around y_{ij} (see Figure 2 for a display of a regular lattice). In the SINAR(1,1) a unilateral spatial structure is assumed, i.e. spatial spillovers are considered to move in one direction across the lattice. The SINAR(1,1) model is given by

$$y_{ij} = \alpha_1 \circ y_{i-1,j} + \alpha_2 \circ y_{i,j-1} + \alpha_3 \circ y_{i-1,j-1} + \epsilon_{i,j} \tag{36}$$

where $\circ$ is the binomial thinning operator with $\alpha_1 \circ y_{i-1,j} = \sum_{k=1}^{y_{i-1,j}} Z_k$ and $Z_k \sim Ber(\alpha_1)$. $\alpha_1, \alpha_2, \alpha_3 \in [0, 1)$ and $\alpha_1 + \alpha_2 + \alpha_3 < 1$ ensure the positivity of the mean of y. $\epsilon_{i,j}$ is a sequence of i.i.d. integer-valued random variables. The model is estimated using Yule-Walker estimators and applied to Student's classic yeast cell count data set. In a later article, a conditional maximum likelihood estimator is proposed for the SINAR(1,1) model (Ghodsi, 2015).

The design of the SINAR(1,1) model stems from a different viewpoint than the previous models and does not fit into the idea of a spatial econometric model with a spatial autocorrelation parameter and explanatory variables. But it accounts very well for the count nature of the data and its application to an economic problem with a spatial process that has one source from which it spreads is not implausible. Brännäs (2013, 2014) propose a more general extension of the INAR model with their simultaneous integer-valued autoregressive model of order one (SINAR(1)) which also includes explanatory variables and models the spatial structure with one or two parameters:

$$y_t = A \circ y_t + B \circ y_{t-1} + \epsilon_t \tag{37}$$

where y_t is a $n \times 1$ vector of counts. The elements of the matrices A and B, α_{ij} and β_{ij}, are parameters which are interpreted as probabilities ($\alpha_{ij} \in [0, 1]$, $\beta_{ij} \in [0, 1]$). Also the elements on the principal diagonal of A (i.e. $\alpha_{ii} \forall i$) are equal to zero. The elements in A and B can contain covariates, e.g. in a logistic form (Brännäs, 1995): $a_{ij,t} = 1/(1 + \exp(x_{ij,t}\theta))$. Similarly, they can contain the spatial distance of units in the form $a_{ij,t} = 1/(1 + \exp(\alpha_1 w_{ij}))$, $i \neq j$

(Brännäs, 2013, p. 8) or $a_{ij,t} = 1/(1 + \exp(\alpha_0 + \alpha_1 w_{ij}))$, $i \neq j$ (Brännäs, 2014, p. 6) where w_{ij} is the respective element of a spatial inverse distance matrix W. The inclusion of a spatial distance measure in this way reduces the number of unknown parameters from n^2 in A to one or two (α_0 and α_1), respectively (Brännäs, 2013, p. 6). The authors do not give an empirical application but make some comments on IV and GMM estimation.

3 Investigation and Extension of the Poisson SAR Model

3.1 Introduction

After summarizing the literature on spatial count data modelling, the remainder of this thesis is concerned with count data models incorporating a spatial autoregressive component. The starting point is the exploration of the spatial autoregressive Poisson model (P-SAR) of Lambert et al. (2010). This model seems to be the most promising observation-driven attempt so far, with respect to a model which is straightforwardly applicable for empirical economists. First, the range of the spatial autocorrelation parameter is not restricted to negative values as it is the case in the auto-models. Second, its estimation does not require any computationally extensive methods like other proposed models. But, Lambert et al.'s article leaves open the question why or if at all full information maximum likelihood estimation (FIML) is not applicable and the proposed limited information maximum likelihood (LIML) estimator is the better choice. The authors claim that in repeated Monte Carlo trials "[t]he usual optimization algorithms were too frequently unsuccessful [...]" (Lambert et al., 2010, p. 244). The FIML and LIML estimation results for the P-SAR model will be compared in a Monte Carlo study (Section 3.4) to verify this statement. Additionally, the effect of ignored spatial correlation or dispersion in the P-SAR model is investigated in the study. In a second step the model and some extensions, which are introduced in Section 3.3, are used to estimate spillover effects in the counts of start-up firm births in the manufacturing sector of the United States (Section 3.6). For evaluation of the empirical results, scoring rules are employed, which are discussed in Section 3.5. Before starting with the Monte Carlo study, a closer description of the model and the LIML estimation procedure takes place in the following section.

3.2 The Poisson Autoregressive Model and Limited Information Likelihood Estimation

The P-SAR model of Lambert et al. (2010) is given in Equations (23) and (24). This model translates the spatial autoregressive (SAR) model for continuous data to counts by including the spatially lagged logarithm of the conditional expectation μ into the intensity equation. But the reduced form in Equation (24) highlights that, unlike in the continuous SAR model, this way of introducing spatial dependence only allows for spatial dependence in the regressors, not in the unexplained part of the observations. In this model the spatial correlation parameter λ measures the spatial correlation between the conditional expectations of the dependent variable in a spatial unit and its neighbors.

Due to the spatial component as well as the nonlinearity in parameters, the parameter estimates cannot be interpreted directly. This makes the calculation of marginal effects of a change in a regressor necessary (see Section 1.2 and LeSage and Pace (2009)). For the P-SAR model the direct marginal effects, which are comparable to the marginal effects in a non-spatial model, are obtained by

$$\frac{\partial \mu_{y_i|X}}{\partial x_{ik}} = a_{ii} \exp(A_{i.} X\beta)\beta_k \tag{38}$$

where A denotes the Leontief inverse, i.e. $A = (I - \lambda W)^{-1}$, and a_{ii} the respective element of this matrix. They give the marginal change in conditional expectation μ_i if regressor x_{ik} changes.
Because of the spatial dependence, the dependent variable also changes if a neighbor's regressor changes. This is denoted as indirect effects whose sum over all neighbors equals

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\partial \mu_{y_i|X}}{\partial x_{jk}} = \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij} \exp(A_{i.} X\beta)\beta_k \tag{39}$$

Accordingly, the total marginal effect, i.e. the sum of the effects of a change in each element in vector $X_{.k}$ is

$$\sum_{j=1}^{n} \frac{\partial \mu_{y_i|X}}{\partial x_{jk}} = \sum_{j=1}^{n} a_{ij} \exp(A_{i.} X\beta)\beta_k \tag{40}$$

where $A_{i.}$ denotes the i^{th} row of the Leontief inverse.

The non-spatial Poisson model is typically estimated using (quasi) maximum likelihood estimation. This estimator also seems to be appropriate for the P-SAR model at first sight. Lambert et al. (2010, p. 244), however, claim that full information maximum likelihood (FIML) estimation is numerically infeasible for estimating the P-SAR model and suggest a limited information maximum likelihood (LIML) approach instead. This estimation procedure consists of two steps. First, the logarithmized counts (zeros are transformed beforehand either by an inverse hyperbolic sine transformation or by adding 0.5) are regressed on spatial instruments via ordinary least squares, which yields:

$$\hat{\delta} = (Q'Q)^{-1}Q'Wg(y_i^*)$$

with the instrumental variables $Q = [X, W\bar{X}, WW\bar{X}]$, $\bar{X}$ containing all regressors of the original problem excluding the constant, and $g(y_i^*)$ being the logarithmized counts with transformed zeros.
The predicted values from step one, $\widehat{g(y_i^*)} = Q\hat{\delta}$, replace y_i in the Poisson probability density function

$$f(y_i|X_{i.}, W, Q_i\hat{\delta}; \beta\lambda) = \frac{\exp(X_{i.}\beta + \lambda Q_i'\hat{\delta})^{y_i} \exp(-\exp(X_{i.}\beta + \lambda Q_i'\hat{\delta}))}{y_i!}$$

and corresponding likelihood function

$$\log L = \sum_{i=1}^{N} y_i(X_{i.}\beta + \lambda Q_i'\hat{\delta})) - \exp(X_{i.}\beta + \lambda Q_i'\hat{\delta})) - \log y_i! \tag{41}$$

The likelihood function is maximized using standard methods, e.g. Newton-Raphson, to obtain the parameter estimates. Lambert et al. (2010) also give formulas to obtain standard errors for estimators of both steps (see pp. 244).

Aside from the statements in the article, there seems to be no obvious reason why full information maximum likelihood estimation should not work for this model. Therefore, Section 3.4 contains a Monte Carlo study which compares the performance of the LIML and FIML estimators. Before turning to the Monte Carlo Study Section 3.3 introduces several extensions of the Poisson SAR model.

3.3 Extensions of the Poisson SAR Model

The spatial structure of the P-SAR model of Lambert et al. (2010) can straightforwardly be transferred to other count data models because it only alters the intensity equation of the models. As a first, obvious extension, a negative binomial spatial autoregressive model (NB-SAR) is introduced. The Poisson model assumes the data to be equidispersed, i.e. to have equal (conditional) mean and variance. The negative binomial (NB) model removes this constraint and allows for overdispersion, which leads to more efficient estimation of overdispersed data. Nevertheless, the variance is still a function of the mean in negative binomial models (with several proposed forms). Here, the negative binomial 2 specification is chosen, whose variance function is $Var[y_i|X_{i.}] = \mu_i + \alpha\mu_i^2$ with α being an additional dispersion parameter. The density of the NB model is then

$$f(y_i|\mu_i,\alpha) = \frac{\Gamma(\alpha^{-1}+y_i)}{\Gamma(\alpha^{-1})\Gamma(y_i+1)}\left(\frac{\alpha^{-1}}{\mu_i+\alpha^{-1}}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\mu_i+\alpha^{-1}}\right)^{y_i} \tag{42}$$

For obtaining the NB-SAR model, the intensity parameter μ is modelled as in Equation (24) which equals, like in the P-SAR model, the conditional expectation of y. Estimation is usually carried out using maximum likelihood estimation with log likelihood function

$$\begin{aligned}\log L \quad = \quad & \sum_{i=1}^{n}\left(\sum_{j=0}^{y_i-1}\log(\alpha^{-1}+j)\right) - \log(y_i!) \\ & -(y_i+\alpha^{-1})\log(\,1+\alpha\mu_i) + y_i\log(\alpha) + y_i\log(\mu_i)\end{aligned} \tag{43}$$

Many other extensions of the Poisson model have been introduced in the literature, the most prevalent ones being zero-inflated Poisson (ZIP) and Hurdle Poisson (HP) models. Both have been developed to take special care of excess zeros in the data, but by doing so they also introduce a greater variability into the model, which in any case allows a better fit to the data structure.

The zero-inflated Poisson models are two-part models which have an inflation process additionally to the Poisson count process. Both processes generate zeros as outcomes, i.e. the inflation process adds additional probability mass to the outcome zero as compared to the standard Poisson model. This means in turn that the other process can assign less probability mass to zeros. Therefore, the probability mass can be shifted to the positive outcomes making the zero-inflated Poisson model more flexible as compared to the standard Poisson model. Especially, it has an improved ability to account for very large outcomes when at the same time many zeros are present. The inflation process usually follows a binary distribution, a typical choice is logit. This is combined with a spatially lagged intensity in the Poisson process to obtain the ZIP-SAR regression model

$$\begin{aligned} f(y_i|X,Z) &= \frac{\exp(Z_{i.}\gamma)}{1+\exp(Z_{i.}\gamma)}\,(1-min\{y_i,1\}) \\ &\quad +\left(1-\frac{\exp(Z_{i.}\gamma)}{1+\exp(Z_{i.}\gamma)}\right)\frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!} \\ \text{with} &\qquad \log\mu = (I-\lambda W)^{-1}X\beta \end{aligned} \tag{44}$$

Z are the regressors of the inflation process and γ their parameters.
The conditional expectation of y_i in the ZIP-SAR model with logit inflation process is

$$E[y_i|X,Z] = \left(1-\frac{\exp(Z_{i.}\gamma)}{1+\exp(Z_{i.}\gamma)}\right)\exp(\mu_i) \tag{45}$$

Zero-inflated models are usually estimated via maximum likelihood. The according log likelihood function of the ZIP-SAR model with logit inflation process is given by

$$\begin{aligned} \ell(\gamma,\beta) &= \sum_{y_i=0}\log(\exp(Z_{i.}\gamma)+\exp(-\mu_i)) + \sum_{y>0} y_i\log(\mu_i)-\mu_i-\log(y_i!) \\ &\quad -\sum_{i=1}^{n}\log(1+\exp(Z_{i.}\gamma)) \end{aligned} \tag{46}$$

Hurdle models are motivated by the idea that a certain threshold has to be crossed before positive outcomes can be observed. The two-part model consists of two completely separable processes, a dichotomous one which generates the zeros by determining whether the hurdle is crossed or not (i.e. the hurdle process) and a second one which gives the probabilities of positive outcomes (i.e. the parent process). The latter one usually follows a truncated count data distribution. The hurdle Poisson model employed here has a Poisson model right censored at one as the hurdle part and a truncated Poisson as the parent process. If the probability of not crossing the hurdle is larger than the probability for a zero in the parent process (i.e. the not truncated Poisson distribution), then the parent process alone has a larger mean than the overall model. This means that the range will be larger than in a classical Poisson model and the hurdle model can better adapt to very large observations if many zeros are present at the same time.

This hurdle Poisson SAR (HP-SAR) model consists of two Poisson processes, the hurdle Poisson is modelled without spatial dependency, the Poisson parent process incorporates the spatial autoregressive structure of the intensity, which leads to following regression model:

$$f(y_i|X,Z) = [\exp(-\exp(Z_{i.}\gamma))]^{(1-min\{y_i,1\})} \times \times \left[\left(\frac{1-\exp(-\exp(Z_{i.}\gamma))}{1-\exp(-\mu_i)}\right)\frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}\right]^{min\{y_i,1\}} \quad (47)$$

$$\text{with} \quad \log\mu = (I-\lambda W)^{-1}X\beta \quad (48)$$

The conditional expectation of y_i in this model is given by

$$E[y_i|X,Z] = \frac{1-\exp(-\exp(Z_{i.}\gamma))}{1-\exp(-\mu_i)}\mu_i \quad (49)$$

Estimation can be carried out with maximum likelihood estimation. The likelihood function of the Poisson-Poisson hurdle model is

$$\ell(\gamma,\beta) = \sum_{i=0}^{n} -\exp(Z_{i.}\gamma)\,(1-min\{y_i,1\}) +[\log(1-\exp(-\exp(Z_{i.}\gamma))) - \log(1-\exp(-\mu_i))]\,min\{y_i,1\} +[y_i\log(\mu_i) - \mu_i - \log(y_i!)]\,min\{y_i,1\} \quad (50)$$

Z denote the regressors of the hurdle process with parameters γ. The likelihood function can also be separated into two independent parts, one being a function of γ and one a function of β, which can be optimized separately.
Figure 3 gives an overview of all proposed models and their connections. With exception of the zero-inflated Poisson model, all models nest at least the non-spatial Poisson.[4]

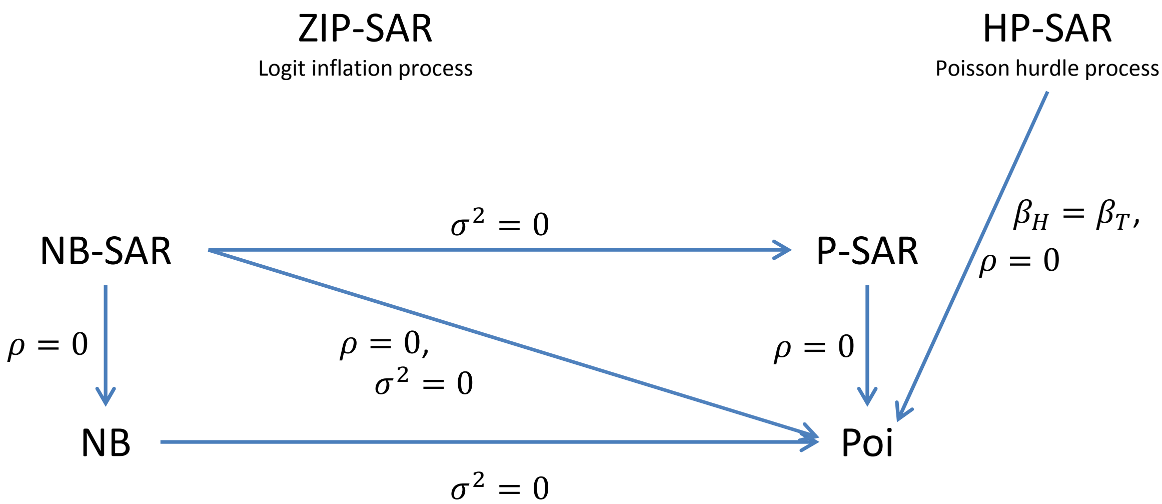


Figure 3: Extensions of the P-SAR model. β_H are the parameters of the hurdle process, β_P the parameters of the parent process. Models to which an arrow points are nested.

[4]For more details on the underlying non-spatial models see e.g. Cameron and Trivedi (2013), Hilbe (2011) or Winkelmann (2010).

3.4 Monte Carlo Study

Before applying the various specifications introduced in the last section to an empirical example, a Monte Carlo study is conducted. Motivated by the argument of Lambert et al. (2010) that the Poisson SAR model could not be estimated consistently via full maximum likelihood, the main goal of the study is to investigate and compare the asymptotic properties of the two maximum likelihood estimators for the P-SAR model. Additionally to that, the NB-SAR model is considered for both data generating process (DGP) and regression model to analyze the influence of overdispersion on the spatial estimation. To my knowledge, there are no studies which analyse the consequences of neglecting the spatial autocorrelation of count data. Therefore, this study is extended to address the question whether the P-SAR and NB-SAR model estimators are able to detect non-spatial data and whether there are consequences if the spatial structure in the data is ignored, i.e. if spatial data is estimated using a non-spatial Poisson or negative binomial model.

3.4.1 Data Generating Processes and Study Setup

Data are generated from the two most popular count data processes, namely the Poisson process and the negative binomial 2 process, both extended with the spatial structure proposed by Lambert et al. (2010). In addition, data from a Poisson distribution without any spatial structure, i.e. non-spatial data, are generated for comparison. From each data generating process 20 different combinations of the spatial autoregressive parameter $\lambda = \{0, 0.2, 0.4, 0.6, 0.8\}$ and sample size $n = \{1000, 5000, 25000, 50000\}$ with each 100 samples (repetitions) are produced. The study is restricted to the case of positive spatial autocorrelation since real-world examples of negative spatial autocorrelation are hard to find in econometrics. The spatial weight matrix used is a contiguity matrix generated by using the function `xy2cont` of the Spatial Statistics Toolbox for MATLAB 2.0 (Pace, 2003) on randomly generated coordinates. The regressors include a constant (X_0), a uniformly distributed variable $X_1 \sim U(0, 2)$ and a normally distributed variable $X_2 \sim N(1, 2)$. The corresponding parameters are $\beta = [\beta_0, \beta_1, \beta_2] = [0.1, 0.1, 0.1]$.

The P-SAR data originates from the data generating process (23) and (24). For the NB-SAR data the overdispersion parameter α takes the values 1/8 and 1/2, respectively. The data is generated from:

$$
\begin{aligned}
y|\mu &\sim NB(\mu, \alpha) && (51)\\
\log \mu &= (I - \lambda W)^{-1} X\beta && (52)\\
\alpha &= \{1/8, 1/2\} && (53)
\end{aligned}
$$

Finally, the non-spatial data is produced using a standard Poisson model (Equation (16)).

In this Monte Carlo study, the P-SAR data and the NB-SAR data are estimated using the P-SAR model (23)-(24) as well as the NB-SAR model (42). Additionally, the spatial data sets are estimated using the non-spatial Poisson (16). All spatial models are estimated with LIML and FIML (i.e. quasi maximum likelihood). The non-spatial model is only estimated

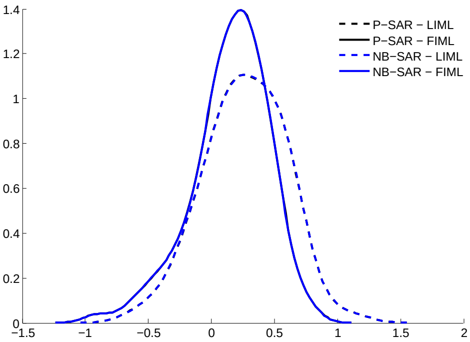
		Model and Estimation Method				
		P-SAR LIML	P-SAR FIML	NB-SAR LIML	NB-SAR FIML	Poisson QML
	P-SAR	x	x	x	x	x
Data	NB-SAR ($\alpha = 1/8$)	x	x	x	x	x
	NB-SAR ($\alpha = 1/2$)	x	x	x	x	x

Table 1: SAR models Monte Carlo setup.

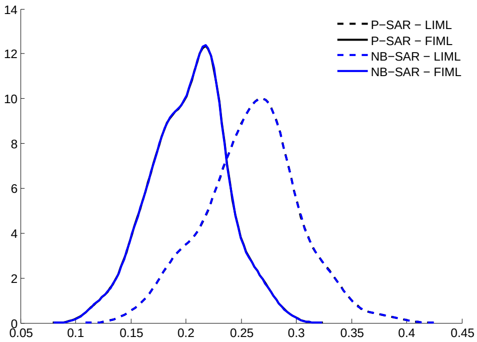
with the quasi maximum likelihood procedure commonly used for this model. Table 1 gives an overview of the estimated combinations of data and models.

3.4.2 Monte Carlo Parameter Estimates

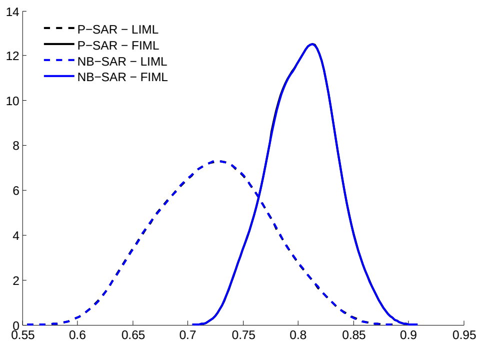
In the following, the estimation results from 100 simulated samples each for $n = \{1000, 50000\}$ and $\lambda = \{0.2, 0.8\}$ for each model and data combination are discussed. Bias, relative bias and root mean squared error (RMSE) of all remaining specifications are reported in Appendix A.1.



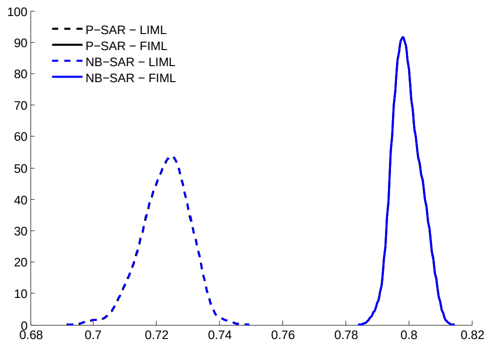
(a) $\lambda = 0.2$, $n = 1000$



(b) $\lambda = 0.2$, $n = 50000$



(c) $\lambda = 0.8$, $n = 1000$



(d) $\lambda = 0.8$, $n = 50000$

Figure 4: Monte Carlo results of probability density estimates for $\hat{\lambda}$ of generated P-SAR data. Dashed and solid lines, respectively, are overlapping. Please note the different scaling on the x-axis.

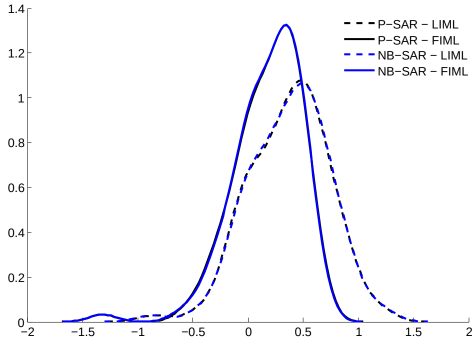
	P-SAR LIML		**P-SAR FIML**		**NB-SAR LIML**		**NB-SAR FIML**		**Non-spatial Poisson**	
	RMSE	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**
					$\lambda = 0.2$, $n = 1000$					
$\hat{\lambda}$	0.325	0.067	0.306	−0.051	0.220	0.038	0.262	−0.051		
$\hat{\beta}_0$	0.089	0.018	0.126	0.028	0.055	0.005	0.114	0.026	0.106	0.084
$\hat{\beta}_1$	0.052	−0.003	0.051	−0.004	0.031	0.002	0.045	−0.006	0.051	−0.002
$\hat{\beta}_2$	0.013	−0.002	0.013	−0.002	0.010	0.000	0.011	−0.002	0.012	0.000
$\hat{\alpha}$					0.009	0.003	0.013	0.006		
					$\lambda = 0.2$, $n = 50000$					
$\hat{\lambda}$	0.073	0.059	0.033	0.003	0.073	0.059	0.033	0.003		
$\hat{\beta}_0$	0.017	0.012	0.014	−0.002	0.017	0.012	0.014	−0.002	0.073	0.073
$\hat{\beta}_1$	0.007	0.001	0.007	0.001	0.007	0.001	0.007	0.001	0.007	0.002
$\hat{\beta}_2$	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.001
$\hat{\alpha}$					0.003	0.002	0.003	0.002		
					$\lambda = 0.8$, $n = 1000$					
$\hat{\lambda}$	0.090	−0.077	0.029	0.004	0.073	−0.049	0.025	0.000		
$\hat{\beta}_0$	0.224	0.214	0.038	−0.004	0.182	0.144	0.034	0.001	1.154	1.153
$\hat{\beta}_1$	0.026	−0.003	0.017	−0.001	0.022	−0.001	0.015	0.000	0.036	0.024
$\hat{\beta}_2$	0.009	−0.001	0.008	−0.001	0.007	−0.001	0.007	−0.002	0.019	0.018
$\hat{\alpha}$					0.005	0.002	0.003	0.001		
					$\lambda = 0.8$, $n = 50000$					
$\hat{\lambda}$	0.078	−0.077	0.004	−0.001	0.078	−0.077	0.004	−0.001		
$\hat{\beta}_0$	0.212	0.212	0.005	0.001	0.212	0.212	0.005	0.001	1.169	1.169
$\hat{\beta}_1$	0.004	0.000	0.003	0.000	0.004	0.000	0.003	0.000	0.022	0.021
$\hat{\beta}_2$	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.021	0.021
$\hat{\alpha}$					0.002	0.001	0.001	0.001		

Table 2: Monte Carlo results of parameter estimates for generated P-SAR data with $\lambda = \{0.2, 0.8\}$ and $n = \{1000, 50000\}$. Values reported as 0.000 are smaller than 0.0005.

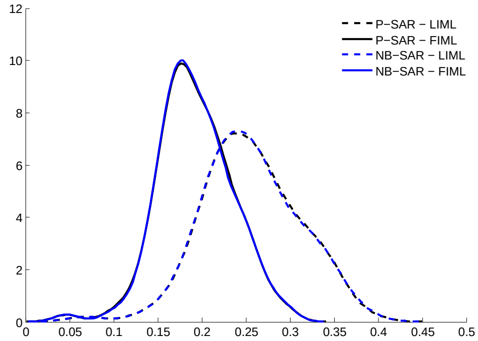
Figure 4 displays the sampling distribution plots of $\hat{\lambda}$ for the P-SAR data set and FIML and LIML estimation.[5] The corresponding values for bias and RMSE are summarized in Table 2. The dashed lines in Figure 4 correspond to the LIML estimates, the solid lines to the FIML ones. The distribution plots of the P-SAR model estimation and the NB-SAR model estimation are practically identical for both estimation methods. This has been expected since the NB-SAR model nests the P-SAR model and, furthermore, Gourieroux et al. (1984) showed for the non-spatial case that a correctly specified conditional mean is sufficient for the consistency of the quasi maximum likelihood estimator. These distribution plots support this result also for the spatial model discussed here. But the results differ considerably between LIML and FIML estimation. Whereas both methods produce more precise estimates, i.e. the probability mass of the sampling distribution is stronger concentrated if the true spatial autocorrelation is larger, the behavior for increasing n differs between FIML and LIML.[6] For the smallest sample size ($n = 1000$) the bias of λ is of the same magnitude for LIML and FIML estimates (though in opposite directions). The bias of FIML decreases with increasing sample size, indicating the consistency of the estimator, whereas the bias of the

[5] Calculated using the MATLAB function `ksdensity`.

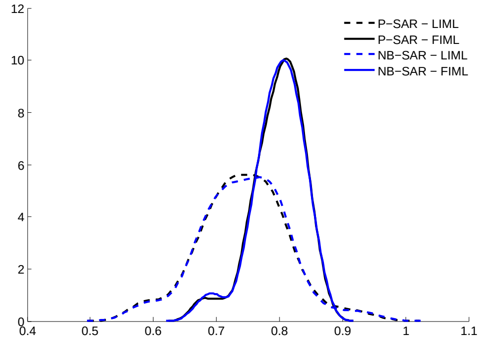
[6] Please note that the scaling of the x-axis differs between the graphs to allow for a better visualization.



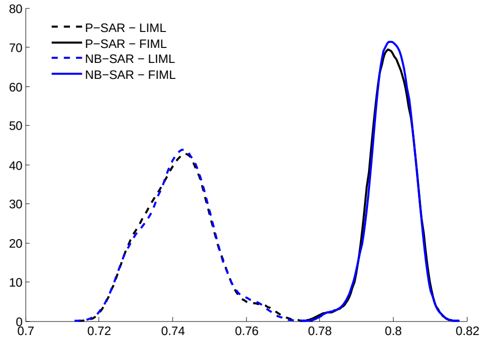
(a) $\lambda = 0.2$, $n = 1000$



(b) $\lambda = 0.2$, $n = 50000$



(c) $\lambda = 0.8$, $n = 1000$

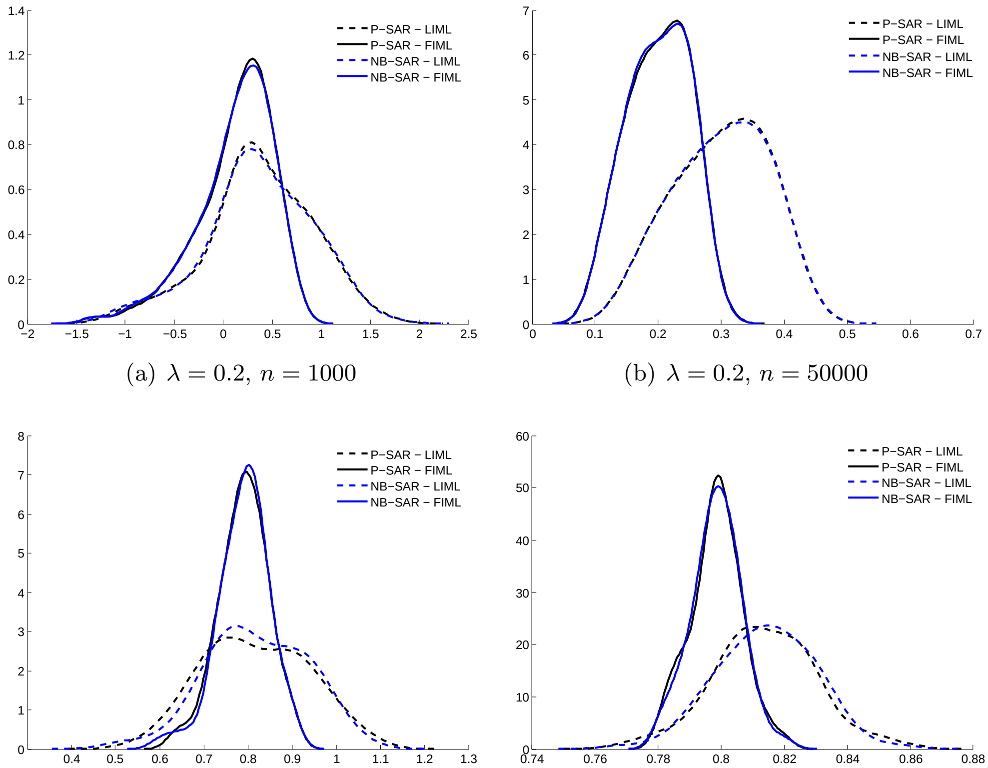


(d) $\lambda = 0.8$, $n = 50000$

Figure 5: Monte Carlo results of probability density estimates for $\hat{\lambda}$ of generated NB-SAR data, $\alpha = 1/8$. Please note the different scaling on the x-axis.

LIML estimation remains at the same level. Also, the distributions of the FIML estimates are more peaked than the ones of the LIML, especially for large λ. The smaller variation in the estimates, visualized by the stronger concentrated sampling distribution plots and documented in the smaller RMSE, shows that the model is estimated more precisely using FIML, indicating the higher efficiency of the FIML estimates compared to the LIML estimates. Generally, this would not be surprising since using the full information available should always lead to results which are at least as good (and often better) as results from another approach using less information. For this model, Lambert et al. (2010) argue that FIML estimation of the P-SAR model is infeasible. With regard to the estimation of the spatial autocorrelation parameter λ, the results just presented contradict this statement and show the clear superiority of the FIML estimation.

The difference between LIML and FIML are less pronounced for the estimates of the parameter vector β, whose bias and RMSE are also displayed in Table 2. The largest bias and RMSE is obtained for the constant β_0. The coefficients of the regressors X_1 and X_2 are generally estimated more precisely, i.e. with much smaller bias and RMSE. Differences between FIML and LIML are observed for data with large spatial correlation, for which the bias and RMSE of the LIML estimations do not decrease, and in some cases even increase, with increasing sample size. The bias and RMSE of the FIML estimation stay at least constant with increasing sample size, in most cases they decrease. Again, FIML clearly outperforms LIML in terms of accuracy of the estimation.



(a) $\lambda = 0.2$, $n = 1000$ (b) $\lambda = 0.2$, $n = 50000$

(c) $\lambda = 0.8$, $n = 1000$ (d) $\lambda = 0.8$, $n = 50000$

Figure 6: Monte Carlo results of probability density estimates for $\hat{\lambda}$ of generated NB-SAR data, $\alpha = 1/2$. Please note the different scaling on the x-axis.

The two columns on the right hand side of Table 2 give the results for the non-spatial Poisson estimation. The bias of the β's (except the constant β_0) increases with increasing spatial correlation in the data. The RMSE is of the same magnitude for the small sample size but decreases less with increasing n for data with a high value of λ. The constant β_0 shows a more severe bias than the other ones. This indicates that the missing spatial component is mostly absorbed by this estimate and only slightly influences the estimates of the other coefficients. Having discussed the simulation results for the parameter vector β, it must be noted that the parameter estimates themselves are only of limited use for interpreting the influence of explanatory variables on the dependent variable. To evaluate the influence of a regressor on the outcomes in a count data model, which models only the dependence of the conditional expectation of y on the explanatory variables X, it is necessary to calculate the marginal effects, which are nonlinear functions of the parameters. This also means that the ability of an estimation procedure to correctly estimate the effects of explanatory variables on the dependent variable should be evaluated through the estimated marginal effects. Because of this, the simulation results of the marginal effects are presented in the next section to gain a better understanding of the implications of an ignored spatial dependence in the data.

Before getting to the marginal effects, Figures 5 and 6 and Tables 3 and 4 display the results for the NB-SAR data estimations with $\alpha = 1/8$ and $\alpha = 1/2$, respectively. Again, the P-SAR and the NB-SAR model estimations lead to almost identical distribution plots

	P-SAR LIML		**P-SAR FIML**		**NB-SAR LIML**		**NB-SAR FIML**		**Non-spatial Poisson**	
	RMSE	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**
					$\lambda = 0.2,\ n = 1000$					
$\hat{\lambda}$	0.394	0.171	0.309	−0.023	0.395	0.174	0.309	−0.022		
$\hat{\beta}_0$	0.096	0.000	0.137	0.017	0.096	−0.001	0.137	0.017	0.103	0.082
$\hat{\beta}_1$	0.048	−0.005	0.049	−0.005	0.047	−0.005	0.049	−0.005	0.049	−0.004
$\hat{\beta}_2$	0.014	−0.002	0.014	−0.002	0.014	−0.002	0.014	−0.002	0.014	0.000
$\hat{\alpha}$					0.035	−0.003	0.035	−0.002		
					$\lambda = 0.2,\ n = 50000$					
$\hat{\lambda}$	0.078	0.056	0.042	−0.007	0.078	0.056	0.042	−0.007		
$\hat{\beta}_0$	0.024	0.019	0.017	0.004	0.024	0.019	0.017	0.004	0.075	0.075
$\hat{\beta}_1$	0.007	0.000	0.007	0.000	0.007	0.000	0.007	0.000	0.007	0.001
$\hat{\beta}_2$	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000
$\hat{\alpha}$					0.005	0.000	0.005	0.000		
					$\lambda = 0.8,\ n = 1000$					
$\hat{\lambda}$	0.089	−0.057	0.041	−0.002	0.088	−0.056	0.041	−0.003		
$\hat{\beta}_0$	0.243	0.229	0.054	0.003	0.241	0.228	0.053	0.003	1.147	1.146
$\hat{\beta}_1$	0.033	0.003	0.023	−0.001	0.033	0.003	0.022	−0.001	0.042	0.028
$\hat{\beta}_2$	0.011	0.001	0.009	0.001	0.011	0.001	0.009	0.001	0.021	0.019
$\hat{\alpha}$					0.014	0.000	0.014	−0.001		
					$\lambda = 0.8,\ n = 50000$					
$\hat{\lambda}$	0.059	−0.059	0.005	−0.001	0.059	−0.059	0.005	−0.001		
$\hat{\beta}_0$	0.235	0.235	0.007	0.001	0.235	0.235	0.007	0.001	1.169	1.169
$\hat{\beta}_1$	0.004	0.000	0.003	0.000	0.004	0.000	0.003	0.000	0.021	0.021
$\hat{\beta}_2$	0.002	0.000	0.001	0.000	0.002	0.000	0.001	0.000	0.021	0.021
$\hat{\alpha}$				0.000	0.003	0.001	0.002	0.000		

Table 3: Monte Carlo results of parameter estimates for generated NB-SAR data with $\alpha = 1/8$, $\lambda = \{0.2, 0.8\}$ and $n = \{1000, 50000\}$. Values reported as 0.000 are smaller than 0.0005.

for $\hat{\lambda}$, although here small differences are visible. The sampling distributions of the FIML estimates are more peaked than the ones of the LIML, especially for large λ. There seems to be no systematic differences in RMSE and bias between data with stronger and weaker overdispersion. In both cases, LIML leads to stronger biased estimates for λ, especially for the smaller sample sizes with relatively weak spatial dependence in the data. Generally, the biases become smaller with increasing sample size, but again, as in the case of the P-SAR data, the bias of the FIML estimations decreases stronger whereas the bias of the LIML estimations stays of the same magnitude. In summary, the results confirm the ones of the P-SAR model regarding the predominance of the FIML estimation. The presence of overdispersion in different strength does not influence this. For the estimates of the non-spatial Poisson estimations we can again note that the constant β_0 is much stronger biased than the other parameter estimates, which seem not to be much influenced by the ignored spatial correlation in the data. As noted above, the estimates for marginal effects should be calculated in order to investigate the ability of the estimations to correctly capture the effects of the regressors on the dependent variable.

	P-SAR LIML		**P-SAR FIML**		**NB-SAR LIML**		**NB-SAR FIML**		**Non-spatial Poisson**	
	RMSE	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**
					$\lambda = 0.2$, $n = 1000$					
$\hat{\lambda}$	0.562	0.175	0.396	0.002	0.568	0.167	0.402	−0.085		
$\hat{\beta}_0$	0.109	0.018	0.163	−0.006	0.111	0.020	0.165	0.043	0.109	0.081
$\hat{\beta}_1$	0.064	−0.003	0.063	0.007	0.065	−0.004	0.064	−0.005	0.065	−0.002
$\hat{\beta}_2$	0.019	−0.007	0.019	−0.004	0.019	−0.007	0.019	−0.007	0.018	−0.004
$\hat{\alpha}$					0.058	−0.021	0.058	−0.020		
					$\lambda = 0.2$, $n = 50000$					
$\hat{\lambda}$	0.123	0.100	0.047	0.001	0.123	0.100	0.047	0.000		
$\hat{\beta}_0$	0.027	0.023	0.019	−0.001	0.027	0.023	0.018	0.000	0.074	0.073
$\hat{\beta}_1$	0.009	0.001	0.009	0.001	0.009	0.001	0.009	0.001	0.009	0.002
$\hat{\beta}_2$	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.001
$\hat{\alpha}$					0.008	0.001	0.008	0.001		
					$\lambda = 0.8$, $n = 1000$					
$\hat{\lambda}$	0.116	0.018	0.056	−0.005	0.113	0.020	0.058	−0.009		
$\hat{\beta}_0$	0.295	0.268	0.070	0.008	0.288	0.264	0.070	0.014	1.155	1.153
$\hat{\beta}_1$	0.053	0.001	0.039	−0.001	0.052	0.003	0.038	−0.001	0.058	0.025
$\hat{\beta}_2$	0.018	−0.002	0.015	0.000	0.018	−0.002	0.015	0.000	0.023	0.017
$\hat{\alpha}$					0.034	−0.003	0.034	−0.004		
					$\lambda = 0.8$, $n = 50000$					
$\hat{\lambda}$	0.021	0.014	0.008	−0.002	0.021	0.014	0.008	−0.002		
$\hat{\beta}_0$	0.276	0.276	0.011	0.003	0.275	0.275	0.011	0.002	1.170	1.170
$\hat{\beta}_1$	0.008	−0.001	0.005	0.000	0.008	−0.001	0.005	0.000	0.021	0.020
$\hat{\beta}_2$	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000	0.021	0.021
$\hat{\alpha}$					0.005	0.002	0.005	0.000		

Table 4: Monte Carlo results of parameter estimates for generated NB-SAR data with $\alpha = 1/2$, $\lambda = \{0.2, 0.8\}$ and $n = \{1000, 50000\}$. Values reported as 0.000 are smaller than 0.0005.

3.4.3 Monte Carlo Estimates of Marginal Effects

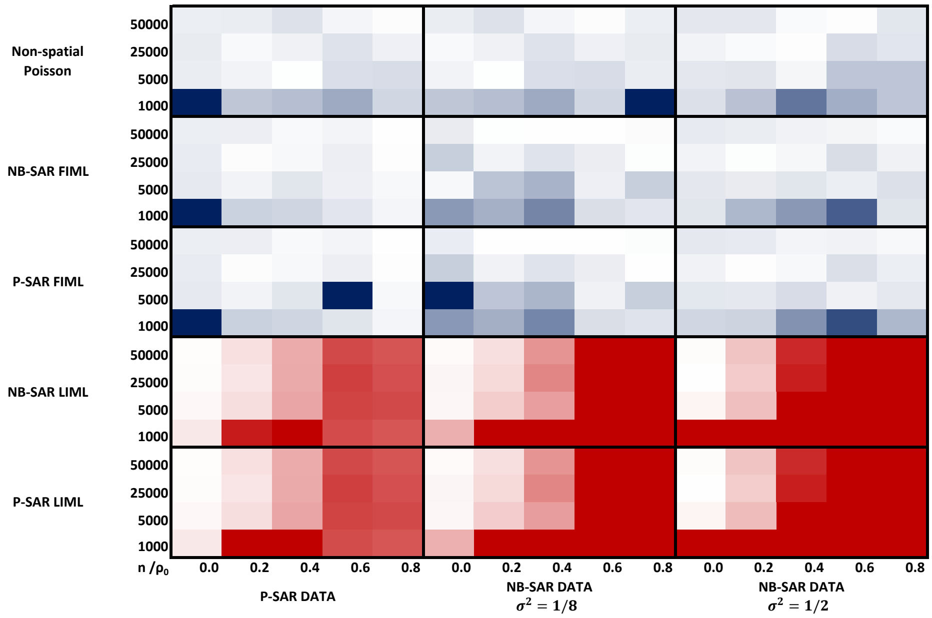
Figures 7 and 8 and Tables 5 to 7 show RMSE, bias and relative bias of the median estimated direct and total marginal effects for X_1 and X_2 in the P-SAR and NB-SAR simulations. They are calculated using Equations (38) and (40) for the P-SAR and NB-SAR model, respectively. For the standard Poisson model the marginal effects are:

$$\frac{\partial \mu_{y_i|X}}{\partial x_{ik}} = \exp(X_{i.}\beta)\beta_k \tag{54}$$

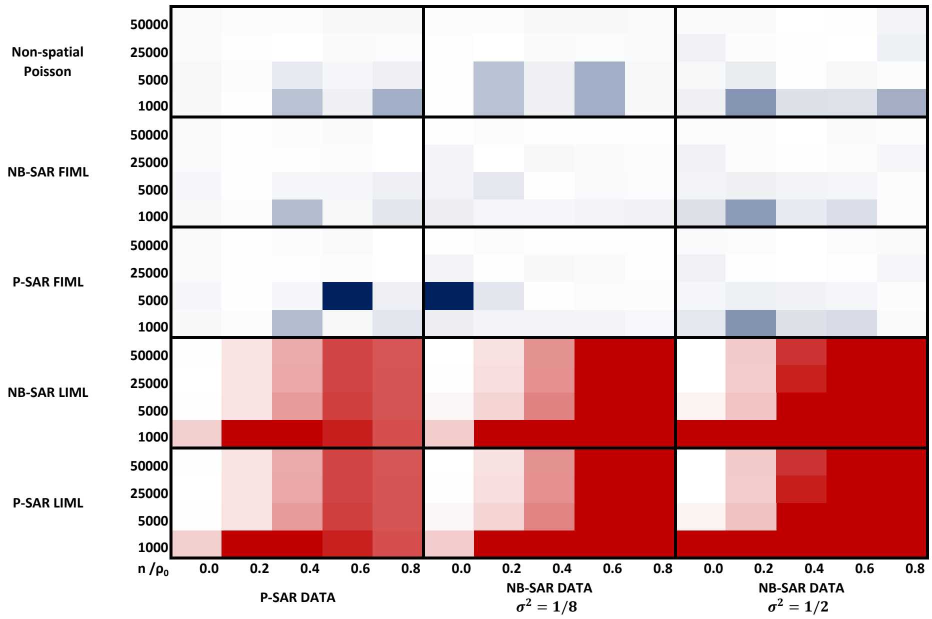
with $X_{i.}$ being the i^{th} row of the explanatory variables matrix X. The relative bias is obtained as the bias of the median estimated marginal effect divided by the true marginal effect according to the DGP.

Figures 7 and 8 show heat grids of the relative bias in the median estimated direct and total marginal effects. Each rectangle represents one specification consisting of DGP, sample size n, strength of spatial autocorrelation λ, and estimated model. The red shading indicates the size of the deviation from zero on a scale which takes the darkest color for absolute deviations of 0.5 and white for values close to zero. This scale is used to visualize the results of the LIML estimations. Because the deviations in the estimated marginal effects from the FIML estimations are in general smaller, a finer scale is necessary to make differences among the FIML estimation results visible. Therefore, the blue shading indicates the size of the deviation from zero on a scale which takes the darkest color for absolute deviations of 0.1 and again white for values close to zero. The visualization of Monte Carlo results in spatial econometrics through heat grids has been proposed by Arribas-Bel et al. (2012) under the name "nested spatial maps."

In most cases the relative bias of the marginal effect of X_1 is larger than the one of X_2, which stems from the higher variance in the DGP of X_2. The two lower rows of blocks display the results for the LIML estimations which are much worse for all cases except for $\lambda = 0$. There, LIML outperforms FIML in terms of bias for some specifications. Additionally, LIML estimation seems to be very sensitive to small sample sizes and large spatial correlation. After the decrease in bias from $n = 1000$ to $n = 5000$, the bias stays of the same magnitude when the sample size increases further. Interestingly, the estimations for P-SAR data are better than the ones for NB-SAR data but still worse than the complementary FIML estimations.



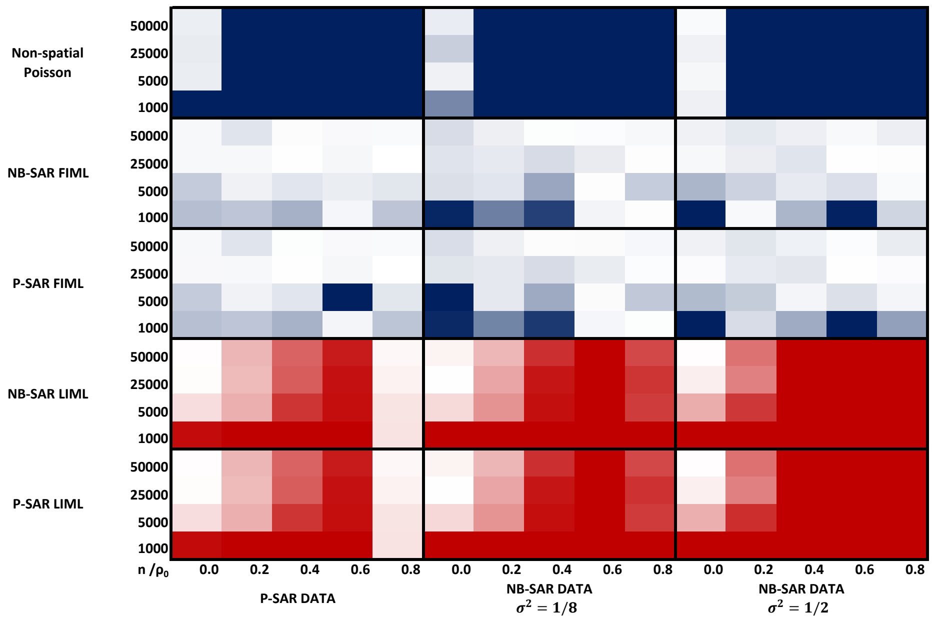
(a) Relative bias of the estimated direct marginal effects of X_1



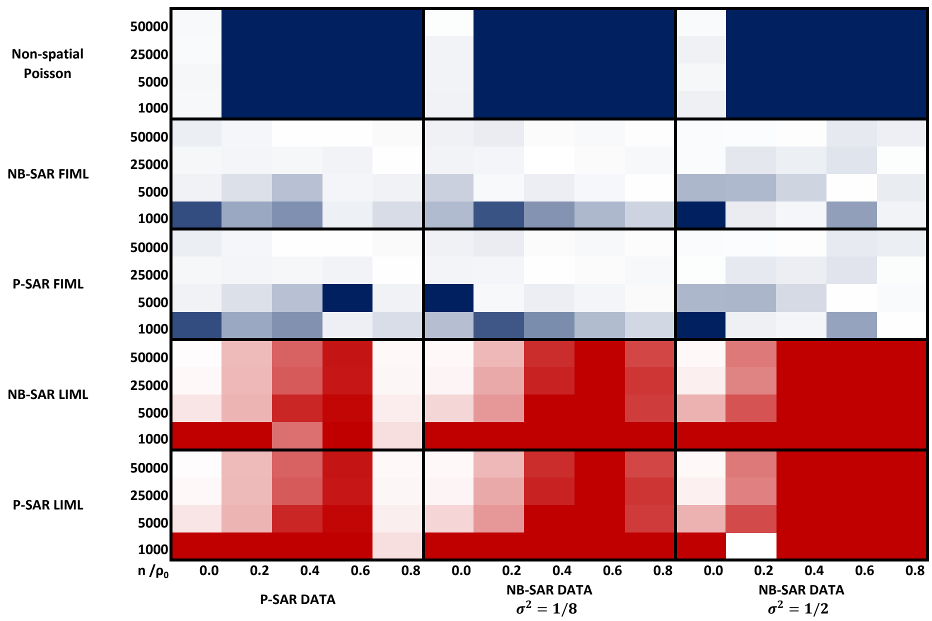
(b) Relative bias of the estimated direct marginal effects of X_2

Figure 7: Heat maps for the relative bias of estimated median direct marginal effects for all generated data sets. The blue shading ranges from -0.1 to 0.1, the red shading ranges from -0.5 to 0.5.

The next two blocks above the LIML blocks display the results for the FIML estimation of the P-SAR and NB-SAR models. The relative biases here are generally much smaller than the corresponding ones of the LIML estimations and show the expected behavior of decreasing bias with increasing sample size. The strength of the spatial correlation in the data has no clear effect on the bias of the estimated marginal effects, however.



(a) Relative bias of the estimated total marginal effects of X_1



(b) Relative bias of the estimated total marginal effects of X_2

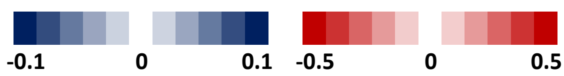


Figure 8: Heat maps for the relative bias of estimated median total marginal effects for all generated data sets. The blue shading ranges from -0.1 to 0.1, the red shading ranges from -0.5 to 0.5.

Finally, the upper block of rows shows the results from the non-spatial Poisson estimation. With regard to the direct effects, the relative biases are of the same magnitude than the ones of the FIML estimations of the spatial models. The total effects are naturally underestimated by the non-spatial model, in cases with positive spatial correlation, since it does not allow for any indirect effects of neighboring observations (see Chapter 1.2 for definitions of direct and indirect impacts).

Tables 5 to 7 give an overview of the calculated RMSE, bias and relative bias of the marginal effects. Results for all specifications can be found in Appendix A.2. LIML also performs worse than FIML in terms of RMSE. With exception of the small n, large λ case with NB-SAR data, the average RMSE of the non-spatial Poisson estimation is even smaller than for the (correct) spatial models.

The numbers presented in this section confirm the results from the previous one in the sense that the estimates of the marginal effects from the FIML estimation are clearly superior to those of the LIML estimation. As outlined in the previous section, this is not surprising, since FIML uses more information in the estimation and the computation of large inverses is manageable with the computational power available nowadays. Misspecification of the dispersion, i.e. estimation of Poisson data with a negative binomial model and vice versa, has no relevant consequences on the consistency and precision of the estimates from a FIML estimation. This is in line with the theoretical result of Gourieroux et al. (1984) for non-spatial models. The estimated direct marginal effects of the non-spatial models have slightly higher biases than the ones of the spatial models, but still, the non-spatial model is able to estimate the marginal effects well, even though the spatial structure in the data is ignored. This indicates that the spatial structure of the DGP could also be regarded as a nuisance (and for example modelled in a spatial error model) if one is only interested in the effects of the explanatory variables. The estimates of the total marginal effects obviously cannot be as accurately estimated by a non-spatial model, since such a model does not allow indirect impacts.

	P-SAR - FIML Total Effects			**NB-SAR - FIML Total Effects**			**Non-spatial Poisson Direct Effects**		
	RMSE	**Bias**	**ReBias**	**RMSE**	**Bias**	**ReBias**	**RMSE**	**Bias**	**ReBias**
$\lambda = 0.2$		$n = 1000$							
X_1	0.117	0.005	0.026	0.117	0.005	0.026	0.077	−0.004	−0.025
X_2	0.069	0.007	0.039	0.069	0.007	0.039	0.018	0.000	0.001
$\lambda = 0.2$		$n = 50000$							
X_1	0.014	0.002	0.013	0.014	0.002	0.013	0.010	0.001	0.008
X_2	0.009	0.001	0.004	0.009	0.001	0.004	0.003	0.000	−0.001
$\lambda = 0.8$		$n = 1000$							
X_1	0.482	0.061	0.026	0.482	0.061	0.027	0.120	0.010	0.019
X_2	0.243	0.036	0.016	0.243	0.036	0.016	0.044	−0.020	−0.036
$\lambda = 0.8$		$n = 50000$							
X_1	0.067	−0.005	−0.002	0.067	−0.005	−0.002	0.016	−0.001	−0.001
X_2	0.034	−0.004	−0.002	0.034	−0.005	−0.002	0.005	−0.002	−0.003

Table 5: Monte Carlo results of median marginal effects for P-SAR data, $\lambda = \{0.2, 0.8\}$, $n = \{1000, 50000\}$. The relative bias (ReBias) is calculated as the bias divided by the true marginal effect according to the DGP. Values reported as 0.000 are smaller than 0.0005.

	P-SAR - FIML Total Effects			NB-SAR - FIML Total Effects			Non-spatial Poisson Direct Effects		
	RMSE	**Bias**	**ReBias**	**RMSE**	**Bias**	**ReBias**	**RMSE**	**Bias**	**ReBias**
$\lambda = 0.2$		$n = 1000$							
X_1	0.127	0.010	0.056	0.126	0.011	0.057	0.073	−0.007	0.029
X_2	0.074	0.014	0.076	0.074	0.015	0.078	0.022	0.000	0.028
$\lambda = 0.2$		$n = 50000$							
X_1	0.523	−0.002	−0.006	0.015	−0.001	−0.007	0.011	0.000	−0.013
X_2	0.327	0.041	−0.008	0.009	−0.002	−0.008	0.003	0.000	−0.001
$\lambda = 0.8$		$n = 1000$							
X_1	0.015	−0.001	−0.001	0.515	−0.002	−0.001	0.144	0.025	−0.107
X_2	0.009	−0.002	0.018	0.334	0.046	0.020	0.045	−0.014	0.004
$\lambda = 0.8$		$n = 50000$							
X_1	0.077	−0.007	−0.003	0.077	−0.007	−0.003	0.018	−0.002	0.008
X_2	0.046	−0.002	−0.001	0.045	−0.002	−0.001	0.007	−0.001	−0.002

Table 6: Monte Carlo results of median marginal effects for NB-SAR data, $\lambda = \{0.2, 0.8\}$, $n = \{1000, 50000\}$, $\alpha = 1/8$. The relative bias (ReBias) is calculated as the bias divided by the true marginal effect according to the DGP. Values reported as 0.000 are smaller than 0.0005.

	P-SAR - FIML Total Effects			NB-SAR - FIML Total Effects			Non-spatial Poisson Direct Effects		
	RMSE	**Bias**	**ReBias**	**RMSE**	**Bias**	**ReBias**	**RMSE**	**Bias**	**ReBias**
$\lambda = 0.2$		$n = 1000$							
X_1	0.148	0.003	0.016	0.150	−0.001	−0.003	0.097	−0.004	−0.027
X_2	0.076	−0.001	−0.007	0.077	−0.002	−0.008	0.028	−0.007	−0.048
$\lambda = 0.2$		$n = 50000$							
X_1	0.018	0.002	0.012	0.018	0.002	−0.019	0.014	0.002	0.012
X_2	0.011	0.000	0.002	0.011	0.000	0.006	0.004	0.000	−0.002
$\lambda = 0.8$		$n = 1000$							
X_1	0.838	−0.097	−0.042	0.817	−0.043	0.011	0.243	0.015	0.026
X_2	0.477	0.002	0.001	0.485	0.013	0.002	0.076	−0.020	−0.036
$\lambda = 0.8$		$n = 50000$							
X_1	0.130	−0.019	−0.008	0.126	−0.017	−0.007	0.033	−0.006	−0.011
X_2	0.069	−0.018	−0.008	0.067	−0.016	−0.007	0.010	−0.003	−0.005

Table 7: Monte Carlo results of median marginal effects for NB-SAR data, $\lambda = \{0.2, 0.8\}$, $n = \{1000, 50000\}$, $\alpha = 1/2$. The relative bias (ReBias) is calculated as the bias divided by the true marginal effect according to the DGP. Values reported as 0.000 are smaller than 0.0005.

3.5 Model Selection: Scoring Rules

With the help of scoring rules, the fit of the predictive distribution to the observed data can be compared among models. Like information criteria which are based on the likelihood values of the estimates, scoring rules negatively penalize a worse fit of the model to the data and are therefore to be minimized. This method of model selection is adopted from the times series of counts literature (see Czado et al. (2009) and Jung et al. (2016)). The score is averaged over all observations for model selection. Several scores are suggested in the count data literature. In the following, the logarithmic, quadratic and ranked probability scores are presented, which place their emphasis on different aspects of the estimated distributions.

The logarithmic score focusses on the predicted probability of the observed count and penalizes predictive distributions for which the observation has a small probability. The average logarithmic score is

$$LogS = \frac{1}{n}\sum_{i=1}^{n} - \log(\hat{P}_i(y_i)) \tag{55}$$

where $\hat{P}_i(y_i)$ is the predictive probability distribution of observation y_i.

The quadratic score also considers the whole estimated probability distribution by adding the squared probabilities of all possible outcomes:

$$QS = \frac{1}{n}\sum_{i=1}^{n} -2\hat{P}_i(y_i) + \sum_{j=0}^{\infty} \hat{P}_i(j)^2 \tag{56}$$

As a third scoring rule the ranked probability score is used which especially penalizes a flat estimated distribution:

$$RPS = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{\infty}(F(j) - \mathbb{I}(y_i \leq j))^2 \tag{57}$$

An important property of scoring rules is propriety, i.e. the score takes its worst value for the worst possible fit and its best value for the best. This property holds for all three scoring rules presented here (Czado et al. (2009) and Jung et al. (2016)).

3.6 Empirical Application: Start-up Firm Births

After the Monte Carlo study has shown that the FIML estimation works well for this model type, the P-SAR model and its extensions introduced in Section 3.3 are applied to the start-up firm births data set also used by Lambert et al. (2010).

3.6.1 Data

For the application of the SAR count models the cross-sectional data of Lambert et al. (2010) on aggregated firm births in the manufacturing sector of the United States between 2000 and 2004 is revisited. The data contains the number of start-up firms during this time period for 3078 U.S. counties as well as variables measuring the location factors like market structure, labor market, and infrastructure for each county (see Table 8). Figure 9 displays the spatial structure of the number of firm births. Large numbers of new firms (marked dark red) can be found at the west and east coast and around the Great Lakes. Almost no start-up firm births (dark blue areas) were registered in the Midwest. The frequency distribution for values up to 100 is shown in Figure 10. The data does not exhibit excess zeros but an extraordinarily large range (from 0 to 6938) for a counts process. Descriptive statistics for the variables in the data set are shown in Table 60 in Appendix A.3.

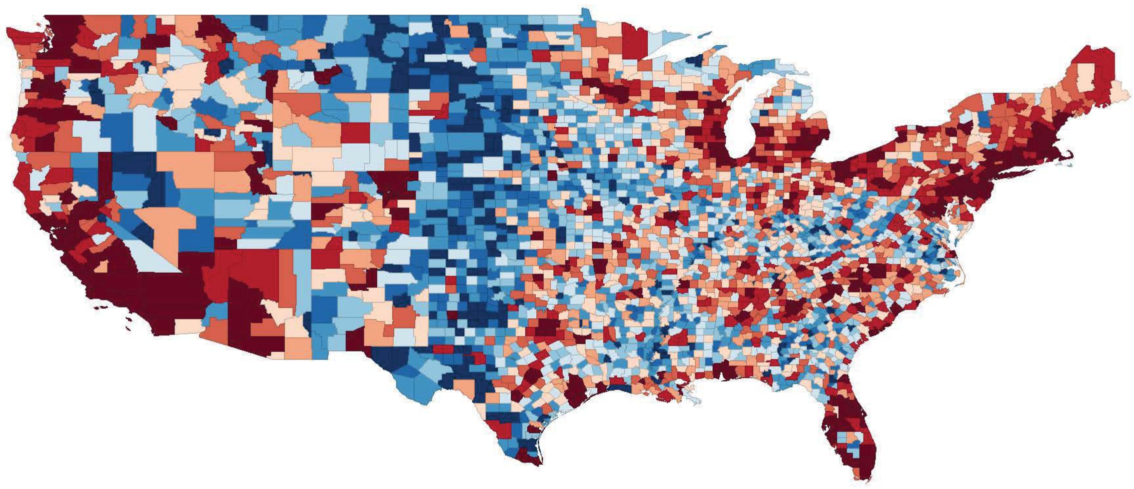


Figure 9: Map of observations for *subirth* in deciles with dark blue representing the lowest values and dark red representing the highest ones.

In the empirical analysis four different spatial weight matrices are used to check the robustness of the results against the choice of the spatial structure.
First, an 8 nearest neighbors inverse distance matrix (W_{dnn}) is implemented which has also been employed in the empirical investigation of Lambert et al. (2010). It is obtained by calculating the inverse distance matrix of the centroids of the counties, then keeping only

	subirths	Single unit start-ups in the lower 48 United States during 2000-2004 in the manufacturing sector (NAICS 31-33)
Agglomeration economies	msemp	Manufacturing share of employment
	tfdens	Total establishment density (in 100s)
	pel10emp	Percent of manufacturing establishments with less than 10 employees
	pem100emp	Percent of manufacturing establishments with more than 100 employees
Market structure	mhhi	Median household income (in 1000s)
	pop	Population (in 10000s)
	cclass	Share of workers in creative occupations
Labor availability and cost	uer	Unemployment rate
	pedas	Percent of adults with an associate's degree
	awage	Average wage per job (in 1000s)
	netflow	Net flow of wages per commuter (in 1000s)
Infrastructure	proad	Public road density
	interst	Interstate highway miles
	hwypc	Government expenditures on highways per capita (in 100s)
	avland	Percent of farmland to total county
Fiscal policy	educpc	Government expenditures on education per capita (in 100s)
	bci	State tax business climate index (higher values indicate more favorable business climates)
Area	metro	Dummy variable identifying counties as belonging to metropolitan areas
	micro	Dummy variable identifying counties as belonging to micropolitan areas

Table 8: Description of start-up firm births data. For detailed information on data sources see Lambert, Brown, and Florax (2010), pp. 249.

the values for the 8 nearest neighbors of each observation and setting all other elements to zeros.

Next, a queen contiguity matrix (W_{con}) which only indicates whether two counties share a common border or vertex, a nearest neighbors matrix (W_{nn}) which has a positive value (i.e. 1) for each of the 8 nearest neighbors, and a full inverse distance matrix which consists of the inverse distances (W_d) of each pair of counties in the sample, are considered. All matrices are row-standardized. Table 9 gives some descriptives for each weight matrix and Moran's I of the dependent variable using the respective weight matrix.

Moran's I (introduced by Moran (1950) for binary weights and generalised for arbitrary weight matrices by Cliff and Ord (1981, p. 17)) is the most prevalent measure of spatial association. It quantifies the dependence across the complete data set by summarising cross-products of deviations from the mean:

$$I = \frac{n}{W_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{58}$$

	W_{dnn}	W_{con}	W_{nn}	W_d
Mean	8	5.985	8	3077
Min	8	3	8	3077
Max	8	12	8	3077
Sum of Relations	24624	18422	24624	9471006
Moran's I	0.3115	0.2471	0.2331	0.0224

Table 9: Number of neighbors in different spatial weight matrices for the start-up firm births data.

where W is the assumed spatial weight matrix and $W_0 = \sum_{i=1}^{n} \sum_{j \neq i}^{n} w_{ij}$, which equals n for row-standardized weight matrices.
Unlike its counterparts from time series analysis, its mean is not 0 but $-\frac{1}{(n-1)}$ and it is not bounded between -1 and 1. Unfortunately, there is no distributional theory available for Moran's I calculated from count data. If one assumed an asymptotic normal distribution (as it is done for the measure calculated from normally distributed data, see Cliff and Ord (1981, pp. 19)), the standardized I using W_{dnn}, W_{con} or W_{nn} as spatial structure would lie far above any relevant significance levels and vote for significant spatial association in the data. The small value of Moran's I for the full inverse distance matrix is remarkable. Apparently, the size of the map with 3078 counties is too large to employ a full weight matrix which leads to more than 9 million relations between the counties.

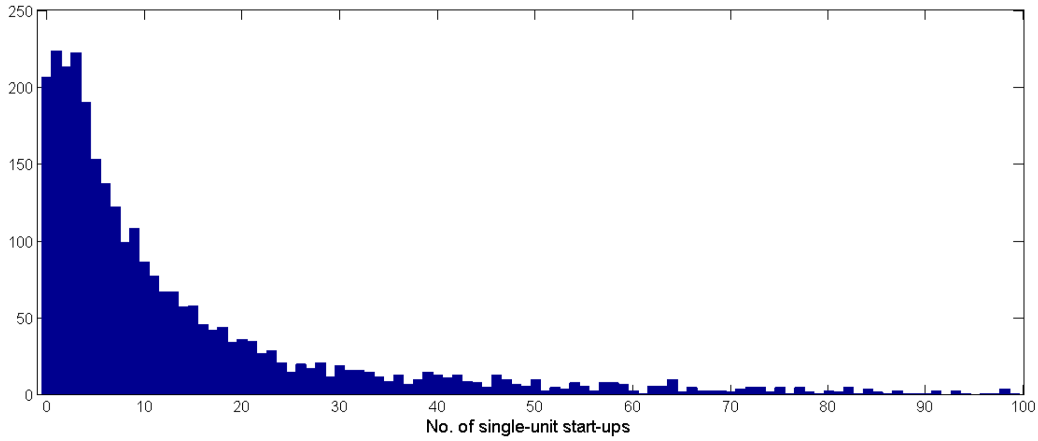


Figure 10: Histogram of start-up firm birth counts in the manufacturing sector, U.S. counties, 2000-2004, x-axis is cut at 100.

3.6.2 Empirical Results

Table 10 gives an overview of all estimates for the spatial correlation parameter λ.[7] If an estimate is given in brackets, convergence could not be achieved with standard methods for the respective combination of model and spatial weight matrix. The estimates of the specifications with nearest neighbors and contiguity matrices are of the same magnitude for the same regression model and range from about 0.17 to 0.29. The estimates from NB-SAR models are the highest. The estimates from the P-SAR model are between 0.23 and 0.27 and larger than the ones of the zero-inflated and hurdle Poisson models which give the smallest values. Estimations using the full inverse distance matrix turned out to be difficult. Convergence might have been reached for three of the four models, but results differ in a way that no clear statement can be made. It seems likely that the extremely large number of spatial relationships imposed by this matrix at a sample size of 3078 causes difficulties when estimating the model, even though the weights are diminishing for increasing distance between the units. Therefore, this spatial weight matrix will not be considered any further. Tables with all parameter estimates can be found in Appendix A.4.

	W_{dnn}	W_{con}	W_{nn}	W_d
P-SAR	0.2533***	0.2774***	0.2306***	−0.0353
	(0.0634)	(0.0655)	(0.0537)	(0.5407)
NB-SAR	0.2803***	0.2902***	0.2932***	0.5416***
	(0.0297)	(0.0321)	(0.0257)	(0.0534)
ZIP-SAR	0.1685**	0.2000***	0.1878***	[0.0028]
	(0.0728)	(0.0784)	(0.0559)	
HP-SAR	0.1686**	0.2001***	0.1879***	0.0019
	(0.0727)	(0.0786)	(0.0559)	(0.3201)

Table 10: Estimation results for λ from SAR models for the start-up firm births data. Standard errors in parentheses, ** and *** denote a 5% and 1% significance, respectively. Convergence was not achieved if result is given in brackets.

Table 11 gives the scoring rules of the P-SAR, NB-SAR, ZIP-SAR, and HP-SAR estimations. Log score and quadratic score are smallest for the negative binomial specifications for each spatial weight matrix as well as among the non-spatial models. The rank probability score is smallest for the zero-inflated Poisson specifications, again regardless of the choice of the weight matrix. Between weight matrices the score values are similar. Log score and ranked probability score slightly prefer the 8 nearest neighbors matrix (W_{nn}), the ranked probability score is smallest for the 8 nearest neighbors inverse distance matrix (W_{dnn}). Because of these numbers, the hurdle model is excluded from the further analysis since it is outperformed by the other models with regard to all scoring rules employed here. Among the remaining models, no clear choice can be made since depending on the chosen scoring rule either the NB-SAR or the ZIP-SAR model is preferred.

[7] All spatial regressions and the non-spatial hurdle Poisson model are computed using MATLAB code written by myself, the remaining non-spatial regressions are executed using the build-in procedures in STATA.

		P-SAR	NB-SAR	ZIP-SAR	HP-SAR
W_{dnn}	LogS	9.720	3.356	9.157	9.180
	QS	−0.020	−0.073	−0.047	0.284
	RPS	15.182	25.967	11.374	43.770
W_{con}	LogS	9.692	3.356	9.041	9.066
	QS	−0.021	−0.073	−0.048	0.296
	RPS	15.158	27.210	11.380	55.543
W_{nn}	LogS	9.697	3.354	9.038	9.063
	QS	−0.021	−0.073	−0.048	0.290
	RPS	15.293	31.604	11.441	89.305
Non-spatial	LogS	9.313	3.379	9.306	9.313
	QS	−0.017	−0.070	−0.047	0.541
	RPS	15.244	37.858	11.731	190.447

Table 11: Scoring rules of SAR model estimates for the start-up firm births data.

Figure 11 compares the relative frequencies in the data (gray bins) to the predicted probabilities from the P-SAR, NB-SAR and ZIP-SAR estimations with W_{dnn} as the spatial weight matrix, which will also be used for the analyses in the following chapter. The zeros are naturally best met by the ZIP-SAR model, which then leads to an underestimation of the probabilities for values between 1 and 4. The P-SAR model predicts the probabilities for small values larger than 0 best. The NB-SAR predicts a larger probability for zeros than the P-SAR model, which is more suitable to the data, but performs worse for small positive outcomes. But then, it presents a smoother decrease in probabilities for large values than the Poisson models. To sum up, this figure shows that none of the models employed here is able to accurately capture the structure in the data, i.e. the large number of observations smaller than 10, as well as the extraordinarily large observations up to 6000.

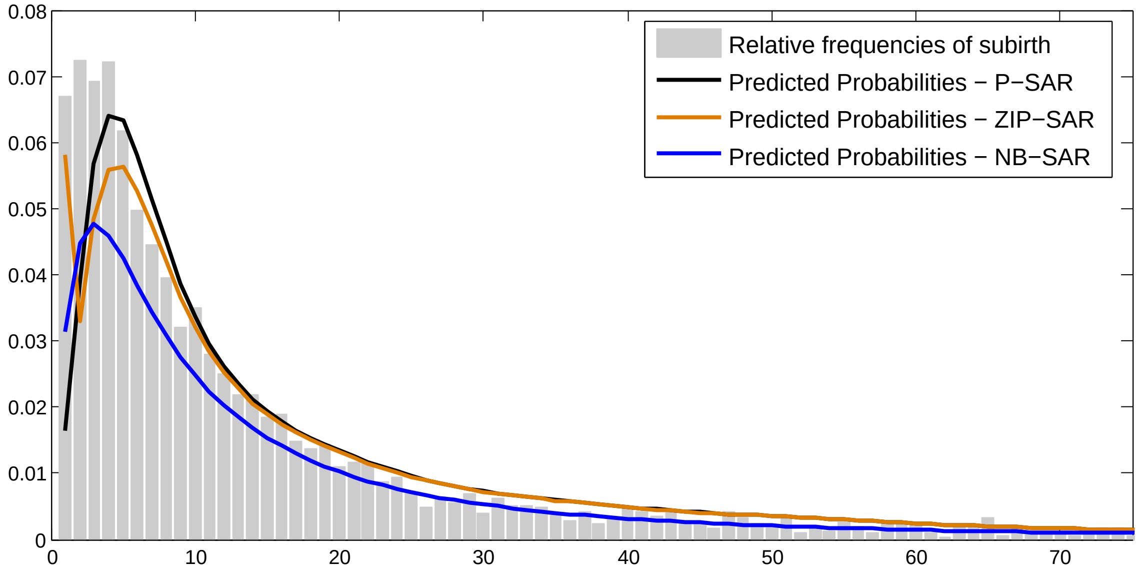


Figure 11: Predicted probabilities of P-SAR, ZIP-SAR, NB-SAR for the start-up firm births data and weighting matrix W_{dnn} .

Finally, the predicted marginal effects of the regressors are computed. The formulas of the marginal effects for the P-SAR and NB-SAR models are given in Equations (38) - (40). The marginal effects of the ZIP-SAR model are obtained by differentiating Equation (45) with respect to X. In this application the regressors of the inflation process and the Poisson process are identical. For this special case, the marginal effects take the form

$$\frac{\partial E[y_i|X]}{\partial x_{ik}} = \frac{\gamma_k \exp(X_{i.}\gamma)}{(1+\exp(X_{i.}\gamma))^2}\Big(a_{ii}\exp(A_{i.}X\beta)\beta_k - \exp(A_{i.}X\beta)\Big) \tag{59}$$

for the direct effects where A denotes the Leontief inverse, i.e. $A = (I - \lambda W)^{-1}$. $A_{i.}$ denotes the i^{th} row and a_{ii} the respective element of this matrix.
The indirect effects summarized over all neighbors are given by

$$\sum_{\substack{j=1\\ j\neq i}}^{n} \frac{\partial E[y_i|X]}{\partial x_{jk}} = \sum_{\substack{j=1\\ j\neq i}}^{n} \frac{1}{1+\exp(X_{i.}\gamma)}\Big(a_{ij}\exp(A_{i.}X\beta)\beta_k\Big) \tag{60}$$

and the total marginal effects equal the sum of direct and indirect effects.

Table 20 gives the results for the W_{dnn} matrix, which is chosen for better comparability to the following chapter. The predictions using W_{con} and W_{nn} do not differ relevantly and are reported in Appendix A.4, Tables 65 and 66. Following LeSage and Pace (2009, pp. 39), the standard errors for the marginal effects can be obtained using their sample counterparts of draws from the asymptotic joint distribution of the parameter estimators (this procedure is also applied e.g. by Baltagi et al. (2014), Fischer et al. (2009), and Liesenfeld et al. (2016b)). The largest number of significant marginal effects is obtained in the NB-SAR model, closely followed by the P-SAR model. The ZIP-SAR specification leads to only few significant effects. In the first two models the direct effects are in general larger than the indirect ones. The opposite holds for the ZIP-SAR model. Aside from the effect of being in a metro- or micropolitan as opposed to a rural area, the largest effects on the number of start-up firms in the P-SAR model are predicted for education (percentage of adults with an associate's degree, *pedas*), public road density (*proad*) and the business climate (state tax business climate index, *bci*). The ZIP-SAR shows a similar pattern in the total effects, but also predicts a relatively large influence of the unemployment rate (*uer*). For the NB-SAR model the highest values are obtained for unemployment rate, public road density, and the share of workers in creative occupations (*cclass*).

3.7 Summary

The presented Monte Carlo results show that FIML estimation clearly outperforms LIML estimation. Furthermore, the LIML estimators do not show consistency. FIML also shows considerable bias for λ in small sample sizes ($n = 1000$) but it is in general decreasing for increasing true λ and increasing sample size, respectively. Ignoring the spatial structure of the kind employed in this chapter does not lead to a larger bias or RMSE in the direct marginal effects. Additionally, data without any spatial structure ($\lambda = 0$) is correctly identified by the spatial models and for sample sizes of at least 5000 even the LIML procedure performs well in this case. Looking at the FIML results, there are no relevant differences in bias of $\hat{\lambda}$ and of the marginal effects observed between the estimation results of NB-SAR data with $\alpha = 1/2$ and $\alpha = 1/8$, respectively.

In the empirical application, the SAR count structure has been transferred to a zero-inflated Poisson and a hurdle Poisson specification. The employed scoring rules as well as a visual inspection of the predicted probability in comparison to the observed relative frequencies do not lead to a clear choice for the best model. The scoring rules choose either the NB-SAR or ZIP-SAR model. The predicted probabilities plot indicates that the two Poisson specifications are in general better in reproducing the distribution for small values whereas the NB-SAR model better predicts the probabilities of very large outcomes. With exception of the full inverse distance matrix, with which estimation is difficult, the choice of the spatial weight matrix had no relevant influence on the results.

Variable	P-SAR Total M.E.	P-SAR Direct M.E.	P-SAR Indirect M.E.	NB-SAR Total M.E.	NB-SAR Direct M.E.	NB-SAR Indirect M.E.	ZIP-SAR Total M.E.	ZIP-SAR Direct M.E.	ZIP-SAR Indirect M.E.
msemp	0.471***	0.302***	0.149***	0.720***	0.411***	0.234***	0.146***	0.008**	0.090**
	(0.075)	(0.041)	(0.049)	(0.047)	(0.021)	(0.033)	(0.045)	(0.003)	(0.036)
pelt10	−0.012	−0.007	−0.004	0.084***	0.048***	0.027***	−0.005	0.001	−0.007
	(0.034)	(0.021)	(0.011)	(0.021)	(0.012)	(0.008)	(0.010)	(0.001)	(0.009)
pemt100	−0.447***	−0.287***	−0.141***	−0.271***	−0.155***	−0.088***	−0.105**	0.002	−0.114***
	(0.070)	(0.037)	(0.047)	(0.037)	(0.020)	(0.017)	(0.042)	(0.002)	(0.043)
tfdens	0.078	0.050	0.025	−0.985***	−0.562***	−0.320***	−0.039	−0.072	0.026
	(0.184)	(0.111)	(0.066)	(0.287)	(0.164)	(0.100)	(0.073)	(0.045)	(0.041)
mhhi	−0.202	−0.130	−0.064	0.146	0.083	0.048	0.037	0.000	0.038
	(0.121)	(0.075)	(0.045)	(0.117)	(0.068)	(0.037)	(0.042)	(0.002)	(0.040)
pop	0.036***	0.023***	0.012**	0.248***	0.142***	0.081***	0.037***	0.015	0.008**
	(0.009)	(0.004)	(0.005)	(0.041)	(0.023)	(0.016)	(0.013)	(0.009)	(0.004)
cclass	0.825***	0.529***	0.261***	1.362***	0.776***	0.443***	0.174**	0.009	0.110
	(0.201)	(0.126)	(0.096)	(0.127)	(0.051)	(0.078)	(0.075)	(0.005)	(0.064)
uer	0.999***	0.641***	0.316**	0.901***	0.514***	0.293***	0.230**	0.002	0.215
	(0.336)	(0.208)	(0.141)	(0.208)	(0.125)	(0.071)	(0.111)	(0.004)	(0.111)
pedas	1.742***	1.116***	0.551***	0.553***	0.315***	0.180***	0.315**	−0.002	0.325**
	(0.292)	(0.153)	(0.189)	(0.120)	(0.073)	(0.041)	(0.139)	(0.003)	(0.141)
awage	0.317***	0.203***	0.100**	−0.613***	−0.350***	−0.199***	−0.003	0.001	−0.005
	(0.113)	(0.077)	(0.041)	(0.112)	(0.064)	(0.042)	(0.042)	(0.002)	(0.040)
netflow	0.035	0.022	0.011	−0.273***	−0.156***	−0.089***	−0.032	0.002	−0.037
	(0.038)	(0.023)	(0.013)	(0.050)	(0.027)	(0.019)	(0.018)	(0.001)	(0.019)
proad	1.110***	0.712***	0.351***	0.840**	0.479**	0.273**	0.211**	−0.007	0.236**
	(0.265)	(0.186)	(0.116)	(0.390)	(0.227)	(0.129)	(0.105)	(0.009)	(0.108)
interst	0.103***	0.066***	0.032***	0.067***	0.038***	0.022***	0.026***	0.001	0.019***
	(0.015)	(0.011)	(0.009)	(0.012)	(0.007)	(0.005)	(0.008)	(0.001)	(0.007)
avland	−0.085***	−0.054***	−0.027***	−0.069***	−0.039***	−0.022***	−0.021***	0.000	−0.023***
	(0.018)	(0.011)	(0.009)	(0.009)	(0.006)	(0.004)	(0.008)	(0.000)	(0.008)
bci	1.455***	0.933***	0.460**	0.480**	0.274	0.156**	0.243**	0.000	0.244**
	(0.440)	(0.281)	(0.186)	(0.236)	(0.140)	(0.075)	(0.122)	(0.004)	(0.121)
educpc	0.045	0.029	0.014	0.059	0.034	0.019	0.006	0.000	0.007
	(0.032)	(0.020)	(0.011)	(0.042)	(0.025)	(0.014)	(0.008)	(0.000)	(0.007)
hwypc	−0.418	−0.268	−0.132	−0.456	−0.260	−0.148	−0.084	−0.001	−0.076
	(0.313)	(0.194)	(0.112)	(0.340)	(0.199)	(0.110)	(0.075)	(0.001)	(0.073)
metro	28.028***	10.309***	17.488***	14.503***	5.207***	9.136***	26.135***	13.494***	12.709***
	(4.942)	(2.013)	(5.673)	(1.192)	(0.488)	(0.746)	(4.092)	(2.361)	(4.662)
micro	10.230***	3.824***	6.293***	8.031***	2.943***	5.100***	−0.606	4.281***	−2.804
	(1.883)	(0.661)	(2.121)	(1.477)	(0.466)	(1.408)	(1.489)	(0.709)	(1.454)

Table 12: Median marginal effects of P-SAR, NB-SAR, ZIP-SAR for the start-up firm births data with weight matrix W_{dnn}. For the dummy variables *metro* and *micro* the effect of a change from 0 to 1 is given. ** and *** denote a 5% and 1% significance, respectively, Standard errors (in brackets) are estimated using their sample counterparts of 2000 draws of the asymptotic joint distribution of the coefficients. Values reported as 0.000 are smaller than 0.0005.

4 The Spatial Linear Feedback Model

4.1 Introduction

In the P-SAR model of Lambert et al. (2010), which is discussed in Chapter 3, the intensity of the firm birth process at location i is related to the intensity of its neighboring regions. This allows the authors to obtain a reduced form of the conditional mean function of the counts and to derive a spatial multiplier term, i.e the Leontief inverse (Equation (24)). The interpretation of the spatial dependence between neighboring regions in this model is less intuitive than in a continuous SAR model, especially because the intensity is an unobserved measure. Therefore, the spatial dependency is only part of an unobserved process. In contrast, a spatial linear feedback model (SLFM) is proposed in this chapter which allows for an interpretation of the spatial dependence parameter closer to that of the continuous SAR model. The spatial dependence parameter of the SLFM denotes the average change in the conditional expectation of the count in region i given a one unit change in the (observable) counts of the neighboring regions (row-standardized weight matrix assumed). Therefore, the spatial dependency is not completely driven by an unobservable process.

Moreover, the model of Lambert et al. (2010) allows for spatial dependence in the exogenous regressors only, since its conditional expectation does not contain an error term. The latter criticism has been taken up in the paper by Liesenfeld et al. (2016b) where the P-SAR model is extended with an additional error term allowing for spatial dependence in the unexplained part of the conditional expectation, too. While the P-SAR model of Lambert et al. (2010) is easily estimated via maximum likelihood, the estimation of the model proposed by Liesenfeld et al. (2016b) is not straightforward and their proposal to use efficient importance sampling estimation is not routinely available.

The SLFM is proposed as an alternative which is convenient to estimate and to interpret for empirical economists interested in studying spatially correlated count data. In particular, it overcomes the numerical difficulties associated with the full information maximum likelihood estimation of the P-SAR model caused by the need to invert a transformation of the $n \times n$ spatial weight matrix, where n denotes the sample size. The SLFM is based on a Poisson regression model (P-SLFM). As a second variant, a negative binomial version (NB-SLFM) is introduced. Unlike the Poisson model, its variance function is more flexible and allows for overdispersion which is able to capture unobserved heterogeneity in the data.

Finally, two visual diagnostic tools and model validation methods are adopted from the time series literature of counts to assess the adequacy of a fitted model and to compare two or more competing model specifications: a suitably adjusted variant of the probability integral transform (PIT) and a relative deviations plot. Additionally, the scoring rules already discussed in Section 3.5 are used for model evaluation.

4.2 Modelling Approach

To overcome the shortcomings of the two models discussed above, a model specification is proposed that follows up Besag's (1974) model in some sense but introduces spatially lagged counts additively into the conditional expectation equation of a Poisson regression or negative binomial regression, respectively. The conditional expectation of the SLFM is then given as

$$y_i|\mu_i \sim D(\mu_i) \tag{61}$$

$$\mu_i = E[y_i|Y_{-i}, X] = \lambda \sum_{\substack{j=1 \\ j\neq i}}^{n} w_{ij} y_j + \exp(X_{i.}\beta) \tag{62}$$

with $X_{i.}$ denoting the i^{th} row of the matrix of exogenous regressors, Y_{-i} the observations of all neighbors of i, and w_{ij} an element of the spatial weight matrix W. In terms of the model classification introduced in Chapter 2.4, this model classifies as an observation-driven model since the neighboring observations enter the latent process, which includes no additional random process as opposed to parameter-driven models. Several count data distributions can be assumed for D, the most prominent being the Poisson. The resulting model will be denoted as P-SLFM. Further, to deal with unobserved heterogeneity in the counts, a negative binomial distribution is used and the resulting model will be denoted as NB-SLFM.

Because a reduced form is not readily available for these SLF models and due to the lack of an operational multivariate count distribution for large n, the true likelihood function cannot be obtained. For estimation of the model the full likelihood is approximated with a pseudo conditional likelihood. The idea to compose a pseudo likelihood function of conditional probability functions like the ones described in Equations (61)-(62) stems from Besag (1975), who proposed this technique for the auto-normal schemes introduced in Besag (1974). He also sketches a proof of the estimator's consistency relating it to the coding technique: Besag points out that the obtained estimator can be thought of as a weighted average of coding estimators. Since these are consistent, the estimator of the conditional pseudo likelihood approach is consistent as well (under suitable regularity conditions).

In case of a P-SLFM specification, the corresponding pseudo log likelihood function is given by

$$\begin{aligned} \log L_C &= \sum_{i=1}^{n} y_i \log\left(\lambda \sum_{\substack{j=1 \\ j\neq i}}^{n} w_{ij} y_j + \exp(X_{i.}\beta)\right) \\ &\quad - \left(\lambda \sum_{\substack{j=1 \\ j\neq i}}^{n} w_{ij} y_j + \exp(X_{i.}\beta)\right) - \log(y_i!) \end{aligned} \tag{63}$$

In case of the NB-SLF model, the pseudo log likelihood function takes the form

$$\begin{aligned}\log L_C &= \sum_{i=1}^{n}\left(\sum_{j=0}^{y_i-1}\log(\alpha^{-1}+j)\right)-\log(y_i!)\\ &-(y_i+\alpha^{-1})\log\left(1+\alpha\left(\lambda\sum_{\substack{j=1\\ j\neq i}}^{n}w_{ij}y_j+\exp(X_{i.}\beta)\right)\right)\\ &+y_i\log(\alpha)+y_i\log\left(\lambda\sum_{\substack{j=1\\ j\neq i}}^{n}w_{ij}y_j+\exp(X_{i.}\beta)\right)\end{aligned} \quad (64)$$

The parameter estimates from the pseudo likelihood for both models can straightforwardly be obtained using standard numerical optimization methods. The asymptotic properties of the estimators as well as their behavior in small samples are investigated in the Monte Carlo study in Section 4.4.

Due to the row-standardization of W, the spatial autocorrelation parameter λ can be interpreted as the absolute change in the conditional expectation of y_i given that the observations of all neighbors of i change by one unit. Accordingly, the change of the observation of a single neighbor j of i causes a change of $E[y_i|Y_{-i},X]$ by λw_{ij}. In spatial models for continuous data the row standardization leads to a parameter range for λ between the smallest eigenvalue of the standardized weights matrix and 1. Since the parameter μ of a Poisson or negative binomial distribution can only take positive values, the spatial autocorrelation parameter of the spatial linear feedback models must additionally fulfill a possibly more restrictive condition ensuring the positiveness of μ:

$$\lambda_i > -\frac{\exp(X_{i.}\beta)}{W_i y} \quad \forall i \quad (65)$$

where W_i and $X_{i.}$ denote the ith row of the respective matrix. The lower bound for λ will therefore be negative, although it might be greater than -1 restricting the strength of negative spatial correlation in the data for which the model is suitable.[8] In the empirical example in Chapter 4.5, however, it is far smaller than -1.

With help of the equality $y = E[y|Y_{-i},X]+\epsilon$, Equation (62) can be reformulated as

$$E[y|Y_{-i},X] = (I-\lambda W)^{-1}\exp(X\beta)+(I-\lambda W)^{-1}\lambda W\epsilon \quad (66)$$

Using this expression it is possible to derive the marginal effects for continuous regressors x_{ik} of the spatial linear feedback model:

$$\frac{\partial\mu_{y_i|X}}{\partial x_{ik}} = a_{ii}\exp(X_{i.}\beta)\beta_k \quad (67)$$

[8] The unlikely case in which the observations of all neighbors of i equal 0 can be ignored since then, λ does not affect its μ.

where a_{ii} is the respective element from $A = (I - \lambda W)^{-1}$ and $X_{i.}$ denotes the i^{th} row of the matrix X. The interpretation of the average of the direct impacts over all units i corresponds to the interpretation of a coefficient in a linear non-spatial model or a marginal effect in a non-linear non-spatial model. It is the average change in the dependent variable i if regressor x_{ik} changes (LeSage and Pace, 2009). But in this spatial model it also includes feedback loops (see Chapter 1.2).

Unlike in non-spatial models, a change of the explanatory variable x_{jk} will affect the dependent variable observation of unit i if unit j is a neighbor or a neighbor's neighbor etc. due to spillover effects and feedback loops (as discussed in Chapter 1.2), i.e.

$$\frac{\partial \mu_{y_i|X}}{\partial x_{jk}} = a_{ij} \exp(X_{j.}\beta)\beta_k \tag{68}$$

The sum of all these effects over all units except i is the indirect impact (or indirect marginal effect) on y_i of a change in the k^{th} regressor in each of the neighboring units:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\partial \mu_{y_i|X}}{\partial x_{jk}} = \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij} \exp(X_{j.}\beta)\beta_k \tag{69}$$

Finally, these two impacts sum up to the total impact or total marginal effect (LeSage and Pace, 2009):

$$\sum_{j=1}^{n} \frac{\partial \mu_{y_i|X}}{\partial x_{jk}} = A_{i.} \exp(X\beta)\beta_k \tag{70}$$

where $A_{i.}$ denotes the i^{th} row of A.

Average marginal effects can be computed in the usual fashion. Standard errors for the marginal effects are again calculated using their sample counterparts of draws from the asymptotic joint distribution of the parameter estimators (LeSage and Pace, 2009, pp. 39).

4.3 Diagnostics

This section introduces two visual diagnostic tools adopted from time series models of counts to evaluate estimation results from the SLF models. Although the general importance of evaluating predictions is obvious, diagnostic tools specially designed for count data are rarely considered in the spatial literature. Together with the scoring rules already used in Chapter 3, this is a first attempt to fill this gap by employing the non-randomized probability integral transform histogram and a relative deviations plot.

The probability integral transform (PIT) which was originally proposed for continuous data by Dawid (1984) and whose idea dates back to Rosenblatt (1952) is an informal way to check the calibration of a model. In the continuous case the PIT is calculated as the values of the predictive cumulative distribution function (CDF) for the observations. If the (continuous) predictive CDF equals the data generating process of the observations the obtained values follow a standard uniform distribution. This can then be checked graphically, for example by plotting a histogram of the PIT values. The advantage of this visual inspection as opposed to a formal test is, that the shape of the histogram also gives guidance as to *why* the uniformity condition is not met (Diebold et al., 1998, p. 869). If the histogram is u-shaped instead of showing a uniform distribution it indicates overdispersion in the data (compared to the predictive distribution) and if it is inverse u-shaped underdispersion in the data is indicated.

For the discrete case, this concept is not directly applicable because the predictive CDF will not be a continuous function. In the case of count data it is a step function meaning that the calculated PIT values will not follow a standard uniform distribution even when DGP and estimation model are identical. Two adaptations of the PIT for time series of counts are discussed in the literature. The first is the randomized PIT (e.g. Denuit and Lambert (2005), Frühwirth-Schnatter (1996), Liesenfeld et al. (2006)), which spreads out the PIT by adding a random term ν weighted with the probability of the observation i.e.

$$u_i^+ = P_{y_i-1} + \nu(P_{y_i} - P_{y_i-1}) \qquad 0 \leq u \leq 1 \tag{71}$$

where P_{y_i} is the predictive CDF of observation y_i, evaluated at y_i. The PIT of these modified observations follows a standard uniform distribution if the predictive CDF is the true data generating process. However, Jung and Tremayne (2011) point out that the shape of the PIT histogram can change considerably between different sets of random terms ν.

The second adaptation of PIT for count data, introduced by Czado et al. (2009), is the nonrandomized PIT which avoids the additional error source of adding a random term. The nonrandomized PIT approach uses the predictive CDF for each observed count y to obtain the distribution of the PIT values directly:

$$F(u|y) = \begin{cases} 0, & u \leq P_{y-1} \\ (u - P_{y-1})/(P_y - P_{y-1}), & P_{y-1} \leq u \leq P_y \\ 1, & u \geq P_y \end{cases} \tag{72}$$

In practice this CDF is evaluated at $u = j/J$ to obtain a chosen number of bins J in the resulting nonrandomized PIT histogram. The n resulting distributions are then averaged and plotted to obtain the histogram. Like in the original case, the PIT histogram should display a uniform distribution if the observations are draws from the predictive distribution.

A method for evaluating the whole predictive distribution graphically is plotting the relative deviations of the estimated probability function and the observed frequencies. Similar plots can be found in Long (1997), for example. This graphic compares the predicted probabilities $\hat{P}_i(k)$ for each predictive distribution $i = 1, \dots, n$ and possible outcome $k = 0, 1, 2, \dots$ with the observed frequencies $h(k)$ in the data set and averages over the n predictive distributions:

$$RelP = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{P}_i(k) - h(k)}{h(k)} \tag{73}$$

with

$$\hat{P}_i(k) = \frac{\hat{\mu}_i^k e^{-\hat{\mu}_i}}{k!} \qquad k = \{0, 1, 2, 3, \ldots\} \tag{74}$$

and $\hat{\mu}_i$ being the estimate for the conditional expectation of observation y_i.
Other than the logarithmic score, this measure takes the whole predictive distribution into account. In addition to the information in the quadratic score and the ranked probability score, which also take into account the whole estimated probability function, the deviation plot gives a visual impression how well the predictive distributions display the different features of the data like mode, tails, etc.

4.4 Monte Carlo Study

4.4.1 Data Generating Process and Study Setup

The data is generated from the conditional model specification in Equations (61)-(62) using a Poisson and negative binomial (2) distribution, respectively, and the Gibbs sampling algorithm of Geman and Geman (1984). That is, by iteratively drawing from the conditional distributions of y_i, $i = 1, \ldots, n$ given all other y_j's draws from the joint distribution of $y_1, \ldots y_n$ are obtained asymptotically. In the k^{th} iteration of this procedure the draws stem from the following distributions:

$$\begin{aligned}
y_1^{(k)} &\sim P(y_1 | y_2^{(k-1)}, \ldots, y_n^{(k-1)}) \\
y_2^{(k)} &\sim P(y_2 | y_1^{(k)}, y_3^{(k-1)}, \ldots, y_n^{(k-1)}) \\
&\vdots \\
y_n^{(k)} &\sim P(y_2 | y_1^{(k)}, \ldots, y_{n-1}^{(k)})
\end{aligned}$$

where P is the conditional probability distribution as specified in Equations (61)-(62) using a Poisson or negative binomial (2) distribution. The start values $y_i^{(0)}$ are drawn from a non-spatial Poisson and negative binomial (2) distribution, respectively, with $\mu_i = \exp(X_i\beta)$. For the negative binomial case the dispersion parameter α is set to 0.2.
An 8-nearest neighbors inverse distance matrix is used for the spatial weighting matrix. To calculate this matrix, first n random coordinates are generated using random numbers from a $U(0, 1)$ distribution. Then, the Euclidian inverse distance matrix for these points, which represent the n spatial units of the simulated data set, is computed using the function `make_nnw` of the spatial econometrics library for MATLAB by LeSage (1999). After selecting the 8 nearest neighbors of each unit, all other entries of the matrix are set to zero and the resulting matrix is finally row-standardized.
The regressor matrix X consists of a constant, $X_1 \sim U(0, 2)$, and $X_2 \sim N(1, 2)$. The parameter vector β is set to $[0.5, 0.5, 0.5]'$. Data is generated for $\lambda = \{0, 0.2, 0.4, 0.6, 0.8\}$ and

$n = \{100, 1000, 5000, 10000\}$. As mentioned in Chapter 3.4, real-life examples of negative spatial autocorrelation are rare in econometrics and therefore this Monte Carlo study focuses on the case of positive spatial autocorrelation.

4.4.2 Monte Carlo Results

The pseudo likelihood estimation introduced above is conducted for both the P-SLFM and the NB-SLFM. The results are displayed in Tables 13 and 14. With exception of the parameter of the constant (β_0) the root mean squared errors and biases for P-SLFM are in general small. The estimates of the spatial autocorrelation parameter λ are biased downward though, especially in the case of a small sample and a weak spatial autocorrelation in the data.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	**100**	**1000**	**5000**	**10000**	**100**	**1000**	**5000**	**10000**
RMSE								
0	0.035	0.010	0.004	0.002	0.149	0.047	0.019	0.012
0.2	0.045	0.013	0.006	0.004	0.188	0.050	0.024	0.016
0.4	0.054	0.015	0.006	0.004	0.255	0.064	0.028	0.019
0.6	0.040	0.013	0.006	0.004	0.243	0.081	0.035	0.024
0.8	0.023	0.009	0.005	0.004	0.299	0.102	0.045	0.033
Bias								
0	−0.005	−0.001	0.000	0.000	0.005	0.002	−0.001	0.000
0.2	−0.005	−0.002	−0.001	−0.001	0.009	0.004	0.004	0.004
0.4	−0.008	−0.002	−0.001	−0.001	0.009	0.006	0.005	0.006
0.6	−0.004	−0.001	−0.001	−0.001	0.000	0.007	0.006	0.005
0.8	−0.004	−0.003	−0.003	−0.003	−0.019	−0.003	0.003	0.003
Relative bias								
0	-	-	-	-	0.009	0.004	0.002	0.000
0.2	0.026	0.008	0.005	0.005	0.018	0.008	0.008	0.008
0.4	0.019	0.004	0.002	0.003	0.017	0.012	0.009	0.011
0.6	0.007	0.002	0.002	0.001	0.000	0.014	0.012	0.011
0.8	0.005	0.004	0.004	0.004	0.038	0.006	0.005	0.006
	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	**100**	**1000**	**5000**	**10000**	**100**	**1000**	**5000**	**10000**
RMSE								
0	0.069	0.021	0.009	0.006	0.027	0.008	0.003	0.002
0.2	0.084	0.022	0.011	0.007	0.031	0.008	0.004	0.003
0.4	0.079	0.026	0.011	0.008	0.049	0.010	0.005	0.003
0.6	0.091	0.031	0.013	0.010	0.043	0.012	0.005	0.004
0.8	0.107	0.039	0.017	0.012	0.042	0.013	0.006	0.005
Bias								
0	0.003	0.000	0.001	0.000	−0.002	0.000	0.000	0.000
0.2	−0.001	−0.001	−0.001	−0.001	−0.002	−0.001	−0.001	−0.001
0.4	−0.001	−0.001	−0.001	−0.001	−0.002	−0.001	−0.001	−0.001
0.6	0.002	−0.001	−0.002	−0.001	−0.001	−0.001	−0.001	−0.001
0.8	0.003	0.002	−0.001	−0.001	0.002	0.000	0.000	0.000
Relative bias								
0	0.007	0.000	0.001	0.000	0.005	0.001	0.000	0.000
0.2	0.003	0.001	0.002	0.001	0.004	0.002	0.001	0.001
0.4	0.002	0.003	0.002	0.002	0.003	0.002	0.002	0.002
0.6	0.005	0.002	0.003	0.001	0.001	0.002	0.002	0.002
0.8	0.006	0.003	0.002	0.002	0.005	0.000	0.001	0.001

Table 13: Monte Carlo results for P-SLFM. The bias is calculated as the average difference between estimates and true parameter value. The relative bias is the absolute value of the bias divided by the true parameter value. Values reported as 0.000 are smaller than 0.0005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.044	0.009	0.004	0.003	0.216	0.062	0.027	0.021
0.2	0.087	0.023	0.012	0.010	0.349	0.107	0.059	0.047
0.4	0.090	0.027	0.014	0.012	0.399	0.134	0.077	0.074
0.6	0.105	0.030	0.013	0.010	0.823	0.197	0.104	0.090
0.8	0.077	0.029	0.013	0.010	2.641	0.324	0.168	0.132
Bias								
0	−0.007	−0.001	0.000	0.000	0.006	0.000	−0.001	0.000
0.2	−0.023	−0.009	−0.008	−0.008	0.058	0.043	0.042	0.038
0.4	−0.025	−0.009	−0.007	−0.008	0.069	0.054	0.054	0.060
0.6	−0.024	−0.008	−0.004	−0.005	0.066	0.086	0.069	0.071
0.8	−0.014	−0.009	−0.006	−0.006	−0.175	0.101	0.103	0.095
Relative bias								
0	-	-	-	-	0.013	0.000	−0.002	0.000
0.2	−0.113	−0.043	−0.040	−0.038	0.116	0.085	0.084	0.077
0.4	−0.062	−0.023	−0.017	−0.021	0.139	0.107	0.108	0.120
0.6	−0.040	−0.013	−0.007	−0.008	0.132	0.171	0.137	0.143
0.8	−0.017	−0.012	−0.007	−0.007	−0.350	0.201	0.205	0.190
	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.116	0.036	0.016	0.011	0.047	0.014	0.006	0.005
0.2	0.165	0.046	0.022	0.016	0.068	0.021	0.012	0.009
0.4	0.202	0.056	0.027	0.022	0.075	0.026	0.015	0.014
0.6	0.351	0.076	0.036	0.027	0.116	0.036	0.019	0.018
0.8	0.986	0.128	0.061	0.043	0.325	0.057	0.030	0.024
Bias								
0	0.004	0.002	0.001	0.000	−0.005	−0.001	0.000	0.000
0.2	−0.015	−0.009	−0.008	−0.007	−0.013	−0.008	−0.008	−0.008
0.4	−0.017	−0.007	−0.010	−0.011	−0.013	−0.012	−0.011	−0.012
0.6	−0.010	−0.018	−0.015	−0.015	−0.023	−0.018	−0.014	−0.015
0.8	0.049	−0.015	−0.022	−0.019	−0.008	−0.023	−0.021	−0.019
Relative bias								
0	0.008	0.005	0.001	0.000	−0.010	−0.002	0.000	0.000
0.2	−0.030	−0.018	−0.016	−0.014	−0.027	−0.015	−0.017	−0.015
0.4	−0.034	−0.014	−0.021	−0.023	−0.026	−0.023	−0.022	−0.024
0.6	−0.019	−0.036	−0.030	−0.031	−0.046	−0.036	−0.028	−0.030
0.8	0.099	−0.031	−0.045	−0.038	−0.017	−0.045	−0.042	−0.038

Table 14: Monte Carlo results for NB-SLFM. The bias is calculated as the average difference between estimates and true parameter value. The relative bias is the absolute value of the bias divided by the true parameter value. Values reported as 0.000 are smaller than 0.0005.

For a better perception of the size of the average biases the relative bias is also reported, which is the absolute value of the bias divided by the true parameter value. The relative bias of λ ranges from 0.2% to 2.6% with the highest values associated with small sample size and weak spatial autocorrelation. β_1 and β_2 show relative biases smaller than 1% throughout the study. Only the parameter of the constant β_0 shows higher relative biases which lie between 0.2% and 3.8%.

Generally, the results for the NB-SLFM are worse than for the P-SLFM, which is not surprising since the P-SLFM is the more parsimonious model. In the results for the NB-SLFM the constant β_0 shows again the largest RMSE and bias. The estimates of λ are biased downward like in the P-SLFM. The relative bias of λ ranges from 0.7% to 11.3% with the largest values belonging again to estimations of data with a small sample size and a small true value of λ.

$\lambda\backslash n$	100		1000		5000		10000	
	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
Bias of average total marginal effect								
0	0.029	−0.015	−0.001	−0.003	0.005	0.001	0.001	0.000
0.2	−0.012	−0.015	−0.008	−0.011	−0.010	−0.006	−0.007	−0.006
0.4	−0.016	−0.024	−0.021	−0.012	−0.009	−0.009	−0.012	−0.011
0.6	0.047	−0.005	−0.024	−0.021	−0.030	−0.017	−0.011	−0.018
0.8	−0.237	−0.231	−0.216	−0.276	−0.299	−0.286	−0.300	−0.284
Relative bias of average total marginal effect								
0	0.008	−0.004	0.000	−0.001	0.001	0.000	0.000	0.000
0.2	−0.003	−0.003	−0.002	−0.002	−0.002	−0.001	−0.001	−0.001
0.4	−0.003	−0.004	−0.003	−0.002	−0.001	−0.001	−0.002	−0.002
0.6	0.005	−0.001	−0.003	−0.002	−0.003	−0.002	−0.001	−0.002
0.8	−0.010	−0.010	−0.011	−0.014	−0.016	−0.015	−0.016	−0.015
Bias of average direct marginal effect								
0	0.042	−0.002	0.003	0.001	0.004	0.000	0.001	0.001
0.2	0.004	0.003	−0.001	−0.003	−0.004	0.000	−0.001	0.000
0.4	0.015	0.003	−0.004	0.001	0.000	0.000	−0.001	0.000
0.6	0.043	0.010	−0.001	0.001	−0.006	0.000	0.002	−0.001
0.8	0.004	−0.002	−0.001	−0.015	−0.020	−0.017	−0.018	−0.014
Relative bias of average direct marginal effect								
0	0.011	0.000	0.001	0.000	0.001	0.000	0.000	0.000
0.2	0.001	0.001	0.000	−0.001	−0.001	0.000	0.000	0.000
0.4	0.004	0.001	−0.001	0.000	0.000	0.000	0.000	0.000
0.6	0.011	0.003	0.000	0.000	−0.001	0.000	0.000	0.000
0.8	0.001	0.000	0.000	−0.003	−0.004	−0.004	−0.004	−0.003
Bias of average indirect marginal effect								
0	−0.013	−0.013	−0.004	−0.004	0.000	0.000	0.000	0.000
0.2	−0.015	−0.018	−0.008	−0.008	−0.006	−0.005	−0.006	−0.006
0.4	−0.031	−0.028	−0.017	−0.014	−0.008	−0.008	−0.011	−0.011
0.6	0.004	−0.015	−0.023	−0.022	−0.024	−0.017	−0.013	−0.016
0.8	−0.241	−0.229	−0.215	−0.261	−0.279	−0.270	−0.282	−0.270
Relative bias of average indirect marginal effect								
0	-	-	-	-	-	-	-	-
0.2	−0.018	−0.021	−0.009	−0.009	−0.007	−0.006	−0.006	−0.006
0.4	−0.014	−0.012	−0.007	−0.006	−0.003	−0.003	−0.005	−0.005
0.6	0.001	−0.003	−0.004	−0.004	−0.005	−0.003	−0.002	−0.003
0.8	−0.013	−0.013	−0.014	−0.017	−0.019	−0.019	−0.020	−0.019

Table 15: Monte Carlo results of median marginal effects for P-SLFM data. The bias is calculated as the average difference between estimates and true value. The relative bias is the absolute value of the bias divided by the true value. Values reported as 0.000 are smaller than 0.0005.

Overall, the estimations show satisfactory results for all sample sizes. The Monte Carlo results support the arguments of Besag (1975) regarding the consistency of the estimator, even though there seems to be a small downward bias in the estimates of λ for the sample sizes considered here. With regard to the empirical application which will be presented in Chapter 4.5 the simulation shows that for a sample size of about 3000 observations the estimation based on the presented pseudo likelihoods works well.

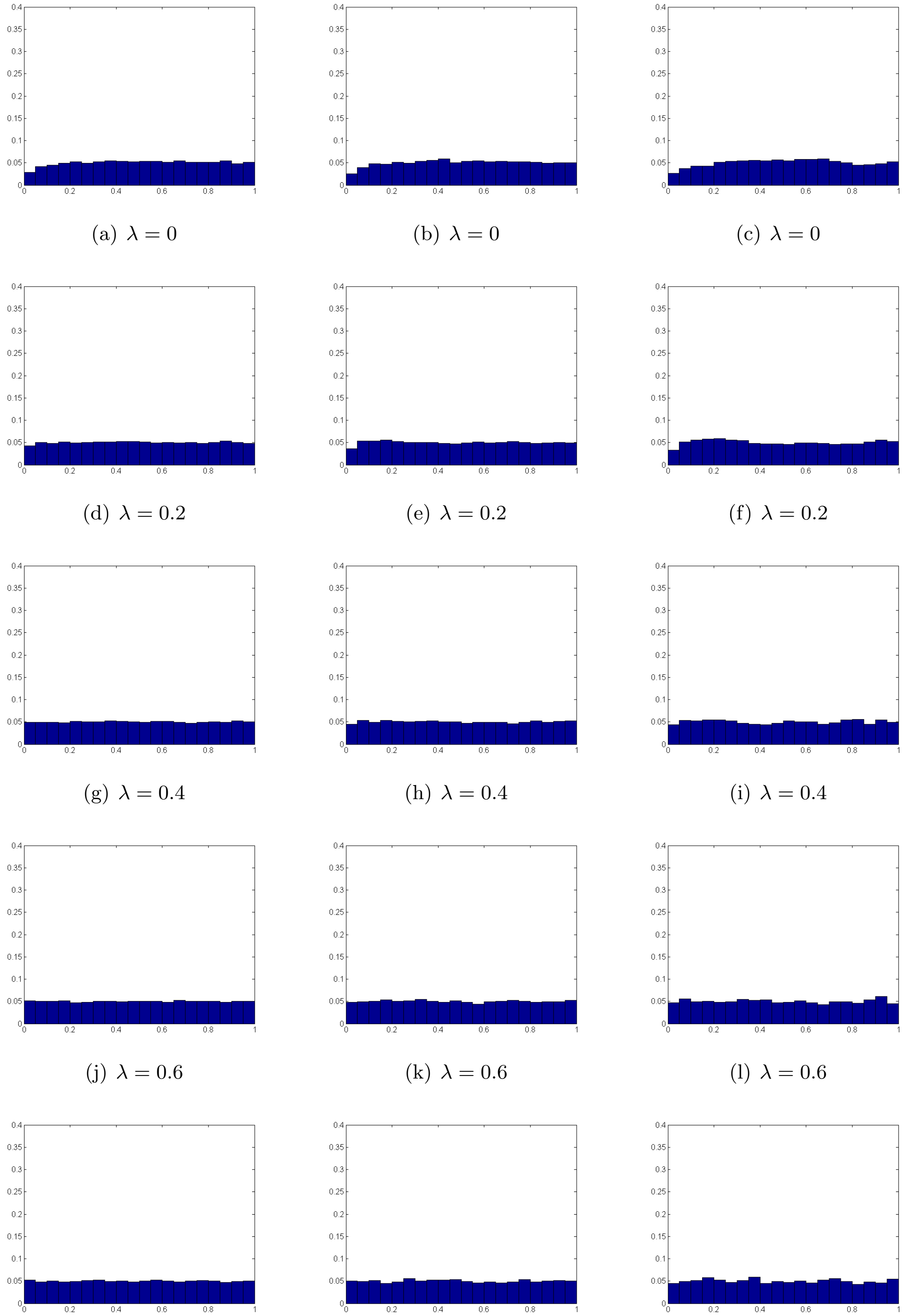
In the following, additional Monte Carlo results are presented. First, the performance of the spatial linear feedback models with regard to marginal effects is evaluated. Then, the suitability of PIT histograms as a visual investigation tool of model fit is investigated. And finally, simulation results of several specification alternatives are reported to show their influence on the results.

$\lambda\backslash n$	100		1000		5000		10000	
	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
Bias of average total marginal effect								
0	0.048	−0.018	0.021	−0.007	0.008	0.004	−0.001	−0.002
0.2	−0.101	−0.087	−0.084	−0.072	−0.090	−0.093	−0.080	−0.085
0.4	−0.145	−0.107	−0.092	−0.151	−0.139	−0.146	−0.155	−0.166
0.6	−0.043	−0.268	−0.325	−0.324	−0.311	−0.292	−0.309	−0.304
0.8	2.338	0.489	−0.674	−0.948	−1.094	−1.027	−0.943	−0.958
Relative bias of average total marginal effect								
0	0.013	−0.005	0.005	−0.002	0.002	0.001	0.000	0.000
0.2	−0.025	−0.022	−0.017	−0.015	−0.019	−0.019	−0.016	−0.017
0.4	−0.023	−0.017	−0.015	−0.024	−0.022	−0.023	−0.025	−0.026
0.6	−0.005	−0.032	−0.036	−0.036	−0.032	−0.030	−0.032	−0.032
0.8	0.095	0.020	−0.036	−0.051	−0.056	−0.053	−0.048	−0.049
Bias of average direct marginal effect								
0	0.070	−0.003	0.025	−0.003	0.009	0.005	−0.001	−0.002
0.2	−0.036	−0.020	−0.031	−0.021	−0.036	−0.039	−0.030	−0.034
0.4	0.006	0.010	−0.008	−0.046	−0.048	−0.052	−0.050	−0.057
0.6	−0.072	−0.078	−0.089	−0.087	−0.100	−0.091	−0.097	−0.095
0.8	0.689	−0.069	−0.055	−0.125	−0.177	−0.158	−0.134	−0.138
Relative bias of average direct marginal effect								
0	0.018	−0.001	0.006	−0.001	0.002	0.001	0.000	0.000
0.2	−0.011	−0.006	−0.008	−0.005	−0.009	−0.010	−0.008	−0.009
0.4	0.001	0.003	−0.002	−0.012	−0.012	−0.013	−0.013	−0.015
0.6	−0.020	−0.022	−0.023	−0.022	−0.023	−0.021	−0.023	−0.023
0.8	0.106	−0.011	−0.012	−0.027	−0.036	−0.033	−0.027	−0.028
Bias of average indirect marginal effect								
0	−0.022	−0.015	−0.004	−0.004	0.000	0.000	0.000	0.000
0.2	−0.065	−0.067	−0.053	−0.051	−0.053	−0.054	−0.050	−0.051
0.4	−0.151	−0.117	−0.084	−0.105	−0.091	−0.094	−0.105	−0.109
0.6	0.029	−0.190	−0.236	−0.237	−0.211	−0.201	−0.211	−0.209
0.8	1.649	0.558	−0.619	−0.823	−0.917	−0.869	−0.809	−0.821
Relative bias of average indirect marginal effect								
0	-	-	-	-	-	-	-	-
0.2	−0.090	−0.093	−0.056	−0.054	−0.057	−0.058	−0.053	−0.054
0.4	−0.065	−0.051	−0.035	−0.044	−0.038	−0.039	−0.044	−0.045
0.6	0.006	−0.040	−0.046	−0.046	−0.039	−0.037	−0.039	−0.039
0.8	0.091	0.031	−0.044	−0.059	−0.063	−0.059	−0.055	−0.056

Table 16: Monte Carlo results of median marginal effects for NB-SLFM data. The bias is calculated as the average difference between estimates and true value. The relative bias is the absolute value of the bias divided by the true value. Values reported as 0.000 are smaller than 0.0005.

Marginal effects according to the Monte Carlo estimates are calculated using Equations (67) to (70) introduced above. Tables 15 and 16 display bias and relative bias of the total, direct, and indirect marginal effects for P-SLFM and NB-SLFM. Like for the parameter estimates, the biases are generally small in the P-SLF model. The indirect effect is in most cases slightly underestimated, whereas the direct effect of the change in the regressor on the dependent variable of the same unit tends to be overestimated. This leads in sum to a negatively biased total effect in most cases.

The biases of the marginal effects in the NB-SLF model are larger in absolute terms which is also the case for the parameters. In addition to the indirect effects, the direct effects of the NB-SLFM are underestimated in most cases as well. This leads to larger total biases in absolute terms. To sum up, the results of the Monte Carlo study document the ability of the SLF models to correctly estimate the marginal effects of exogenous regressors.



(a) $\lambda = 0$ (b) $\lambda = 0$ (c) $\lambda = 0$

(d) $\lambda = 0.2$ (e) $\lambda = 0.2$ (f) $\lambda = 0.2$

(g) $\lambda = 0.4$ (h) $\lambda = 0.4$ (i) $\lambda = 0.4$

(j) $\lambda = 0.6$ (k) $\lambda = 0.6$ (l) $\lambda = 0.6$

(m) $\lambda = 0.8$ (n) $\lambda = 0.8$ (o) $\lambda = 0.8$

Figure 12: PIT histograms for Monte Carlo estimates from the NB-SLFM. n=5000.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.074	0.011	0.004	0.003	0.301	0.079	0.035	0.024
0.2	0.086	0.033	0.019	0.017	0.301	0.147	0.099	0.092
0.4	0.112	0.044	0.022	0.020	0.525	0.205	0.145	0.143
0.6	0.161	0.045	0.020	0.016	1.077	0.307	0.190	0.183
0.8	0.160	0.044	0.022	0.018	14.498	0.503	0.310	0.281
Bias								
0	−0.021	−0.001	0.000	0.000	0.044	0.000	0.000	0.000
0.2	−0.018	−0.017	−0.015	−0.015	0.049	0.079	0.082	0.083
0.4	−0.034	−0.016	−0.015	−0.017	0.161	0.109	0.123	0.133
0.6	−0.048	−0.016	−0.010	−0.010	0.110	0.176	0.157	0.166
0.8	−0.052	−0.021	−0.013	−0.013	−1.450	0.300	0.256	0.252
Relative bias								
0	-	-	-	-	0.087	0.001	0.001	0.001
0.2	−0.088	−0.086	−0.075	−0.073	0.097	0.157	0.164	0.166
0.4	−0.085	−0.041	−0.036	−0.042	0.323	0.218	0.247	0.266
0.6	−0.080	−0.026	−0.016	−0.017	0.220	0.352	0.314	0.332
0.8	−0.065	−0.026	−0.016	−0.016	−2.900	0.600	0.512	0.504

	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.171	0.050	0.022	0.016	0.076	0.018	0.008	0.006
0.2	0.189	0.061	0.033	0.025	0.075	0.031	0.020	0.019
0.4	0.259	0.082	0.045	0.038	0.096	0.040	0.028	0.029
0.6	0.435	0.114	0.058	0.050	0.159	0.059	0.038	0.037
0.8	15.898	0.180	0.090	0.072	2.142	0.098	0.060	0.055
Bias								
0	−0.005	0.001	0.001	0.000	−0.013	0.000	0.000	0.000
0.2	−0.009	−0.014	−0.018	−0.016	−0.016	−0.017	−0.017	−0.017
0.4	−0.028	−0.026	−0.028	−0.029	−0.034	−0.023	−0.024	−0.027
0.6	−0.015	−0.043	−0.034	−0.038	−0.030	−0.038	−0.033	−0.035
0.8	−0.487	−0.056	−0.051	−0.049	0.187	−0.066	−0.052	−0.050
Relative bias								
0	−0.009	0.001	0.001	0.000	−0.025	−0.001	0.000	0.000
0.2	−0.019	−0.027	−0.035	−0.032	−0.032	−0.033	−0.034	−0.034
0.4	−0.055	−0.052	−0.056	−0.058	−0.067	−0.045	−0.049	−0.054
0.6	−0.029	−0.087	−0.068	−0.076	−0.059	−0.076	−0.066	−0.069
0.8	−0.974	−0.112	−0.102	−0.098	0.374	−0.132	−0.105	−0.101

Table 17: Monte Carlo results for NB-SLFM, $\alpha = 0.5$. The bias is calculated as the average difference between estimates and true parameter value. The relative bias is the absolute value of the bias divided by the true parameter value. Values reported as 0.000 are smaller than 0.0005.

The appropriateness of a visual inspection of the PIT histograms for comparing the fit of different SLF models is also evaluated using the Monte Carlo estimates. PIT histograms are calculated for each iteration of the Monte Carlo study. To give an impression of the results for the SLF models, three representative PIT histograms for each variant of $\lambda = \{0, 0.2, 0.4, 0.6, 0.8\}$ and $n = 5000$ are displayed.

Figure 20 in Appendix B.1 and Figure 12 show the PIT histograms of the estimations of P-SLFM and NB-SLFM data with the respective models. The histograms are close to uniformity for the majority of simulated samples which has been expected since DGP and estimated model are identical here. Figure 21 in Appendix B.1 shows the PIT histograms of estimations of NB-SLFM data with the P-SLFM model. As expected the histograms exhibit a u-form, indicating that the dispersion in the data is higher than in the estimated model. Since the overdispersion in the generated data is a function of the intensity, which increases with increasing λ given an otherwise unchanged DGP, the fit also gets worse with

			$\hat{\lambda}$				$\hat{\beta}_0$	
$\lambda \backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.046	0.012	0.005	0.004	0.275	0.091	0.039	0.029
0.2	0.091	0.024	0.012	0.010	0.421	0.118	0.062	0.048
0.4	0.094	0.028	0.015	0.011	0.477	0.151	0.085	0.062
0.6	0.102	0.030	0.014	0.010	1.699	0.201	0.109	0.082
0.8	0.091	0.029	0.013	0.011	4.571	0.350	0.171	0.124
Bias								
0	−0.004	−0.001	0.000	0.000	0.001	−0.001	−0.002	0.001
0.2	−0.018	−0.007	−0.007	−0.007	0.025	0.025	0.026	0.024
0.4	−0.021	−0.006	−0.006	−0.007	0.041	0.020	0.032	0.038
0.6	−0.016	−0.008	−0.004	−0.004	−0.044	0.052	0.034	0.041
0.8	−0.018	−0.008	−0.006	−0.006	−0.635	0.037	0.061	0.051
Relative bias								
0	-	-	-	-	0.001	−0.002	−0.005	0.002
0.2	−0.092	−0.036	−0.034	−0.033	0.050	0.050	0.053	0.048
0.4	−0.054	−0.016	−0.016	−0.018	0.082	0.040	0.064	0.076
0.6	−0.027	−0.013	−0.007	−0.007	−0.088	0.104	0.068	0.082
0.8	−0.023	−0.010	−0.008	−0.008	−1.270	0.075	0.122	0.103
			$\hat{\beta}_1$				$\hat{\beta}_2$	
$\lambda \backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.150	0.047	0.020	0.014	0.058	0.020	0.009	0.007
0.2	0.206	0.052	0.027	0.018	0.079	0.024	0.013	0.009
0.4	0.241	0.067	0.030	0.022	0.087	0.031	0.017	0.011
0.6	0.604	0.083	0.041	0.028	0.424	0.036	0.019	0.015
0.8	2.523	0.134	0.065	0.044	0.319	0.063	0.028	0.021
Bias								
0	0.005	0.003	0.001	0.000	−0.007	−0.001	0.001	0.000
0.2	−0.005	−0.005	−0.004	−0.003	−0.008	−0.004	−0.005	−0.004
0.4	−0.015	0.001	−0.006	−0.006	−0.008	−0.004	−0.005	−0.007
0.6	−0.048	−0.010	−0.007	−0.008	0.008	−0.010	−0.006	−0.008
0.8	0.159	−0.001	−0.013	−0.008	0.033	−0.009	−0.011	−0.009
Relative bias								
0	0.009	0.005	0.002	−0.001	−0.013	−0.002	0.001	−0.001
0.2	−0.011	−0.009	−0.008	−0.007	−0.016	−0.007	−0.009	−0.008
0.4	−0.030	0.002	−0.012	−0.013	−0.017	−0.008	−0.010	−0.013
0.6	−0.096	−0.020	−0.013	−0.015	0.016	−0.019	−0.012	−0.015
0.8	0.319	−0.003	−0.027	−0.016	0.065	−0.018	−0.021	−0.018

Table 18: Monte Carlo results for NB-SLFM data estimated with P-SLFM model, $\alpha = 0.2$. The bias is calculated as the average difference between estimates and true parameter value. The relative bias is the absolute value of the bias divided by the true parameter value. Values reported as 0.000 are smaller than 0.0005.

increasing λ as visible in the PIT histograms in Figure 21. These results support that the visual inspection of the PIT histograms works fine for the spatial linear feedback models with regard to a possible misspecification of the dispersion.

Finally, several variations of the standard specifications are presented to investigate the robustness of the estimation. They cover variation in the dispersion parameter α, misspecification due to ignoring the overdispersion in the data, the choice of a different spatial weighting matrix in the data generating process and estimation model, and the increase of the exogenous regressors' influence on the dependent variable.

Increasing the dispersion parameter in the DGP from 0.2 to 0.5 leads to more dispersion in the generated data. Table 17 shows that both the RMSE and the bias are larger in the presence of larger dispersion, especially for small sample sizes and high values of λ. This points to the fact that small sample sizes are more sensitive to outliers, which are more likely if the dispersion is larger, and which enter the intensity equation of the NB-SFLM directly through the spatial term Wy. For sample sizes of at least 5000, estimation works well, nevertheless.

Table 18 contains the simulation results of a P-SLFM estimation with NB-SLFM data (according to the DGP in Section 4.4.1). As noted in the previous chapter, a correctly specified conditional mean function leads to consistent estimates in a non-spatial count data model (Gourieroux et al., 1984) despite misspecification of the conditional variance. Here, the results of the estimates for β support this only for small values of λ. For large values of the true spatial autocorrelation parameter RMSE and bias are considerably larger, especially for the constant β_0. This is surprising regarding the contrary results for the P-SAR model in Chapter 3.4 and indicates a sensitivity to misspecification of the dispersion. Through the neighbor's outcomes of the dependent variable, which enter the intensity equation directly, the conditional expectation contains a part which is influenced by the dispersion, e.g. due to a larger number of outliers. This might in turn lead to less precise estimates and slower convergence rates.

Since the choice of a spatial weight matrix is arbitrary, Table 67 (in Appendix B.2) displays the results for P-SLFM data with a contiguity matrix used in the DGP and in the estimation instead of the 8 nearest neighbors inverse distance matrix used so far. The contiguity matrix is calculated and row-standardized using the function `xy2cont` of the Spatial Statistics Toolbox for MATLAB 2.0 (Pace, 2003) on randomly generated coordinates. Not surprisingly, the RMSE and bias do not differ in magnitude from the ones of the standard specification presented in Table 13. This supports that to the general applicability of the spatial linear feedback models does not depend on the choice of the spatial weight matrices.
Finally, the RMSE and bias of estimation of a P-SLF model with true β coefficients values of 1.5 instead of 0.5 are shown in Table 68 (in Appendix B.2). The large coefficients in combination with the otherwise unchanged DGP lead to larger values of conditional mean and therefore potentially larger values of the dependent variable. This also means an increase of the variation in the data since conditional mean and variance are equal in a Poisson model. All values of RMSE and bias are much smaller than in the standard specification (Table 13) which is not surprising since larger variation in the data, i.e. more informative data, generally leads to more precise estimates.

4.5 Empirical Application: Start-up Firm Births

Using the data from Lambert et al. (2010) on firm births in the United States (see Section 3.6.1), an empirical application of the spatial linear feedback model is conducted in this section. Again, an 8 nearest neighbors matrix is used for the spatial weight matrix which equals the weight matrix used by Lambert et al. (2010). To calculate this matrix the Euclidian inverse distance matrix for the counties is computed first, then the 8 nearest neighbors of each observation are selected, setting all other entries of the matrix equal to zero and finally the resulting matrix is row-standardized.

Table 19 displays the estimation results for the P-SLF and NB-SLF models discussed in Section 4.2 and, for comparison, the non-spatial Poisson and non-spatial NB regression whose conditional expectation is given by Equation (16). In the case of the NB model, the so called negative binomial 2 specification is chosen which exhibits a quadratic variance function. In both spatial models the spatial autocorrelation parameter λ is highly significant, indicating that the proposed spatial structure is present in the data, but it is considerably different in size. Also several other parameter estimates (and marginal effects, respectively) show clear differences between the P-SLF and NB-SLF results. Both the log score and the quadratic score clearly favor the negative binomial models and give the spatial variant a slight preference. The ranked probability score in contrast prefers the spatial Poisson model because it highly penalizes the relatively flat estimated distributions from the negative binomial models. All three scores provide evidence in favor of the spatial model specification, which is also supported by the high statistical significance of λ.

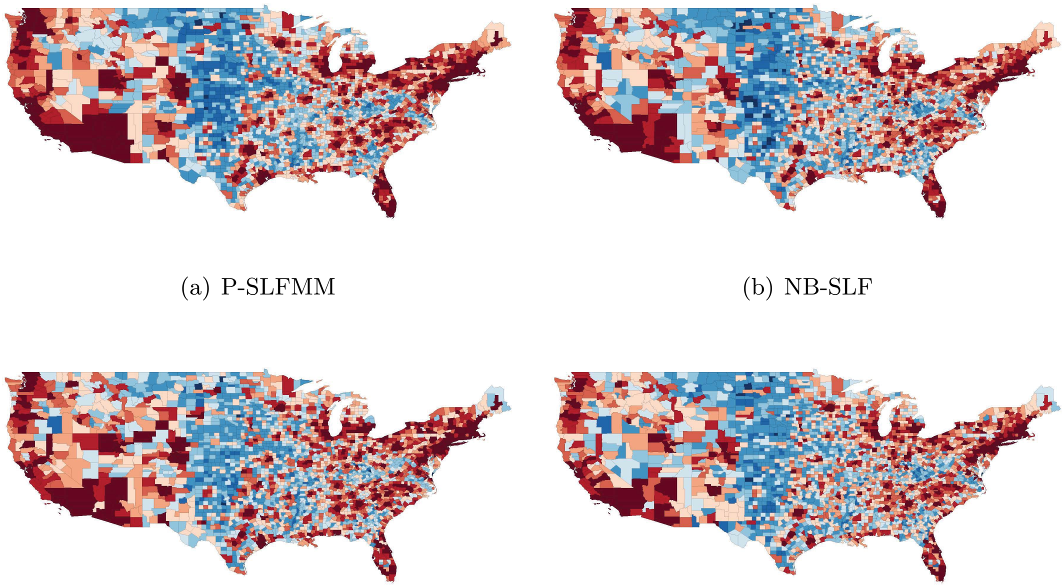
To investigate the robustness of these results with regard to the choice of the spatial weighting matrix, estimations using a contiguity matrix (W_{con}) and an 8 nearest neighbors matrix (W_{nn}), which weights all neighbors equally (instead of taking the distance to the neighbors into account), are conducted additionally. The matrices are calculated from the center's coordinates of the U.S. counties using the already mentioned MATLAB functions `xy2cont` and `make_nnw` (LeSage, 1999). Table 69 (in Appendix B.3) gives the parameter estimates and scores of the P-SLFM and NB-SLFM estimations with these weight matrices which do not differ substantially from the previous results using an 8 nearest neighbors inverse distance matrix in Table 19. The scores indicate a similar, some a slightly worse, fit than the one obtained with the 8 nearest neighbors matrix (Table 19).

To get a first visual impression of the models' fit, Figure 13 shows the estimated conditional expectations of each county for the four models. The classes equal the deciles of the observations displayed in Figure 9 going from dark blue (0 firm births) to dark red (60 to 6938 firm births). On all four maps the clusters of high numbers of firm births at west and east coast as well as at the Great Lakes are recognizable as well as the cluster of lower values in the Midwest. The spatial models, however, predict higher values for many counties and reproduces the map of the observations more properly. Nevertheless, all models tend to overestimate the number of start-up firms in counties with observation 0 (dark blue in Figure 9).

The nonrandomized PIT is calculated to visually compare the calibration of the models and

	SLFM		**Non-spatial**	
Variable	**Poisson**	**NB**	**Poisson**	**NB**
λ	0.288***	0.166***		
	(0.043)	(0.022)		
const	−1.707***	−1.120***	−0.934***	−1.066***
	(0.397)	(0.249)	(0.281)	(0.195)
msemp	0.035***	0.053***	0.031***	0.050***
	(0.006)	(0.003)	(0.004)	(0.002)
pelt10	−0.007**	0.005***	−0.002	0.005***
	(0.003)	(0.001)	(0.002)	(0.001)
pemt100	−0.034***	−0.023***	−0.029***	−0.018***
	(0.006)	(0.003)	(0.004)	(0.002)
tfdens	−0.013	−0.046***	0.006	−0.053***
	(0.011)	(0.016)	(0.010)	(0.013)
mhhi	−0.034***	0.024**	0.000	0.027***
	(0.008)	(0.010)	(0.009)	(0.005)
pop	0.002***	0.017***	0.002***	0.018***
	(0.000)	(0.003)	(0.000)	(0.003)
cclass	0.088***	0.101***	0.048***	0.082***
	(0.011)	(0.007)	(0.013)	(0.005)
uer	0.037	0.076***	0.073***	0.080***
	(0.037)	(0.021)	(0.022)	(0.013)
pedas	0.150***	0.062***	0.130***	0.044***
	(0.022)	(0.011)	(0.021)	(0.009)
awage	0.033***	−0.058***	0.019***	−0.038***
	(0.008)	(0.012)	(0.007)	(0.007)
netflow	0.003	−0.027	0.002	−0.016***
	(0.003)	(0.006)	(0.820)	(0.003)
proad	0.093***	0.084***	0.103***	0.083***
	(0.023)	(0.024)	(0.018)	(0.022)
interst	0.009***	0.005***	0.007***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
avland	−0.007***	−0.006***	−0.009***	−0.007***
	(0.002)	(0.001)	(0.001)	(0.001)
bci	0.128***	0.032	0.080	0.034
	(0.042)	(0.021)	(0.037)	(0.015)
educpc	0.006***	0.006**	0.004	0.004
	(0.002)	(0.003)	(0.002)	(0.003)
hwypc	−0.039	−0.132	−0.030	−0.028
	(0.023)	(0.031)	(0.019)	(0.021)
metro	1.630***	1.017***	1.265***	0.845***
	(0.157)	(0.081)	(0.092)	(0.054)
micro	0.839***	0.645***	0.573***	0.546***
	(0.119)	(0.055)	(0.063)	(0.038)
α		0.403***		0.437***
		(0.026)		(0.024)
Log L	−28149	−10300	−32248	−10401
LogS	9.002	3.348	9.917	3.379
QS	−0.027	−0.073	−0.017	−0.070
RPS	14.035	22.836	15.244	37.858

Table 19: Estimation results from P-SLF, NB-SLF, non-spatial Poisson, and non-spatial NB models for the start-up firm births data. N=3078. Heteroscedasticity robust standard errors in brackets, calculated with the sandwich formula (White, 1980), ** and *** denote a 5% and 1% significance, respectively. Values reported as 0.000 are smaller than 0.0005.



(a) P-SLFMM

(b) NB-SLF

(c) Non-spatial Poisson

(d) Non-spatial NB

Figure 13: Estimated conditional expectations of the number of start-up firm births with dark blue representing the lowest values and dark red the highest ones. The color scheme corresponds to the size classes used in Figure 9, i.e. deciles of the observed data.

to get an indication whether the dispersion allowed by the model is suitable for the data. Figure 14 displays the PIT histograms for the four estimated models. The u-shaped PIT histograms of both Poisson models show clear sign of too little dispersion allowed by the model. The PIT histograms of both negative binomial models have an even shape, but the NB-SLFM shows a decrease for PIT values larger than 0.8. This might be caused by the extreme outliers in the data (> 6000) and at the same time relatively high frequencies of small values (< 3). Single index models as used here generally have difficulties to cope with such a data structure. To sum up, the PIT histograms clearly advocate for the negative binomial models.

A further way to look at the fit of the model is to predict the probabilities of each possible outcome and to compare these visually with the empirical frequencies in a relative deviations plot. The relative differences of predictive probabilities and observed frequency of the four models are displayed in Figure 15. Generally, the two negative binomial models outperform the Poisson models. Small values are underestimated by the Poisson models, large values overestimated by all models. For outcomes between 5 and 25 only small differences between frequencies and predicted distributions of the negative binomial models and the P-SLF model can be observed, whereas the non-spatial Poisson model fits the data less well.

For estimating the impact of the regressors on the number of firm births the marginal effects have to be computed. Table 20 displays the median marginal effects for all four models. The median is reported, rather than the mean, since the data set contains some unusually large observations of the dependent variable (at least in the context of count data). Because of

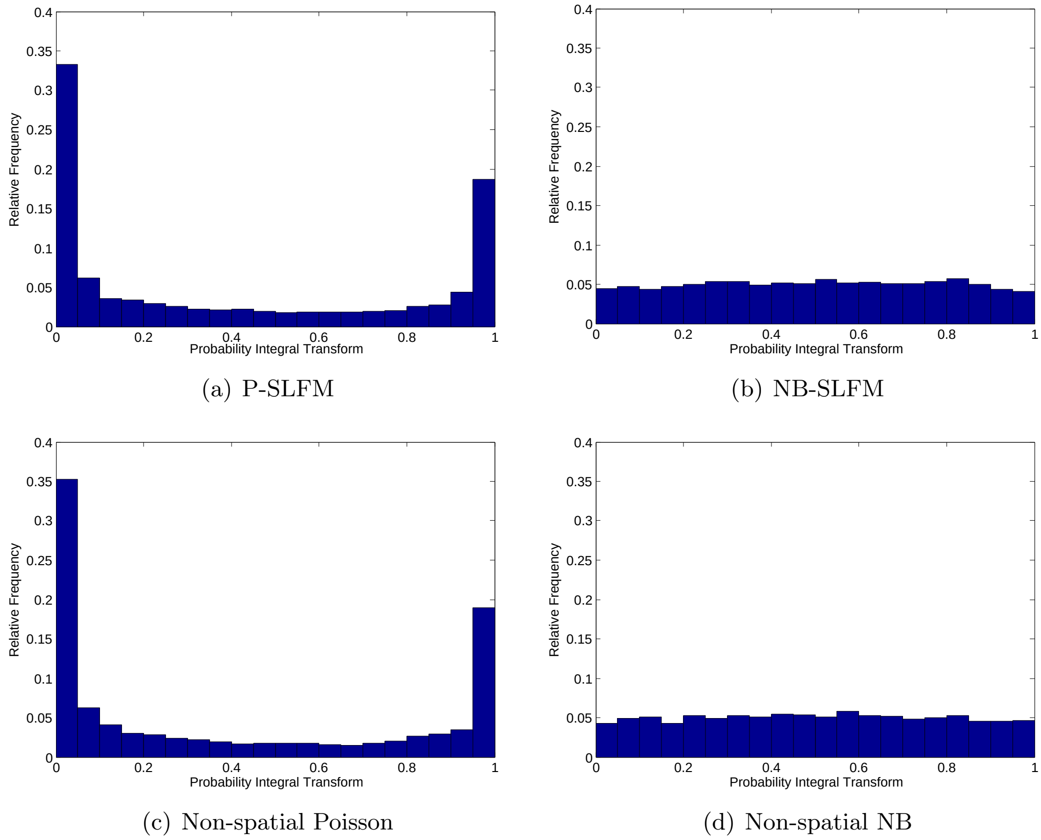


Figure 14: Nonrandomized PIT histograms for SLF models and the start-up firm births data.

that, the predicted effects for these regions distort the mean marginal effects. In contrast, calculating only the marginal effect for a certain (hypothetical) observation would not reflect the spatial nature of the data accurately since only one particular pattern of neighbors could then be considered.

As mentioned in Section 4.2, there exist direct and indirect marginal effects in a spatial model which sum up to the total marginal effects. The direct effect gives the impact of a change of regressor x_{ik} on observation i of the dependent variable and is therefore comparable to a marginal effect in a non-spatial model. The indirect effect is the sum of the impacts of a change in all other regressors x_{jk}, $j \neq i$ on observations i, which does not exist in a non-spatial model. The standard errors for the marginal effects are obtained using $B = 2000$ draws from the asymptotic joint distribution of the parameter estimators (LeSage and Pace, 2009).

As can be seen in Table 20, the majority of effects is significant. In general, more effects are significant in the negative binomial models than in the Poisson models. In the spatial models the indirect effects are smaller than the direct ones. When comparing the impacts between spatial and non-spatial models we see that the total marginal effects are in general of the same magnitude than the marginal effects of the corresponding non-spatial model. This implies that the spatial models decompose in some sense the effects from the non-spatial ones into direct and indirect ones highlighting the influence of neighboring observations. Again,

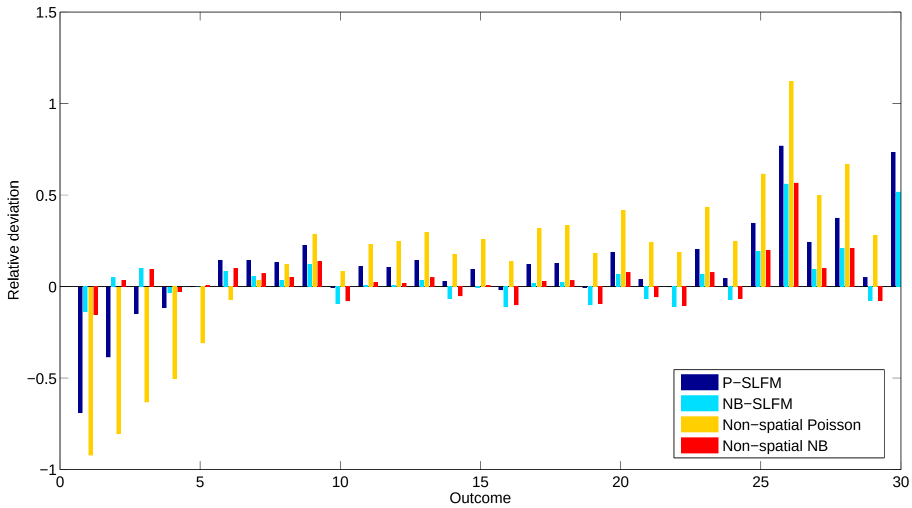


Figure 15: Relative deviations plot for SLF models and the start-up firm births data.

there are similar patterns for non-spatial and spatial versions of the Poisson as well as the negative binomial models. The variables *pedas* (*percent of adults with an associate's degree*), *bci* (*state tax business climate index*), and *proad* (*public road density*) for the P-SLFM and *cclass* (*share of workers in creative occupations*), *hwypc* (*government expenditures on highways per capita*), and *proad* for the NB-SLFM have the largest influence on the conditional mean of the dependent variable (besides the area dummies). The predictions for marginal effects are generally in line with the predicted marginal effects of the models presented in Chapter 3. The signs of significant effects are identical (only for the insignificant marginal effects of *tfdens* (*total establishment density*) in the spatial Poisson models different signs are observed). For most variables also the significance of the effects agrees, but a few additional variables have a significant influence here. The significance differs in some cases, especially the marginal effects of median household income (*mhhi*) and the government expenditures for education (*educpc*) are highly significant in both the P-SLFM and the NB-SLFM but not in the P-SAR and NB-SAR models. In most cases, the marginal effects of the SAR and SLF models are of the same magnitude. The total effects of the dummy variables *metro* and *micro*, which characterize the urbanization of the county, are smaller in the SLF model than in the SAR models, with an additional shift from indirect to direct effects. The latter are even higher than in the SAR models. Apart from that, the in absolute terms much smaller effect of *tfdens* and the larger effect of *hwypc*, both in the NB-SLFM are noticeable.

Variable	P-SLFM Total M.E.	P-SLFM Direct M.E.	P-SLFM Indirect M.E.	NB-SLFM Total M.E.	NB-SLFM Direct M.E.	NB-SLFM Indirect M.E.	Poisson Direct M.E.	NB Direct M.E.
msemp	0.377***	0.198***	0.148***	0.516***	0.339***	0.116***	0.326***	0.463***
	(0.061)	(0.040)	(0.032)	(0.032)	(0.026)	(0.018)	(0.044)	(0.024)
pelt10	−0.065**	−0.034**	−0.025**	0.050***	0.033***	0.011***	−0.022	0.048***
	(0.028)	(0.014)	(0.013)	(0.013)	(0.009)	(0.003)	(0.020)	(0.011)
pemt100	−0.366***	−0.192***	−0.144***	−0.226***	−0.148***	−0.051***	−0.314***	−0.171***
	(0.062)	(0.035)	(0.035)	(0.034)	(0.021)	(0.012)	(0.038)	(0.023)
tfdens	−0.140	−0.074	−0.055	−0.456***	−0.299***	−0.102***	0.066	−0.493***
	(0.123)	(0.065)	(0.050)	(0.157)	(0.109)	(0.034)	(0.107)	(0.125)
mhhi	−0.366***	−0.192***	−0.144***	0.237**	0.155**	0.053**	0.002	0.249***
	(0.095)	(0.046)	(0.049)	(0.103)	(0.064)	(0.026)	(0.091)	(0.049)
pop	0.022***	0.011***	0.008***	0.163***	0.107***	0.037***	0.025***	0.170***
	(0.005)	(0.003)	(0.002)	(0.033)	(0.023)	(0.008)	(0.005)	(0.030)
cclass	0.948***	0.498***	0.373***	0.987***	0.648***	0.222***	0.516***	0.764***
	(0.125)	(0.075)	(0.082)	(0.069)	(0.055)	(0.037)	(0.136)	(0.049)
uer	0.399	0.209	0.157	0.745***	0.489***	0.167***	0.779***	0.750***
	(0.398)	(0.212)	(0.159)	(0.215)	(0.133)	(0.059)	(0.236)	(0.124)
pedas	1.616***	0.848***	0.636***	0.608***	0.399***	0.136***	1.387***	0.415***
	(0.256)	(0.135)	(0.157)	(0.109)	(0.076)	(0.031)	(0.211)	(0.084)
awage	0.356***	0.187***	0.140***	−0.565***	−0.371***	−0.127***	0.198***	−0.354***
	(0.089)	(0.047)	(0.044)	(0.129)	(0.072)	(0.040)	(0.075)	(0.062)
netflow	0.032	0.017	0.013	−0.267***	−0.175***	−0.060***	0.025	−0.146***
	(0.027)	(0.014)	(0.011)	(0.063)	(0.035)	(0.020)	(0.024)	(0.025)
proad	1.002***	0.526***	0.395***	0.825***	0.542***	0.185***	1.097***	0.776***
	(0.251)	(0.147)	(0.114)	(0.236)	(0.155)	(0.064)	(0.194)	(0.213)
interst	0.086***	0.045***	0.034***	0.049***	0.032***	0.011***	0.078***	0.044***
	(0.014)	(0.008)	(0.008)	(0.009)	(0.006)	(0.003)	(0.011)	(0.007)
avland	−0.075***	−0.040***	−0.030***	−0.056***	−0.037***	−0.013***	−0.092***	−0.062***
	(0.018)	(0.010)	(0.008)	(0.009)	(0.006)	(0.003)	(0.016)	(0.007)
bci	1.368***	0.718***	0.539***	0.313	0.206	0.070	0.855**	0.316**
	(0.450)	(0.230)	(0.207)	(0.211)	(0.136)	(0.050)	(0.400)	(0.137)
educpc	0.065***	0.034***	0.025**	0.055**	0.036**	0.012**	0.038*	0.036
	(0.022)	(0.012)	(0.010)	(0.025)	(0.017)	(0.006)	(0.020)	(0.025)
hwypc	−0.420	−0.221	−0.165	−1.294***	−0.850***	−0.290***	−0.320	−0.258
	(0.248)	(0.132)	(0.103)	(0.313)	(0.197)	(0.090)	(0.200)	(0.194)
metro	23.233***	14.520***	7.382***	12.481***	8.632***	2.449***	19.385***	9.909***
	(2.972)	(1.486)	(1.791)	(1.304)	(0.752)	(0.513)	(2.288)	(0.822)
micro	10.901***	4.787***	5.031***	7.154***	4.650***	1.757***	6.200***	5.575***
	(2.349)	(0.639)	(1.677)	(0.759)	(0.397)	(0.378)	(0.788)	(0.434)

Table 20: Median marginal effects of P-SLFM, NB-SLFM, non-spatial Poisson, and non-spatial NB for the start-up firm births data with weight matrix W_{dnn}. For the dummy variables *metro* and *micro* the effect of a change from 0 to 1 is given. ** and *** denote a 5% and 1% significance, respectively. Standard errors (in brackets) are estimated using their sample counterparts of 2000 draws of the asymptotic joint distribution of the coefficients.

In conclusion, the two model classes, SAR and SLF, lead to similar results regarding the influence of the considered exogenous regressor on the number of start-up firm births. The diagnostics above show that none of the models considered here perfectly match the start-up firm births data set of the U.S. counties. Nevertheless, there is clear evidence for spatial dependency of the form considered in the SLF models in these data. Also, the results of the negative binomial models are clearly favoured upon those of the Poisson models. The additional source of dispersion introduced by the additional parameter in the negative binomial specification further improves the fit of the spatial linear feedback model. Therefore, the negative binomial spatial feedback model presents itself as best choice in this analysis.

4.6 Unilateral Modelling with Composite Maximum Likelihood

Because the SLF model does not have a reduced form and there is no operational multivariate count distribution for large n available, a full likelihood function cannot be obtained. In the previous sections it has been shown that the used pseudo likelihood leads to consistent estimates. A way to obtain a full likelihood function is to adopt a unilateral modelling approach (Arbia et al., 2014, pp. 178). In this approach, the units in the data set are ordered linearly resulting in a data structure similar to that of a time series. Correspondingly, the likelihood function is derived as the product of the conditional density functions, with the condition for observation i consisting of all preceding observations according to the applied ordering. According to Arbia et al. (2014, pp. 178) this unilateral model is supposed to be a good approximation to the complete spatial model as long as the spatial effect is isotropic, i.e. there is no directional differences of the spatial dependence, all neighbors influence the dependent variable by the same amount (see Cressie (1993, pp. 60) for a formal definition of isotropy in this context).

For the start-up firm births data set four different order directions based on the coordinates of the county centers (rounded to one digit) are taken into account: (i) from north-west to south-east, (ii) from south-west to north-east, (iii) from west-north to east-south, (iv) from west-south to east-north. For each of these order directions either the western or the eastern neighbors can be considered as preceding ones leading to a total of 8 different estimations. The conditional likelihood function of the unilateral P-SLFM for order direction k equals:

$$\begin{aligned} LogL_k &= \sum_{i=1}^{n} P(y_i|y_j, j \in N_k(i), X) \\ &= \sum_{i=1}^{n} y_i \log(\lambda \, \Sigma_{j \in N_k(i)} w_{ij} y_j + \exp(X_{i.}\beta)) \\ &\quad -\lambda \, \Sigma_{j \in N_k(i)} w_{ij} y_j + \exp(X_{i.}\beta) - \log(y_i!) \end{aligned}$$

where $N_k(i)$ denotes the set of preceding neighbors of observation i under order direction k. The likelihood specification implies that $y_1 \sim Po(\mu = \exp(X_{1.}\beta))$, i.e. the first observation in the respective order has no preceding neighbors. This approach only considers spatial effects from a part of the neighbors, which is not a plausible assumption for all applications. In case of start-up firm births this does not seems to be an intuitive approach. Particularly,

	NW ↔ SE		SW ↔ NE	
	Western Neighbors	**Eastern Neighbors**	**Western Neighbors**	**Eastern Neighbors**
λ	0.371 (0.072)	0.200 (0.037)	0.217 (0.041)	0.366 (0.076)
	WN ↔ ES		**WS ↔ EN**	
	Western Neighbors	**Eastern Neighbors**	**Western Neighbors**	**Eastern Neighbors**
λ	0.397 (0.069)	0.215 (0.040)	0.344 (0.064)	0.263 (0.063)

Table 21: Estimation results of the unilateral modelling approach using P-SLFM for the start-up firm births data, heteroscedasticity robust standard errors in brackets.

it ignores the dependencies between the preceding and following neighbors. Arbia (2014, pp. 178) argue that the estimates for the spatial autocorrelation parameter should be similar for different orders if the data is isotropic. In this case, the estimators of the unilateral model give good approximations for the parameters of the full model.

Table 21 displays the estimation results for the spatial autocorrelation parameter of the unilateral P-SLF models employing the four different orders mentioned above. The estimates vary between 0.200 and 0.397 which, together with the small standard errors, makes the isotropy of the data unlikely and therefore does not justify the usage of unilateral approximation for this application.

4.7 Summary

This chapter has introduced a new model for spatially autocorrelated count data, the spatial linear feedback model. The main advantage of this model is that it includes the neighboring observations linearly in the conditional expectation equation of a standard Poisson or negative binomial regression model. Doing this, a more intuitive interpretation of the spatial autocorrelation parameter is obtained as opposed to the ones in previous models for this data type. Additionally, the model can be estimated via pseudo maximum likelihood making it straightforwardly adoptable for empirically working economists. The asymptotic properties of the pseudo likelihood estimation are investigated in a Monte Carlo study which confirms its applicability. Secondly, evaluation methods have been introduced which are already used in the times series analysis of counts, i.e. PIT histograms and relative deviations plots. In the empirical example it is shown that the data set contains dispersion from more than one source, leading to inferior estimation results if only one source is considered. The estimated marginal effects are of the same magnitude for both spatial and non-spatial models, indicating a general appropriateness of the estimation strategy for the spatial models. In the spatial models, however, these effects can be separated into direct and indirect impacts to reveal the influence of neighboring observations. The best fit for the modelling of firm births in U.S. counties is reached with the negative binomial spatial linear feedback model.

5 Spatial Panel Models: Forecasting Crime Counts

5.1 Introduction

This chapter is concerned with modelling spatial panel data of counts and especially with its forecasting. The first of three approaches discussed in the following is an extension of the spatial linear feedback model of the previous chapter with fixed effects (Section 5.3). This follows the common way of modelling a fixed effects panel model for counts by introducing multiplicative fixed effects which are multiplied with the exponential function of the intensity equation.

The other two models additionally consider serial correlation. The second approach deals with an extension of a dynamic panel count data model. For this, the linear feedback model (LFM) of Blundell et al. (2002) is chosen as a starting point to which spatial terms are added (Section 5.4). The LFM is a fixed effects model with a linearly included serial correlation component and therefore has a structure which is suitable for the linear inclusion of spatial autoregressive terms.

The third modelling approach stems not from a count data model but from a spatial panel model for continuous data. To use this model for count data, the non-negativity of the estimated conditional expectations and especially of the forecasts must be ensured. This is handled with parameter restrictions and through the use of an exponential function for the explanatory variables.

All models are applied to forecasting crime in Pittsburgh's census tracts. The data set is described in the next section. In each of the following three sections one model and its results are discussed.

5.2 Data: Crime in Pittsburgh

The investigated panel data set contains crime counts for the 138 census tracts of Pittsburgh (in the borders for the 2010 census). It covers the months January 2008 to December 2013. 33 different crime types are reported separately, and summarized into two categories according to the U.S. Department of Justice (U.S. Department of Justice, Federal Bureau of Investigation, 2004). Part I crimes are all major crimes namely aggravated assault, burglary, larceny, motor vehicle theft, murder/manslaughter, negligent manslaughter, rape, and robbery. Part II crimes are accordingly the remaining, minor offenses like drunken driving, fraud, and vandalism (see U.S. Department of Justice, Federal Bureau of Investigation (2004, p. 8) for the full list). The logarithm of the lagged Part II crimes will be used as an explanatory variable in the models for the Part I crimes following the Broken Windows theory of crimes (Wilson and Kelling, 1982; Kelling and Coles, 1996). It says that the occurrence of less severe crimes often increases first in a neighborhood before more severe ones take place more often, too.

Figure 16 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 one. The highest number of Part I crimes with an average of 76 per months is observed for census tract 201 ("Downtown") followed by census tract 1702 ("Southside") which lies in 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 Figure 16) and ones with a small population density.

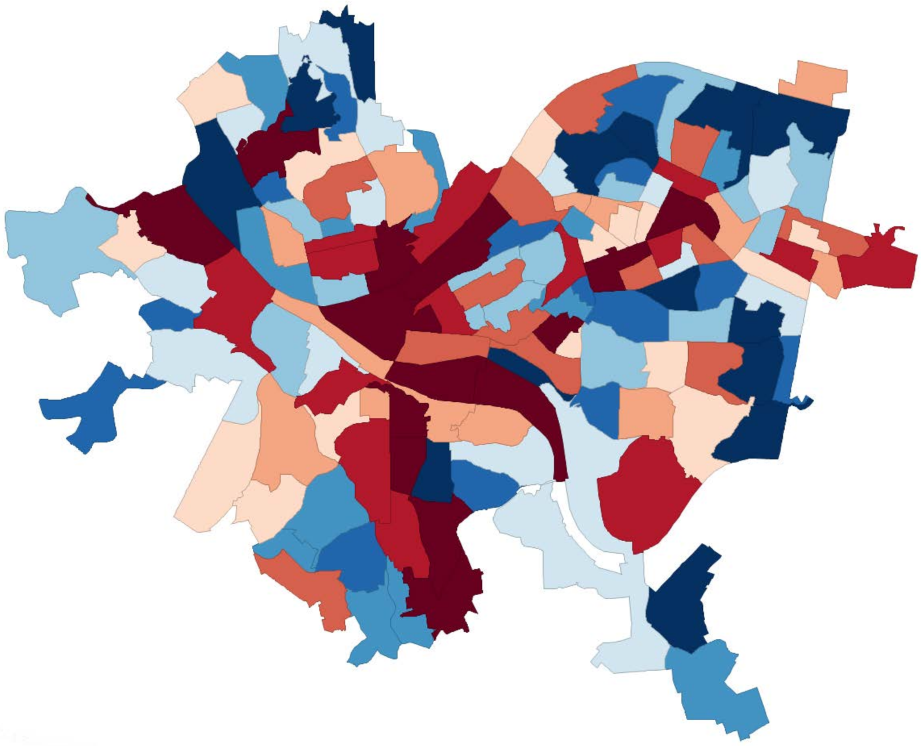


Figure 16: Map of time averages of Part I crimes in Pittsburgh, grouped in deciles. Dark red areas belong to highest deciles, dark blue ones to lowest.

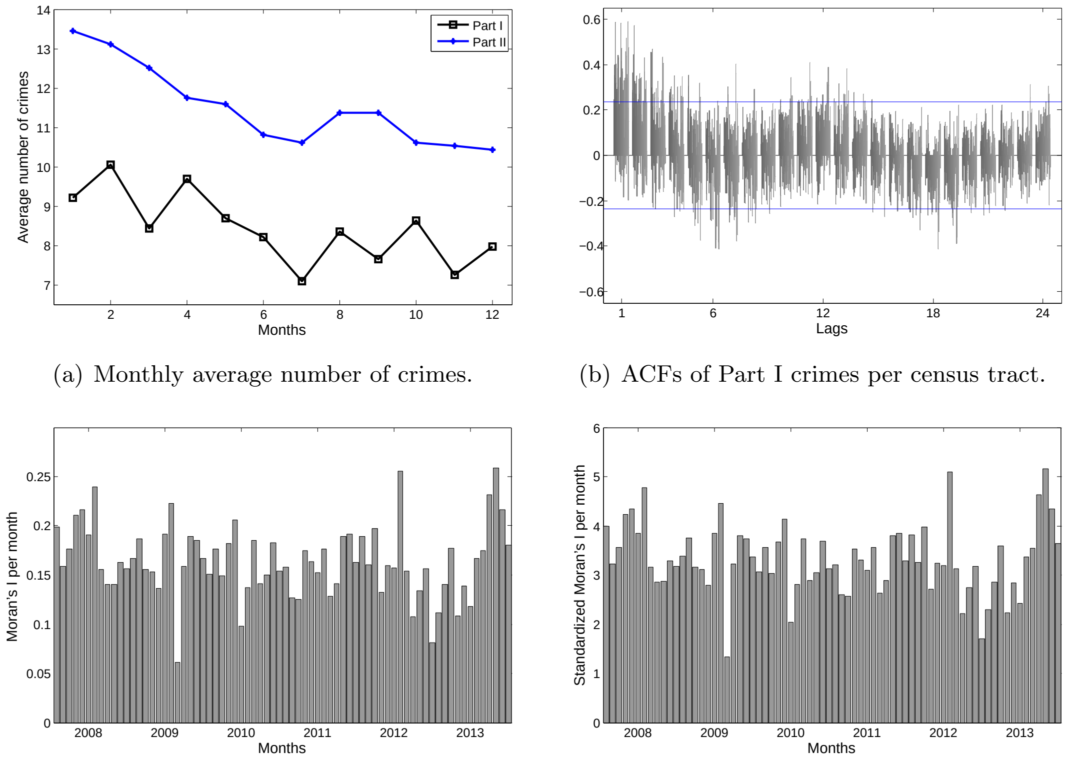
Table 22 gives some descriptive statistics of the variables in the data set for the whole period and for each year separately. On average, more Part II crimes than Part I crimes are observed per month and census tract, although the difference is small given the difference in the number of crime types in the two crime classes. The highest amount of crimes was conducted in 2008. A downward tendency is recognizable through the years in the sample, although the differences between the years are small. 2013 shows the smallest average crime rates.

Figure 17(a) gives a visual impression of the seasonal pattern in the data. Generally, there are more crimes in the first half of the year and also a hump in the Part II crimes during the months August and September. The autocorrelation functions (ACF) of the Part I crimes for each of the 138 census tracts are plotted together in Figure 17(b) (blue lines are the 95 % confidence interval). They indicate autocorrelation with the first lag for the majority of census tracts and with the second and third lag for some of the tracts. Also, the autocorrelation functions point to weak positive or negative autocorrelations with other lags in a few census tracts.

	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

Table 22: Descriptive statistics of the Pittsburgh crime data set.

The spatial weight matrix used for this application is a queen contiguity matrix. Each census tract has on average 5.6 neighbors. As a measure of spatial association Moran's I has been introduced in Section 3.6.1. Figures 17(c) and 17(d) show the values of Moran's I and standardized Moran's I for each month in the data set. The standardized values range from 1.334 to 5.155 with an average of 3.328. The lowest values are obtained for August 2009, June 2010, and December 2012. If an asymptotic normal distribution is assumed, only Moran's I of August 2009 and December 2012 are not significant at the 5% level. Even without an appropriate test at hand, these high values of Moran's I clearly point to the existence of spatial correlation in the data.



(a) Monthly average number of crimes.

(b) ACFs of Part I crimes per census tract.

(c) Moran's I of Part I crimes.

(d) Standard. Moran's I of Part I crimes.

Figure 17: Pittsburgh crime data descriptives.

5.3 The Spatial Linear Feedback Panel Model with Fixed Effects

5.3.1 Specification of the Spatial Linear Feedback Panel Model

In the course of this thesis, the extension of the SLFM for cross-sections, introduced in Chapter 4, to a panel fixed effects model is a straightforward way to estimate and forecast panel data of spatially correlated counts. The introduction of entity fixed effects is done in the usual fashion for count data modelling, i.e. by multiplying the fixed effects ν_i with the intensity equation. Assuming time invariant spatial dependence, the Poisson spatial linear feedback panel model (P-SLFPM) is obtained as:

$$y_{it}|\mu_{it}, \nu_i \sim Po(\nu_i \mu_{it})$$

$$\mu_{it} = \lambda \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_{jt} + \rho \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_{jt-1} + \exp(X_{it.}\beta) \tag{75}$$

with λ denoting the spatial autocorrelation parameter, w_{ij} the corresponding element of a row-standardized spatial weight matrix for the N spatial units, $i = 1, \ldots, N$, $X_{it.} = [x_{it1}, \ldots, x_{itK}]$ the vector of exogenous regressors with parameter vector β, and time periods $t = 1, \ldots, T$. In addition to the spatial dependence in the cross-section, this model also allows for dependence between the observation y_{it} and the neighboring observations in the previous time period. y_{-it} denotes the vector of observations in t of all neighbors to unit i. Because the spatial terms enter the conditional expectation $E[y_{it}|y_{-it}, y_{t-1}, X_{it.}] = \nu_i \mu_{it}$, which has to be non-negative for count data, special restrictions to the parameter spaces of λ and ρ apply. These depend on the values of W, X and β and have to be checked manually for each application. Alternatively, the conditions $\lambda, \rho > 0$ can be set independently of the data, even though the parameter space is then possibly stronger restricted than necessary, since λ and ρ can be allowed to be negative as long as the estimated conditional expectation is positive for each unit.

The resulting model is estimated using the pseudo maximum likelihood approach explained in Chapter 4.2 and the conditional maximum likelihood for Poisson fixed effects models, introduced for this particular case by Palmgren (1981) and Hausman et al. (1984). The log likelihood function resulting from conditioning on the sufficient statistics for ν_i, $\sum_{t=1}^{T} y_{it}$, is:

$$\log L = \sum_{i=1}^{N} \left[\log(\sum_{t=1}^{T} y_{it})! - \sum_{t=1}^{T} \log(y_{it}!) + \sum_{t=1}^{T} y_{it} \log \left(\frac{\mu_{it}}{\sum_{s=1}^{T} \mu_{is}} \right) \right] \tag{76}$$

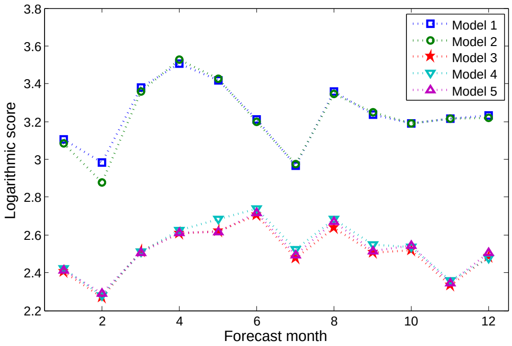
Cluster-robust standard errors are obtained by the sandwich formula using the scores of the log-likelihood function (e.g. described in Cameron and Trivedi (2013, pp. 353)). To obtain one-step ahead point forecasts from this model, the intensity in Equation (75) is evaluated for $t = T + 1$ and multiplied by the estimate of the fixed effect

$$\hat{\nu}_i = \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} \hat{\mu}_{it}} \tag{77}$$

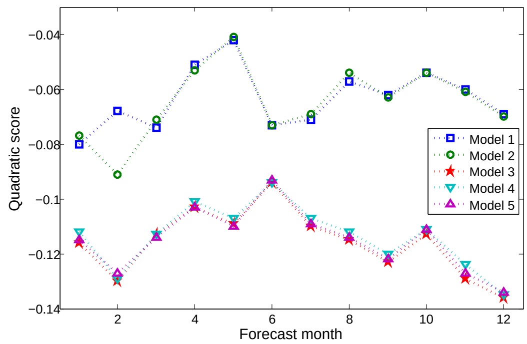
Rounding can take place to obtain integer forecasts. Additionally, the result can be employed as the conditional expectation of a Poisson distribution to gain a forecast density.

5.3.2 Empirical Application: Forecasts for Pittsburgh's Crime Counts

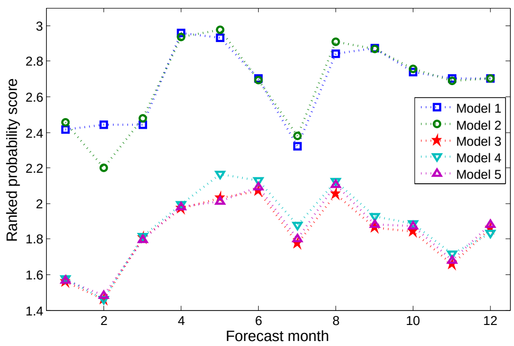
The P-SLFPM model is applied to the crime data set from Pittsburgh. To forecast the number of Part I crimes in the census tracts of Pittsburgh for each month of 2013 a one-step-ahead expanding window forecast is employed. Therefore, the first estimation sample ranges from January 2008 to December 2012. Results of the parameter estimates of this first estimation sample are displayed in Table 23. Model 1 is a pooled estimation using the P-SLF model, Model 2 a panel estimation with monthly time fixed effects and Models 3-5 include entity fixed effects as well as time fixed effects. The simultaneous spatial correlation parameter λ is highly significant in all specifications as well as the time lagged Part II crimes. Also, the serial lagged spatial term is highly significant. Its parameter estimate is relatively small compared to the one of the simultaneous spatial term. The size remains stable, whether or whether not the Part II crimes are included as a regressor, which entails information about the same period as the serial lagged spatial term. Apparently they are independent of each other. The majority of the time dummy parameters turns from negative to positive values when individual fixed effects are included (base category is January). February shows the least criminal activity with a significant, negative coefficient. The summer months have the



(a) Logarithmic score



(b) Quadratic score



(c) Ranked probability score

Figure 18: Scoring rules of density forecasts from the P-SLFPM for the Pittsburgh crime data.

highest coefficient values which supports the hypothesis that criminal activity raises with the temperature, since then, a bad temper is evoked faster (Gorr et al., 2003). Therefore, the inclusion of one dummy for the months May to September, instead of using monthly time dummies, has been considered as well, but does not improve the forecasts. Results for this specification are displayed in Appendix C.1, Tables 70 to 72.

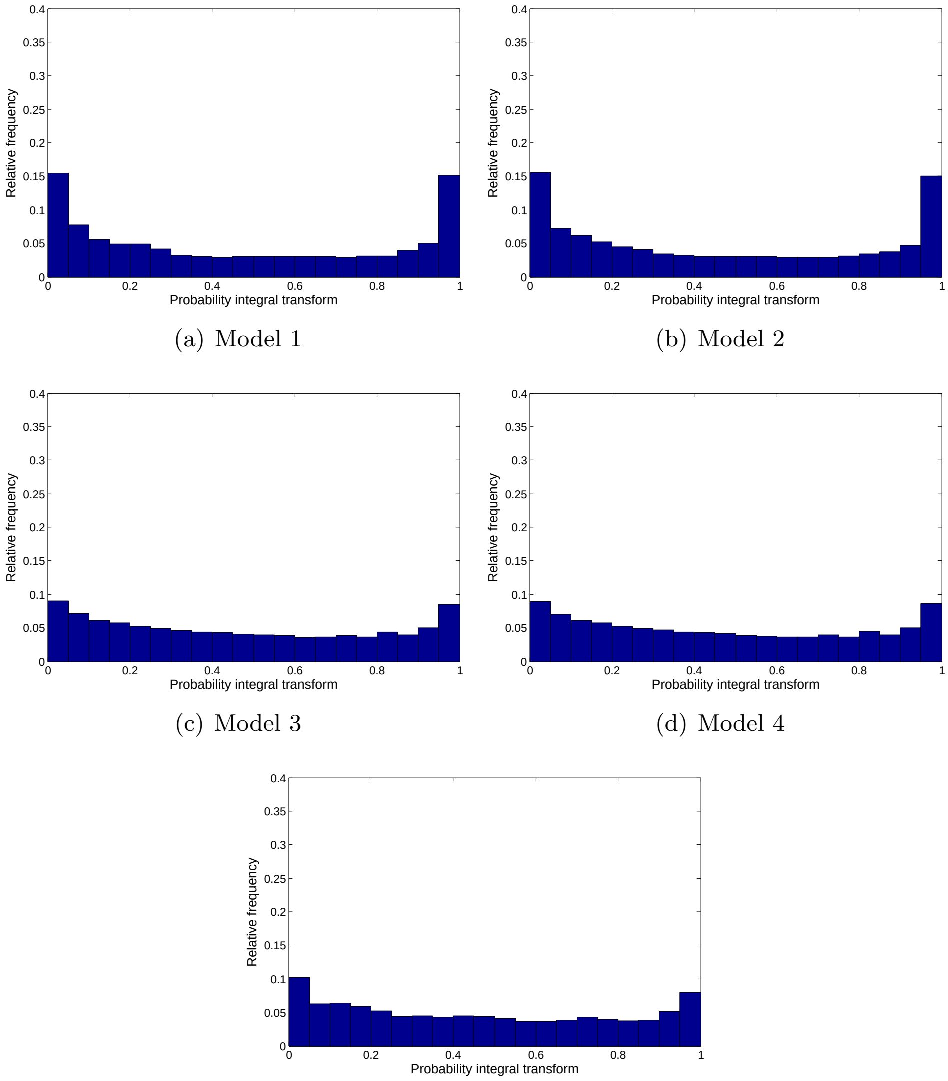
The lower part of Table 23 gives the root mean squared forecast error (RMSFE) and mean absolute forecast error (MAFE) averaged over the 12 months of the forecasting period (January to December 2013). Table 73 in Appendix C.1 displays the forecast results for each month separately. In general, the measures over the single months are close together. The specifications with entity fixed effects clearly outperform the pooled and time fixed effects model in terms of point forecast accuracy. Among those with both types of fixed effects the one including all introduced regressors performs best (even though only with a small lead) with an average mean absolute forecast error of 2.590.

	1 pooled	2 time effects	3 fixed effects	4 fixed effects	5 fixed effects
Estimates for January 2008 to December 2012					
$WPartI_t$	0.247**	0.235***	0.118***	0.112***	0.101***
	(0.115)	(0.018)	(0.020)	(0.027)	(0.020)
$WPartI_{t-1}$			0.043***		0.040**
			(0.015)		(0.016)
$\log(PartII_{t-1})$	0.943***	0.930***	0.090***	0.091**	
	(0.224)	(0.034)	(0.033)	(0.040)	
Jan		−0.629***			
		(0.114)			
Feb		−0.320***	−0.259***	−0.205**	−0.255**
		(0.112)	(0.076)	(0.090)	(0.104)
Mar		−0.450***	0.112	0.045	0.120
		(0.119)	(0.084)	(0.079)	(0.118)
Apr		−0.368***	0.134**	0.099	0.163
		(0.117)	(0.066)	(0.062)	(0.098)
May		−0.398***	0.182***	0.158***	0.206**
		(0.119)	(0.065)	(0.061)	(0.094)
Jun		−0.352***	0.222***	0.198***	0.253***
		(0.121)	(0.062)	(0.057)	(0.090)
Jul		−0.326***	0.259***	0.234***	0.287***
		(0.119)	(0.058)	(0.052)	(0.081)
Aug		−0.421***	0.297***	0.274***	0.327***
		(0.117)	(0.052)	(0.047)	(0.075)
Sep		−0.382***	0.152**	0.162**	0.177**
		(0.117)	(0.066)	(0.065)	(0.090)
Oct		−0.423***	0.139**	0.128**	0.157
		(0.116)	(0.065)	(0.064)	(0.092)
Nov		−0.388***	0.064	0.066	0.079
		(0.112)	(0.072)	(0.072)	(0.097)
Dec		−0.321***	0.413***	0.365***	0.439***
		(0.111)	(0.104)	(0.115)	(0.156)
Constant	−0.449				
	(0.778)				
Log L	−26766	−26650	−223777	−223804	−223789
Average Forecast Results for January to December 2013					
RMSFE	5.523	5.507	3.567	3.660	3.603
MAFE	3.583	3.571	2.590	2.649	2.619

Table 23: Results from the P-SLFPM for the Pittsburgh crime data. Robust standard errors in parentheses; W is a queen contiguity spatial weighting matrix; ** and *** denote 5% and 1% statistical significance.

Additionally to point forecasts, density forecasts can be obtained. Since the point forecast (before rounding) is actually the forecast of the conditional expectation of the respective density, it directly leads to Poisson density forecasts. Figure 18 displays the scoring rules (introduced in Chapter 3.5) of the density forecasts from Model 1 to 5. The numerical values can be found in Appendix C.1, Table 74. The forecast for the summer months are slightly worse, except for July, than for the rest of the year. The entity fixed effect models again outperform the other specifications. Model 3 is evaluated best by all three criteria but is closely followed by the other entity fixed effects specifications.

An alternative way to evaluate density forecast is by looking at their PIT histograms. The



(a) Model 1

(b) Model 2

(c) Model 3

(d) Model 4

(e) Model 5

Figure 19: Nonrandomized PIT histograms of density forecasts from the P-SLFPM for the Pittsburgh crime data.

nonrandomized PIT histograms (for a description of the concept see Chapter 4.3) are shown in Figure 19. For panel data, the PIT residuals are averaged over both dimensions to obtain one graph per forecasting model (Diebold et al., 1998). The histograms again show the better fit of the entity fixed effects models indicated by the almost even shape of the bins. On the contrary, the histograms of Model 1 and 2 have a clear u-shape, which points towards more dispersion in the data than the model is able to forecast. But also none of the histograms for the fixed effects models displays a completely even distribution. This leads to the conclusion that there is some dispersion in the data which is not captured by the forecasts of any of these Poisson models. To improve the forecasting of the crime counts, dynamic models which incorporate a serial lag of the Part I crimes are introduced in the next two sections.

5.4 A Dynamic Spatial Panel Model for Counts with Multiplicative Fixed Effects

5.4.1 Specification of the Model with Multiplicative Fixed Effects

The autocorrelation functions in Section 5.2 indicate that the data inhibits serial correlation. Therefore, the use of a dynamic model may lead to improved forecasts. For this purpose, a dynamic count data model from the literature is extended with spatial terms. The linear feedback model (LFM) of Blundell et al. (2002) is chosen, which is an observation-driven dynamic count data model. Blundell et al. (2002) follow a distribution-free approach and estimate the model using quasi-differenced generalized method of moments (GMM). They argue that the usual specification of the conditional mean of a count variable, including fixed effects ν_i, $E[y_{it}|X_{it.}\nu_i] = \exp(X_{it.}\beta + \eta_i) = \mu_{it}\nu_i$ implicitly gives a regression model of the form

$$y_{it} = \mu_{it}\nu_i + u_{it}$$

Based on this, they introduce the dynamic LFM for count data

$$\begin{aligned} E[y_{it}|y_{it-1}, X_{it.}, \nu_i] &= \gamma y_{it-1} + \mu_{it}\nu_i \\ &= \gamma y_{it-1} + \exp(X_{it.}\beta + \eta_i) \end{aligned}$$

where $X_{it.} = [x_{it1}, \ldots, x_{itK}]$ is the vector of explanatory variables with parameter vector β, $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The serial correlation parameter γ needs to fulfil $\gamma \geq 0$ to ensure positive estimates for the conditional expectation. By including the serial lag of the dependent variable additively instead of including it in the argument of the exponential function, there is no risk of explosive behavior in this model as long as $\gamma < 1$ also holds.

For our purpose, a simultaneous and serially lagged spatial term is added to the LFM to obtain a spatial dynamic fixed effects model:

$$E[y_{it}|Y_{-it}, Y_{t-1}, X_{it.}, \nu_i] = \gamma y_{it-1} + \lambda \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_{jt} + \rho \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_{jt-1} + \exp(X_{it.}\beta + \eta_i) \tag{78}$$

with w_{ij} being elements from a spatial weight matrix W, Y_{-it} are the observations of all neighbors of i for period t, Y_{t-1} is the vector of N observations for period $t-1$. Again, the estimation results from Equation (78) must be positive, which can be reached by restricting $\gamma \geq 0$, $\lambda \geq 0$, and $\rho \geq 0$. Valid exceptions are possible, depending on the values of X, β, and W, as long as Equation (78) is positive.

The corresponding regression model is given by

$$y_{it} = \gamma y_{it-1} + \lambda \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt} + \rho \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt-1} + \exp(X_{it.}\beta + \eta_i) + u_{it} \tag{79}$$

Aside from the restrictions on the serial and spatial autocorrelation parameters that guarantee a positive conditional expectation, the parameters also need to fulfil 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 Elhorst (2012) and Parent and LeSage (2011), which imply that there is a trade-off between the serial and spatial correlation to maintain a stable model (Elhorst, 2008):

$$\gamma < 1 - (\lambda + \rho)\omega_{max} \quad \text{if } \lambda + \rho \geq 0 \tag{80}$$
$$\gamma < 1 - (\lambda + \rho)\omega_{min} \quad \text{if } \lambda + \rho < 0 \tag{81}$$
$$\gamma > (\lambda + \rho)\omega_{max} - 1 \quad \text{if } \lambda - \rho \geq 0 \tag{82}$$
$$\gamma > (\lambda + \rho)\omega_{min} - 1 \quad \text{if } \lambda - \rho < 0 \tag{83}$$

where ω_{min} and ω_{max} denote the smallest and largest characteristic root of the spatial weight matrix W. The largest characteristic root of row-standardized spatial weight matrices equals 1.

5.4.2 Quasi-Differenced GMM Estimation

The model in Equation (79) can be estimated with quasi-differenced GMM as long as the regressors X do not contain endogenous variables (Blundell et al., 2002). If the regressors X contain endogenous variables, the multiplicative specification of the fixed effects together with additive error terms cannot be estimated since no valid moment conditions are available (Windmeijer, 2008). Chamberlain (1992) and Wooldridge (1997) propose quasi-differencing transformations of the errors which eliminate the fixed effects and can therefore be used in the GMM function. The moment conditions based on these are only valid for predetermined or exogenous explanatory variables X. The Wooldridge quasi-differenced errors are used in the following and obtained by

$$\begin{aligned} q_{it} &= \frac{u_{it}}{\exp(X_{it.}\beta)} - \frac{u_{it-1}}{\exp(X_{it-1.}\beta)} \\ &= \frac{y_{it} - \gamma y_{it-1} + \lambda \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt} + \rho \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt-1}}{\exp(X_{it.}\beta)} \\ &\quad - \frac{y_{it-1} - \gamma y_{it-2} + \lambda \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt-1} + \rho \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt-2}}{\exp(X_{it-1.}\beta)} \end{aligned} \tag{84}$$

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

$$E[q_{it}|y_i^{t-2}, x_i^{t-1}] = 0 \tag{85}$$

with $y_i^{t-2} = [y_{i1}, \dots, y_{it-2}]$ and $x_i^{t-1} = [x_{i1}, \dots, x_{it-1}]$.
This allows the choice of instruments for the GMM estimation which are gathered in matrix Z. 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, X_{t-1} and X_{t-2}, form the instrument matrix. Additional exogenous variables added to the model are also included in Z. If the model contains a serially lagged spatial term (Wy_{t-1}), an additional lag of the spatial term, Wy_{t-4}, is inserted in Z. To reduce the instrument count and to limit 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 is asked to minimize for example the condition $\sum_t \Delta y_{i,t-2} q_{it}$ instead of the $T-2$ conditions $\Delta y_{i,t-2} q_{it}$, $t = 2, \dots, T$. Using collapsed instruments leads to a more precise estimation of the optimal weight matrix in the second step and less biased estimates (Roodman, 2009).
The instrument matrix takes the form:

$$Z_i = \begin{bmatrix} y_{i1} & 0 & [Wy_1]_i & 0 & 0 & x_{i2} & x_{i1} \\ y_{i2} & y_{i1} & [Wy_2]_i & [Wy_1]_i & 0 & x_{i3} & x_{i2} \\ y_{i3} & y_{i2} & [Wy_3]_i & [Wy_2]_i & [Wy_1]_i & x_{i4} & x_{i3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{iT-2} & y_{iT-3} & [Wy_{T-2}]_i & [Wy_{T-3}]_i & [Wy_{T-4}]_i & x_{iT-1} & x_{iT-2} \end{bmatrix}$$

where $[Wy_1]_i$ denotes the i^{th} row of the product Wy_1.

The quasi-differenced error terms and the instrument matrix enter the GMM function

$$gmf = \Big(\frac{1}{N}\sum_{i=1}^{N} q_i(\theta)' Z_i\Big) H^{-1} \Big(\frac{1}{N}\sum_{i=1}^{N} Z_i' q_i(\theta)\Big) \tag{86}$$

with $q_i(\theta) = (q_{i3}, q_{i4}, \dots, q_{iT})'$, θ is the vector of parameters to be estimated, and H is a weight matrix. This function is numerically minimized with respect to the parameters θ in two steps to obtain consistent (already achieved in the first step) and efficient estimates. The weight matrix is set to the identity matrix for the first step. Afterwards, the weight matrix for the second step is calculated using the results from the first step, $\hat{\theta}_1$:

$$H(\hat{\theta}_1) = \frac{1}{N}\sum_{i=1}^{N} Z_i' q_i(\hat{\theta}_1) q_i(\hat{\theta}_1)' Z_i \tag{87}$$

The asymptotic variance of the resulting efficient two-step GMM estimator is

$$\widehat{var}(\hat{\theta}_2) = \frac{1}{N}\Big(C(\hat{\theta}_2)' H^{-1}(\hat{\theta}_1) C(\hat{\theta}_2)\Big)^{-1} \tag{88}$$

where

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

One-step ahead forecasts cannot be obtained straightforwardly because unlike in the case of difference GMM for continuous data, the forecast of the quasi-differenced dependent variable 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. From the following equation, which equals the Wooldridge transformation for a static non-spatial count model in $t = T + 1$,

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

an expression for y_{T+1} can be derived:

$$\begin{aligned} y_{iT+1} = & \left(\frac{y_{iT}}{\exp(X_{iT.}\beta)} - A_{i.}\left(\gamma\left(\frac{y_{T-1}}{\exp(X_{iT.}\beta)} - \frac{y_T}{\exp(X_{iT+1.}\beta)}\right)\right.\right. \\ & \left.\left. + \rho W_{i.}\left(\frac{y_{T-1}}{\exp(X_{iT.}\beta)} - \frac{y_T}{\exp(X_{iT+1.}\beta)}\right)\right)\right)\exp(X_{iT+1.}\beta) \\ & - A_{i.}\left(\frac{u_{iT}}{\exp(X_{iT.}\beta)} - \frac{u_{iT+1}}{\exp(X_{iT+1.}\beta)}\right)\exp(X_{iT+1.}\beta) \end{aligned} \tag{89}$$

with $A_{i.}$ being the i^{th} row of the Leontief inverse $A = (I - \lambda W)^{-1}$ and $W_{i.}$ the i^{th} row of the spatial weight matrix. Because of the assumed predetermination of X and Equations (84) and (85) the expected value of Equation (89) is given by

$$E[y_{iT+1}|x^{T+1}] = \left(\frac{y_{iT}}{\exp(X_{iT.}\beta)} - A_{i.}\left(\gamma\left(\frac{y_{T-1}}{\exp(X_{iT.}\beta)} - \frac{y_T}{\exp(X_{iT+1.}\beta)}\right)\right)\right)\exp(X_{iT+1.}\beta) \tag{90}$$

which can serve as a one-step ahead forecast if x_{iT+1} is known at time T.

5.4.3 Illustration with Simulated Data

Before forecasting the crime counts of Pittsburgh, a small simulation study is supposed to give some insights into the behavior of the quasi-differenced GMM estimates. For this purpose, a smaller version of the model is estimated including only the simultaneous spatial term. To get the whole picture of the model's properties, the entire parameter space of the model should be covered by the generated data. For this, the values of the 3 parameters γ, λ, and β should be varied and the resulting combinations be used in samples of different size, obtained by varying N and T. This would lead to a multitude of configurations and the need of time and computational resources for a Monte Carlo study of this size cannot be provided in the context of this thesis. Therefore, 16 exemplary configurations are chosen. The serial and spatial correlation parameters take the values: $(\gamma, \lambda) = \{(0.5, 0.4), (0.2, 0.1), (0.2, 0.6), (0.7, 0.2)\}$. These represent situations with either strong serial or spatial correlation as well as strong serial and spatial correlation and weak serial and spatial correlation. In accordance with the crime data set only large T-samples are considered with $(N, T) = \{(50, 50), (100, 50), (100, 100), (500, 100)\}$. The model contains one additional regressor x whose coefficient β is set to 1.5 in all configurations. For each of these 16 combinations, 1000 repetitions are generated. The outcomes of the dependent variable are generated using an iterative procedure 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. The vector y_t is updated instantly after each step $i = 1 \dots N$. After B iterations over all units i, the last vector is kept as y_t and used as a regressor for generating the outcomes for $t + 1$. The outcomes are calculated using equation

$$y_{i,t} = \gamma y_{j,t-1} + \lambda \sum_{j=1}^{N} w_{ij} y_t + \exp(x_{it}\beta + \eta_i) + e_{i,t} \tag{91}$$

where the error terms $e \sim$ $i.i.d.$ $N(0, 5)$, the individual effects $\eta \sim$ $i.i.d.$ $N(0, 0.5)$, and the $N \times 1$ starting vector $y_1 = \exp(X_{.1}\beta + \eta) + e_1$. The spatial contiguity weight matrix W is generated using the function `xy2cont` of the Spatial Statistics Toolbox for MATLAB 2.0 (Pace, 2003) on randomly generated coordinates. The result of Equation (91) is rounded to the nearest integer. If this returns a negative value, it is set to 0 to meet the required count data nature of the generated data. The $N \times 1$ vector of the regressor $X_{.t}$ is generated iteratively in the style of Blundell et al. (2002) with serial correlation and correlated with the individual effects

$$X_{.t} = \delta X_{.t-1} + \tau\eta + \epsilon_t$$

with random term ϵ $\sim i.i.d.$ $N(0, 0.1)$ and $X_{.1} = \frac{\tau}{1-\delta}\eta + \xi$, ξ $\sim i.i.d.$ $N(0, \frac{0.5}{1-\delta^2})$, $\tau = 0.1$, and $\delta = 0.5$.

The results of the simulations, displayed in Table 24, are mixed. Among all parameters the estimates for γ have the smallest RMSE which decreases with increasing sample size and is mostly independent of the size of the true γ. The RMSE of $\hat{\lambda}$ also decreases with increasing sample size but is in general larger than that of $\hat{\gamma}$. Especially for samples with small true λ, it is considerably large. The worst fit is obtained for estimates of β. The RMSE decreases for increasing sample size but only slightly and is very sensitive to small true values of γ and λ. The biases of all parameter estimates are negative, with very few exceptions, and are sensitive to small values of true γ. Again the results for $\hat{\beta}$ in terms of bias are much worse than for $\hat{\gamma}$ and $\hat{\lambda}$. All in all in the scope of this small illustration, the GMM estimator performs acceptably well for estimating the serial autocorrelation parameter γ and the spatial autocorrelation parameter λ, if their true values are of medium size. But it is not

N	T	γ	λ	RMSE $\hat{\gamma}$	RMSE $\hat{\lambda}$	RMSE $\hat{\beta}$
50	50	0.5	0.4	0.033	0.089	1.155
100	50	0.5	0.4	0.026	0.060	1.036
100	100	0.5	0.4	0.022	0.042	0.736
500	100	0.5	0.4	0.013	0.022	0.462
50	50	0.2	0.1	0.064	0.380	1.080
100	50	0.2	0.1	0.065	0.279	1.092
100	100	0.2	0.1	0.064	0.190	1.042
500	100	0.2	0.1	0.060	0.092	1.018
50	50	0.2	0.6	0.036	0.126	1.050
100	50	0.2	0.6	0.032	0.101	0.891
100	100	0.2	0.6	0.027	0.083	0.814
500	100	0.2	0.6	0.020	0.057	0.704
50	50	0.7	0.2	0.036	0.095	1.194
100	50	0.7	0.2	0.032	0.063	1.070
100	100	0.7	0.2	0.029	0.043	0.860
500	100	0.7	0.2	0.021	0.022	0.593

N	T	γ	λ	Bias $\hat{\gamma}$	Bias $\hat{\lambda}$	Bias $\hat{\beta}$	Relative Bias $\hat{\gamma}$	Relative Bias $\hat{\lambda}$	Relative Bias $\hat{\beta}$
50	50	0.5	0.4	−0.011	−0.004	−0.198	−0.021	−0.010	−0.132
100	50	0.5	0.4	−0.015	−0.011	−0.444	−0.030	−0.028	−0.296
100	100	0.5	0.4	−0.016	−0.016	−0.469	−0.032	−0.041	−0.313
500	100	0.5	0.4	−0.012	−0.010	−0.369	−0.023	−0.025	−0.246
50	50	0.2	0.1	−0.057	0.140	−0.423	−0.283	1.402	−0.282
100	50	0.2	0.1	−0.062	0.041	−0.900	−0.310	0.408	−0.600
100	100	0.2	0.1	−0.063	−0.005	−1.014	−0.313	−0.045	−0.676
500	100	0.2	0.1	−0.060	−0.025	−1.013	−0.298	−0.245	−0.675
50	50	0.2	0.6	−0.020	−0.003	−0.186	−0.099	−0.004	−0.124
100	50	0.2	0.6	−0.023	−0.032	−0.593	−0.116	−0.053	−0.395
100	100	0.2	0.6	−0.023	−0.048	−0.719	−0.114	−0.080	−0.479
500	100	0.2	0.6	−0.019	−0.047	−0.684	−0.094	−0.079	−0.456
50	50	0.7	0.2	−0.017	−0.005	−0.251	−0.025	−0.023	−0.167
100	50	0.7	0.2	−0.023	−0.009	−0.519	−0.033	−0.046	−0.346
100	100	0.7	0.2	−0.025	−0.012	−0.548	−0.036	−0.059	−0.366
500	100	0.7	0.2	−0.020	−0.007	−0.452	−0.028	−0.037	−0.301

Table 24: Monte Carlo results for the dynamic spatial panel model with multiplicative fixed effects. Relative bias is calculated as bias divided by true parameter value. The instrument matrix contains additionally a constant.

able to accurately estimate β, the coefficient of the predetermined regressor X. Nevertheless, the investigation here is very limited and a closer inspection would be necessary to come to general conclusions about the applicability of a quasi-differenced GMM estimator for estimating the spatial version of this model.

5.4.4 Empirical Application: Forecasts for Pittsburgh's Crime Counts

Despite the partly discouraging Monte Carlo results in the previous section, the forecast ability of the model with multiplicative fixed effects is also investigated. For this, the crime data set described in Section 5.2 and a one-step-ahead expanding window forecast are used again. Table 25 displays the estimation results for the data from January 2008 to December 2012 as well as the average RMSFE and MAFE for the one-step-ahead forecasts of January to December 2013. Detailed results of the forecasting exercise are displayed in Appendix C.2, Table 77, as well as the results for estimations with one summer dummy for the months May to September instead of monthly time dummies (Tables 75 and 76).

The relevant stationarity conditions from Equations (80) - (83) are fulfilled by the results in Table 25 ($\omega_{min} = -0.333$, $\omega_{max} = 1$) and the Sargan test does not reject instrument exogeneity for any specification. But none of the regressors is significant in any of the specifications. This stands in strong contrast to the results for the spatial linear feedback panel model in Section 5.3.2 where serial and spatial correlation of the Part I crimes as well as the lagged Part II crimes are significant. The average forecast results displayed in the lower part of Table 25 point to the importance of the simultaneous spatial term Wy_t for forecasting. But even the best specification, which is Model 1 containing all proposed regressors, is clearly worse than the fixed effects models of the panel spatial linear feedback model. Together with the results from the Monte Carlo illustration these outcomes indicate that the estimation approach used here is not reliable. Blundell et al. (2002) also report shortcomings of the quasi-differenced estimation for the non-spatial LFM and propose alternatively a pre-sample mean estimator, which is less biased. But for this estimator a long time series of the dependent variable must be available previous to the actual estimation sample. Therefore, its applicability depends on the data availability which is not given here.

	1	**2**	**3**	**4**	**5**
Estimates for January 2008 to December 2012					
$PartI_{t-1}$	0.031	0.043	0.318	0.023	0.034
	(0.383)	(0.322)	(1.050)	(0.401)	(0.312)
$WPartI_t$	0.606	0.564		0.640	0.575
	(0.407)	(0.425)		(0.419)	(0.594)
$WPartI_{t-1}$	−0.037		0.240	−0.064	
	(0.284)		(0.170)	(0.290)	
$log(PartII_{t-1})$	0.522	0.433	1.542	0.551	0.481
	(2.036)	(1.880)	(5.079)	(1.736)	(2.679)
Feb	−0.407	−0.498	−0.165		
	(2.367)	(2.276)	(2.633)		
Mar	−0.133	−0.027	0.956		
	(2.527)	(2.393)	(2.643)		
Apr	−0.144	−0.054	0.877		
	(3.310)	(2.968)	(3.829)		
May	−0.088	0.016	0.159		
	(3.519)	(3.043)	(3.029)		
Jun	−0.097	0.013	0.175		
	(3.799)	(3.291)	(3.176)		
Jul	0.146	0.252	0.574		
	(3.928)	(3.320)	(4.125)		
Aug	0.103	0.214	0.360		
	(4.341)	(3.624)	(4.377)		
Sep	0.069	0.059	0.459		
	(3.084)	(2.577)	(5.033)		
Oct	0.144	0.135	0.881		
	(2.843)	(2.482)	(5.635)		
Nov	−0.112	−0.103	0.032		
	(2.612)	(2.338)	(2.749)		
Dec	−0.110	−0.135	0.379		
	(2.734)	(2.492)	(3.992)		
Sargan	0.000	0.000	0.002	0.000	0.000
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Average Forecast Results for January to December 2013					
RMSFE	5.866	5.540	17.711	6.001	5.607
MAFE	4.013	3.844	7.094	4.066	3.862

Table 25: Results from the dynamic spatial panel model with multiplicative fixed effects for the Pittsburgh crime data. Standard errors in parentheses; p-value of the Sargan test in brackets; W is a queen contiguity spatial weighting matrix; ** and *** denote 5% and 1% statistical significance. Values reported as 0.000 are smaller than 0.0005.

5.5 A Dynamic Spatial Panel Model for Counts with Additive Fixed Effects

5.5.1 Specification of the Model with Additive Fixed Effects

This section describes an alternative autoregressive model to forecast spatial count data. Instead of extending a multiplicative fixed effects count data model with spatial terms, the spatial model considered here was originally intended for continuous data by Kukenova and Monteiro (2009) and includes the fixed effects additively. When using the model for count data, the conditional expectation of the count variable y_{it} from this model can either entail an exponential term for the explanatory variables X, which is a typical form in count data models, or be entirely linear. In the former case it is given by:

$$E[y_{it}|Y_{-it}, Y_{t-1}, X_{it.}, \eta_i] = \gamma y_{it-1} + \lambda \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt} + \rho \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt-1} + \exp(X_{it.}\beta) + \eta_i \tag{92}$$

with w_{ij} being elements from the spatial weight matrix W. Y_{-it} are the observations of all neighbors of i for period t, Y_{t-1} is the vector of N observations for period $t-1$, $X_{it.} = [x_{it1}, \ldots, x_{itK}]$ is a vector of regressors, η_i an individual fixed effect, $i = 1, \ldots, N$, and $t = 1, \ldots, T$. Together with restrictions on the parameters, namely $\gamma \geq 0$, $\lambda \geq 0$, $\rho \geq 0$, the exponential function increases the chance of a positive estimation result for the expectation.

Alternatively, the conditional expectation can be modelled linearly:

$$E[y_{it}|Y_{-it}, Y_{it-1}, X_{it.}, \eta_i] = \gamma y_{it-1} + \lambda \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt} + \rho \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} y_{jt-1} + X_{it.}\beta + \eta_i \tag{93}$$

However, these specifications may lead to negative estimates of the conditional expectation, even when all parameters are restricted to non-negative values. Either the regressors could cause negativity of the conditional expectation (in the linear variant), or the fixed effects could take on negative values. Since count data is non-negative, a procedure to handle this is necessary. If the outcome of the conditional expectation serves as a prediction or forecast, one solution is to replace negative values with zero. The rationale behind this method is that since only zero or positive values of the dependent variable can be observed, a forecasted negative conditional expectation may be interpreted as a clear sign of a zero outcome being the most likely one. Therefore, zero is taken as the prediction or forecast instead of the value of the conditional expectation.

The regression model corresponding to Equations (92) or (93), respectively, is

$$y_{it} = \gamma y_{it-1} + \lambda \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_{jt} + \rho \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_{jt-1} + g(X_{it.}) + \eta_i + u_{it} \tag{94}$$

where $g(X_{it.}) = \exp(X_{it.}\beta)$ for the exponential specification, $g(X_{it.}) = X_{it.}\beta$ for the linear one and u_{it} is an error term. The conditions in Equations (80) - (83) for spatial and serial stationarity described in Section 5.4.1 apply here as well.

One-step ahead forecasts $\hat{y}_{T+1}$ can be calculated as the sum of observed values of the previous period and the forecasted differences for the forecasting period:

$$\widehat{\Delta y}_{T+1} = (I - \hat{\lambda} W)^{-1} (\hat{\gamma} y_T + g(X_{T+1}\hat{\beta})) \tag{95}$$

This leads to non-integer and potentially negative forecasts. Positive forecasts are rounded to the closest integer. A negative forecast value is interpreted as a zero following the rationale described above. The advantage of this forecasting approach is that it does not to need any distributional assumption, which also has not been made for the estimation since GMM is used. A subsequent distributional assumption to obtain integer forecasts (instead of the rounding described above) would therefore seem artificial.

5.5.2 System GMM Estimation

A GMM estimation of an equivalent linear spatial panel model with fixed effects and endogenous regressors was conducted by Kukenova and Monteiro (2009). They use the system GMM estimator of Blundell and Bond (1998), which estimates the difference equation and the level equation simultaneously to obtain less biased results than a corresponding difference GMM estimator.

For the previously introduced Model (94), the difference equation is:

$$\Delta y_{it} = \gamma \Delta y_{it-1} + \lambda \Delta [\sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_{jt}] + \rho \Delta [\sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} y_{jt-1}] + \Delta g(X_{it.}) + \Delta u_{it} \tag{96}$$

With regard to the moment conditions a distinction is made between exogenous (EX) and endogenous explanatory variables (EN). Possible moment conditions for the difference equation are (see Kukenova and Monteiro (2009, p. 10) and Arellano and Bond (1991)):

$$E(y_{i,t-\tau} \Delta u_{it}) = 0 \quad \text{for } t = 3, \ldots, T \text{ and } 2 \leq \tau \leq t-1 \tag{97}$$

$$E(EX_{i,\tau} \Delta u_{it}) = 0 \quad \text{for } t = 3, \ldots, T \text{ and } 1 \leq \tau \leq T \tag{98}$$

$$E(EN_{i,t-\tau} \Delta u_{it}) = 0 \quad \text{for } t = 3, \ldots, T \text{ and } 2 \leq \tau \leq t-1 \tag{99}$$

$$E([W y_{t-\tau}]_i \Delta u_{it}) = 0 \quad \text{for } t = 3, \ldots, T \text{ and } 2 \leq \tau \leq t-1 \tag{100}$$

And the extra moment conditions for extended GMM are:

$$\begin{aligned} E(\Delta y_{i,t-1}u_{it}) &= 0 \quad \text{for } t = 3,\dots,T & (101) \\ E(\Delta EX_{i,t}u_{it}) &= 0 \quad \text{for } t = 2,\dots,T & (102) \\ E(\Delta EN_{i,t-1}u_{it}) &= 0 \quad \text{for } t = 3,\dots,T & (103) \\ E(\Delta[Wy_{t-1}]_i u_{it}) &= 0 \quad \text{for } t = 3,\dots,T & (104) \end{aligned}$$

As before, collapsed instruments are employed (see also Section 5.4.2). That is, moment conditions are summarized over t so that the estimator is asked to minimize, for example, the condition $\sum_t \Delta y_{i,t-2}u_{it}$ instead of the $T-2$ conditions $\Delta y_{i,t-2}u_{it}$ $t = 2,\dots,T$. Estimates are obtained by numerically minimizing the GMM function:

$$gmf = \left(\frac{1}{N}\sum_{i=1}^{N} V_i' Z_i\right) H^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' V_i\right) \tag{105}$$

with the instrument matrix $Z_i = [Z_i^{(1)} Z_i^{(2)}]$ similar to the one proposed by Kukenova and Monteiro (2009):

$$Z_i^{(1)} = \begin{bmatrix} y_{i1} & 0 & 0 & [Wy_1]_i & 0 & 0 & 0 \\ y_{i2} & y_{i1} & 0 & [Wy_2]_i & [Wy_1]_i & 0 & 0 \\ y_{i3} & y_{i2} & 0 & [Wy_3]_i & [Wy_2]_i & [Wy_1]_i & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{iT-2} & y_{iT-3} & 0 & [Wy_{T-2}]_i & [Wy_{T-3}]_i & [Wy_{T-4}]_i & 0 \\ 0 & 0 & \Delta y_{i2} & 0 & 0 & 0 & [W\Delta y_2]_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \Delta y_{iT-1} & 0 & 0 & 0 & [W\Delta y_{T-1}]_i \end{bmatrix}$$

$$Z_i^{(2)} = \begin{bmatrix} EX_{i3} & EN_{i1} & 0 & 0 \\ EX_{i4} & EN_{i2} & EN_{i1} & 0 \\ EX_{i5} & EN_{i3} & EN_{i2} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ EX_{iT} & EN_{iT-2} & EN_{iT-3} & 0 \\ \Delta EX_{i3} & 0 & 0 & \Delta EN_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ \Delta EX_{iT} & 0 & 0 & \Delta EN_{iT-1} \end{bmatrix}$$

where $y_t = [y_{1t},\dots,y_{Nt}]'$, $EX = [EX_{it1},\dots,EX_{itK}]$ is a vector of exogenous and $EN = [EN_{it1},\dots,EN_{itK}]$ of endogenous regressors. The error vector is defined as
$V_i = [\Delta u_{i2}\dots\Delta u_{iT-1}, u_{i2}\dots u_{iT-1}]'$.

The weighting matrix equals $H = \sum_{i=1}^{N} Z_i' H_i^{(k)} Z_i$ with

$$H_i^{(1)} = \begin{bmatrix} H^{(1)D} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \qquad H^{(1)D} = \begin{bmatrix} 1 & -.5 & 0 & \dots & & 0 \\ -.5 & 1 & -.5 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & -.5 \\ 0 & 0 & 0 & \dots & -.5 & 1 \end{bmatrix}$$

for step 1 and

$$H_i^{(2)} = V_i(\hat{\theta}_1)V_i'(\hat{\theta}_1)$$

for step 2. $\hat{\theta}_1 = \{\hat{\gamma}, \hat{\lambda}, \hat{\rho}, \hat{\beta}\}$ are the estimates from step 1.
The asymptotic variance for $\hat{\theta}_2$ is given by

$$\widehat{var}(\hat{\theta}_2) = \frac{1}{N}\Big(C(\hat{\theta}_2)'(H^{(2)}(\hat{\theta}_1))^{-1}C(\hat{\theta}_2)\Big)^{-1}$$

where

$$C(\hat{\theta}_2) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial Z_i' V_i(\theta)}{\partial \theta}\Big|_{\hat{\theta}_2}$$

5.5.3 Illustration with Simulated Data

Like in the Monte Carlo analysis for the dynamic model with multiplicative fixed effects in Section 5.4.3, only a few exemplary configurations can be considered in this Monte Carlo analysis. The simulated data is generated using a DGP similar to the one described in Section 5.4.3 with further adaption from the DGP of Kukenova and Monteiro (2009). For each time t the N observations are generated iteratively. A burn-in of 50 repetitions for each t and 50 time periods applies.
y_{it} is generated using:

$$y_{it} = \gamma y_{i,t-1} + \lambda \sum_{j=1}^{N} w_{ij} y_t + g(x_{it}) + \eta_i + 0.4\epsilon_{it} + e_{it} \tag{106}$$

with $g(x_{it}) = \exp(x_{it}\beta)$ for the non-linear version and $g(x_{it}) = x_{it}\beta$ for the linear one, results are rounded to the nearest non-negative integer. β is set to 1.5 in all configurations. The individual effects are drawn from $\eta \sim i.i.d.\ U(0,1)$ and the errors from $e \sim i.i.d.\ U(0,1)$ and $\epsilon \sim i.i.d.\ U(0,0.1)$. Again, the spatial contiguity weight matrix W is generated using the function `xy2cont` of the Spatial Statistics Toolbox for MATLAB 2.0 (Pace, 2003) on randomly generated coordinates.

N	T	γ	λ	RMSE $\hat{\gamma}$	RMSE $\hat{\lambda}$	RMSE $\hat{\beta}$			
50	50	0.5	0.4	0.037	0.085	0.431			
100	50	0.5	0.4	0.026	0.055	0.326			
100	100	0.5	0.4	0.019	0.054	0.227			
500	100	0.5	0.4	0.013	0.051	0.097			
50	50	0.2	0.1	0.033	0.375	0.488			
100	50	0.2	0.1	0.023	0.234	0.367			
100	100	0.2	0.1	0.015	0.159	0.225			
500	100	0.2	0.1	0.007	0.088	0.104			
50	50	0.2	0.6	0.032	0.131	0.469			
100	50	0.2	0.6	0.023	0.117	0.369			
100	100	0.2	0.6	0.017	0.082	0.219			
500	100	0.2	0.6	0.010	0.050	0.101			
50	50	0.7	0.2	0.035	0.055	0.353			
100	50	0.7	0.2	0.025	0.069	0.304			
100	100	0.7	0.2	0.018	0.078	0.178			
500	100	0.7	0.2	0.009	0.052	0.085			
				Bias			Relative Bias		
N	T	γ	λ	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\beta}$
50	50	0.5	0.4	−0.011	0.060	0.040	−0.022	0.150	0.026
100	50	0.5	0.4	−0.010	0.047	−0.002	−0.021	0.118	−0.001
100	100	0.5	0.4	−0.010	0.048	0.020	−0.020	0.120	0.013
500	100	0.5	0.4	−0.011	0.036	0.016	−0.022	0.089	0.011
50	50	0.2	0.1	0.004	0.282	0.018	0.019	2.823	0.012
100	50	0.2	0.1	0.001	0.139	0.035	0.006	1.389	0.024
100	100	0.2	0.1	0.000	0.061	0.013	−0.001	0.614	0.009
500	100	0.2	0.1	−0.001	−0.003	−0.002	−0.003	−0.027	−0.002
50	50	0.2	0.6	−0.009	0.110	−0.021	−0.043	0.184	−0.014
100	50	0.2	0.6	−0.008	0.079	0.013	−0.038	0.131	0.009
100	100	0.2	0.6	−0.008	0.040	0.014	−0.039	0.066	0.009
500	100	0.2	0.6	−0.007	0.000	0.006	−0.035	0.000	0.004
50	50	0.7	0.2	0.001	0.042	0.008	0.002	0.208	0.005
100	50	0.7	0.2	0.000	0.044	0.004	0.001	0.218	0.003
100	100	0.7	0.2	0.001	0.042	0.019	0.001	0.208	0.013
500	100	0.7	0.2	−0.002	0.021	0.020	−0.003	0.104	0.013

Table 26: Monte Carlo results for the dynamic non-linear spatial panel model with additive fixed effects and $g(X_{it.}) = \exp(X_{it.}\beta)$. The relative bias is calculated as bias divided by true parameter value. Values reported as 0.000 are smaller than 0.0005.

The generation of y_{it} is done iteratively, updating y_{it} after each step $i = 1, \ldots, N$ in vector y_t instantly. After 50 iterations over all units i the last vector is kept as y_t and used as a regressor for generating the outcomes for $t + 1$. For the time dimension a burn-in of 50 iterations is discarded, too. As starting values for the iteration, the $N \times 1$ vector $y_1 = g(X_1) + \eta + 0.4\epsilon_1 + e_1$ is used.

					RMSE	
N	T	γ	λ	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\beta}$
50	50	0.5	0.4	0.038	0.112	0.791
100	50	0.5	0.4	0.026	0.084	0.523
100	100	0.5	0.4	0.020	0.067	0.378
500	100	0.5	0.4	0.014	0.049	0.181
50	50	0.2	0.1	0.032	0.435	1.033
100	50	0.2	0.1	0.022	0.256	0.730
100	100	0.2	0.1	0.015	0.167	0.420
500	100	0.2	0.1	0.007	0.080	0.195
50	50	0.2	0.6	0.033	0.199	0.895
100	50	0.2	0.6	0.024	0.118	0.658
100	100	0.2	0.6	0.017	0.077	0.400
500	100	0.2	0.6	0.007	0.080	0.195
50	50	0.7	0.2	0.037	0.082	0.639
100	50	0.7	0.2	0.025	0.094	0.499
100	100	0.7	0.2	0.018	0.098	0.313
500	100	0.7	0.2	0.009	0.065	0.150

					Bias			Relative Bias	
N	T	γ	λ	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\beta}$
50	50	0.5	0.4	−0.011	0.101	0.059	−0.023	0.253	0.039
100	50	0.5	0.4	−0.009	0.073	0.159	−0.019	0.184	0.106
100	100	0.5	0.4	−0.010	0.051	0.130	−0.021	0.127	0.087
500	100	0.5	0.4	−0.011	0.019	0.091	−0.022	0.048	0.060
50	50	0.2	0.1	−0.001	0.251	0.055	−0.004	2.510	0.037
100	50	0.2	0.1	0.000	0.153	0.192	0.002	1.532	0.128
100	100	0.2	0.1	−0.001	0.068	0.098	−0.007	0.675	0.066
500	100	0.2	0.1	−0.001	0.011	0.004	−0.004	0.107	0.003
50	50	0.2	0.6	−0.012	0.148	0.131	−0.061	0.247	0.087
100	50	0.2	0.6	−0.010	0.075	0.159	−0.051	0.124	0.106
100	100	0.2	0.6	−0.008	0.031	0.093	−0.039	0.051	0.062
500	100	0.2	0.6	−0.001	0.011	0.004	−0.004	0.107	0.003
50	50	0.7	0.2	0.005	0.070	0.072	0.007	0.350	0.048
100	50	0.7	0.2	0.001	0.064	0.065	0.002	0.319	0.044
100	100	0.7	0.2	0.001	0.042	0.061	0.001	0.210	0.041
500	100	0.7	0.2	−0.003	0.015	0.026	−0.004	0.073	0.017

Table 27: Monte Carlo results for the dynamic linear spatial panel model with additive fixed effects and $g(X_{it.}) = X_{it.}\beta$. The relative bias is calculated as bias divided by true parameter value. Values reported as 0.000 are smaller than 0.0005.

The regressors are generated to be autocorrelated, correlated with the individual effects η and endogenous through ϵ:

$$X_{.t} = \delta X_{.t-1} + \tau\eta + \epsilon_t \tag{107}$$

with starting vector $X_{.1} = \frac{\tau}{1-\delta}\eta + \xi$, $\xi \sim i.i.d.\ N(0, \frac{0.5}{1-\delta^2})$, $\tau = 0.1$, and $\delta = 0.5$.

Tables 26 and 27 display the Monte Carlo results for Model (94) without the serial lagged spatial term $\sum_{j=1}^{N} w_{ij} y_{jt-1}$. In both variants, RMSE and bias decrease with increasing sample size $N+T$. $\hat{\beta}$ shows again the highest RMSE which is partly influenced by the much higher true value of 1.5 compared to γ and λ. The bias of the serial correlation parameter γ is mostly negative, as expected, due to the Nickell bias (which is also reported by Kukenova and Monteiro (2009)). Also expectedly, it diminishes with increasing sample size but in some cases turns from positive to negative. In the model with the exponential function, the largest average relative bias is obtained for the spatial correlation parameter λ, which is more sensitive to relative small sample sizes compared to the other parameter estimators. Additionally, a small true λ leads to stronger biased estimates, but in all cases the bias decreases with increasing sample size. In contrast to the Monte Carlo results of the model with multiplicative fixed effects in Section 5.4.3, here the bias of $\hat{\beta}$ is of acceptable size. This generally also holds for the completely linear version of the model, even though its bias of $\hat{\beta}$ is larger compared to the version with the exponential function. Nevertheless, it is still much smaller than in the model with multiplicative fixed effects.

5.5.4 Empirical Application: Forecasts for Pittsburgh's Crime Counts

The number of Part I crimes in the census tracts of Pittsburgh is again forecasted for each month of 2013 using a one-step-ahead expanding window forecast. The lagged Part II crimes are considered as an endogenous variable here and the instrument matrix for the dynamic panel models with additive fixed effects is built accordingly. The estimation sample of the Pittsburgh crime counts data (2008 to 2012) is estimated with Equation (94) in both variants. The results are reported in Tables 28 and 29. For comparison with the previous panel models, both variants have also been estimated with one summer dummy for the months May to September instead of monthly time dummies, even though the discussion below shows that this is not a useful specification here. The respective results are reported in Appendix C.3, Tables 78 - 81.

The Sargan test does not reject the null hypothesis of exogeneity of the instruments for any of the specifications. With exception of Specification 4 of the exponential variant all specifications are stationary with regard to the conditions in Equations (80) - (83) ($\omega_{min} = -0.333$ and $\omega_{max} = 1$). Condition (80) does not hold for Specification 4 in Table 28. Its estimate of the spatial correlation parameter is large which leads together with the slightly increased estimate for the serial autocorrelation (compared to the other specifications) to an instable model. The serial autocorrelation parameter is significant in all specifications. This confirms the usefulness of a dynamic model for this data set. The exponential variant of the model generally estimates a smaller serial autocorrelation than the linear one. In Specification 1, which incorporates all regressors, the simultaneous spatial term is not significant in both variants, but the lagged spatial term is significant. In the linear variant, the simultaneous spatial term turns significant if the lagged spatial term is excluded (Specification 2). If it is significant, its size is much larger than in the SLFP model estimations in Section 5.3. Altogether, the results for the dynamic model specifications do not allow for a clear statement, whether simultaneous spatial correlation is present in the data.

The lagged Part II crimes are surprisingly only significant if either the simultaneous spatial

	1	2	3	4	5
Estimates for January 2008 to December 2012					
$PartI_{t-1}$	0.044**	0.042**	0.040**	0.074***	0.079***
	(0.021)	(0.020)	(0.020)	(0.018)	(0.018)
$WPartI_t$	0.108	0.231		0.955***	0.875***
	(0.191)	(0.188)		(0.108)	(0.105)
$WPartI_{t-1}$	0.089**		0.095***		0.001
	(0.038)		(0.026)		(0.041)
$log(PartII_{t-1})$	0.771***	0.719***	0.807***	−0.282	
	(0.070)	(0.084)	(0.033)	(0.918)	
Feb	−1.702***	−1.284***	−1.700***		−0.176
	(0.352)	(0.338)	(0.321)		(0.431)
Mar	0.240	0.270	0.558		−0.052
	(0.686)	(0.697)	(0.396)		(0.689)
Apr	−0.317	−0.051	−0.043		−0.085
	(0.745)	(0.772)	(0.545)		(0.859)
May	0.002	0.274	0.336		−0.073
	(0.814)	(0.834)	(0.517)		(0.919)
Jun	−0.217	0.059	0.106		−0.205
	(0.838)	(0.877)	(0.592)		(0.962)
Jul	0.291	0.538	0.694		−0.088
	(0.912)	(0.931)	(0.542)		(0.992)
Aug	0.256	0.543	0.646		−0.150
	(0.927)	(0.960)	(0.577)		(1.033)
Sep	−0.582	−0.127	−0.298		−0.255
	(0.806)	(0.833)	(0.562)		(0.947)
Oct	−0.145	0.165	0.169		−0.218
	(0.761)	(0.789)	(0.482)		(0.881)
Nov	−0.523	−0.207	−0.275		−0.255
	(0.672)	(0.690)	(0.464)		(0.785)
Dec	−0.506	−0.270	−0.306		−0.314
	(0.542)	(0.550)	(0.380)		(0.618)
Sargan	0.035	0.071	0.036	0.053	0.022
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Average Forecast Results for January to December 2013					
RMSFE	4.941	4.745	5.126	7.750	4.794
MAFE	3.678	3.520	3.818	6.397	3.546

Table 28: Results from the dynamic non-linear spatial panel model with additive fixed effects for the Pittsburgh crime data and $g(X_{it.}) = \exp(X_{it.}\beta)$. Standard errors in parentheses; p-value of the Sargan test in brackets; W is a queen contiguity spatial weighting matrix; ** and *** denote 5% and 1% statistical significance.

term is not included in the model or is not significant. Apparently, the triggering effect of the Part II crimes can either be measured by including the lagged Part II crimes themselves or by including the outcomes of the Part I crimes in the neighbor tracts, which might have been exposed to the same change in Part II crimes. In models with additive fixed effects, it is possible to leave out the Part II crimes and still estimate the model with individual fixed effects. This is done in Specification 5 in Table 28 which can be counted to the exponential or linear variants of the model. There, the simultaneous spatial correlation parameter is significant and large, but the lagged spatial correlation parameter is insignificant. Lastly, only the time dummy for February has a significant influence on the number of Part I

	1	2	3	4
Estimates for January 2008 to December 2012				
$PartI_{t-1}$	0.068***	0.067***	0.066***	0.071***
	(0.017)	(0.017)	(0.017)	(0.020)
$WPartI_t$	0.148	0.601**		0.919***
	(0.331)	(0.245)		(0.146)
$WPartI_{t-1}$	0.099**		0.114***	
	(0.050)		(0.038)	
$log(PartII_{t-1})$	2.561**	0.916	3.022***	0.345
	(1.085)	(0.799)	(0.325)	(0.664)
Feb	−1.567***	−0.641	−1.688***	
	(0.546)	(0.391)	(0.455)	
Mar	0.200	0.078	0.434	
	(0.686)	(0.679)	(0.510)	
Apr	−0.069	0.004	0.160	
	(0.788)	(0.799)	(0.657)	
May	0.167	0.092	0.469	
	(0.881)	(0.875)	(0.664)	
Jun	0.077	−0.013	0.380	
	(0.930)	(0.932)	(0.708)	
Jul	0.522	0.227	0.910	
	(1.031)	(0.996)	(0.700)	
Aug	0.493	0.230	0.876	
	(1.059)	(1.037)	(0.717)	
Sep	−0.350	−0.155	−0.114	
	(0.849)	(0.865)	(0.723)	
Oct	−0.048	−0.056	0.226	
	(0.825)	(0.828)	(0.634)	
Nov	−0.366	−0.222	−0.175	
	(0.694)	(0.707)	(0.594)	
Dec	−0.493	−0.321	−0.357	
	(0.538)	(0.544)	(0.481)	
Sargan	0.051	0.078	0.046	0.044
	[1.000]	[1.000]	[1.000]	[1.000]
Average Forecast Results for January to December 2013				
RMSFE	4.786	4.661	4.881	5.589
MAFE	3.580	3.435	3.672	4.359

Table 29: Results from the dynamic linear spatial panel model with additive fixed effects for the Pittsburgh crime data and $g(X_{it.}) = X_{it.}\beta$. Standard errors in parentheses; p-value of the Sargan test in brackets; W is a queen contiguity spatial weighting matrix; ** and *** denote 5% and 1% statistical significance.

crimes, all other time dummies are insignificant in all specifications. The estimation and forecasting of the models with linear additive fixed effects have also been repeated using only the dummy for the month February instead of 11 dummies. This improves the forecast results slightly but otherwise does not lead to relevant changes. Therefore the results are not reported here.

Tables 28 and 29 also give the average RMSFE and MAFE from the one-step ahead forecasts for 2013. Single results for all months are given in Appendix C.3, Tables 82 and 83. The results for the single months are in general similar to each other. The overall best forecasts in terms of RMSE and MAFE are obtained from the linear Specification 2, closely followed by exponential Specification 2 and Specification 5, which does not incorporate the lagged Part II crimes. The worse forecast results of specification 3 underline the importance of the simultaneous spatial correlation term even though its significance could not be shown steadily. Even more, the inclusion of time dummies is crucial. Compared to the forecast results of the spatial linear feedback panel model, the forecasts obtained from the dynamic models employed in this section are clearly worse, even though the additional serial correlation in these models is highly significant. Apparently, this additional, relevant variable is outweighed in terms of forecast ability by the less precise GMM estimation.

5.6 Summary

In this chapter, three different models for spatial panel count data and their forecasting performance are investigated. The first approach is an extension of the SLFM model introduced for cross-sectional data in Chapter 4. Individual fixed effects are added into the exponential function of the model in the usual fashion for count data. Estimation of this spatial linear feedback panel model is based on the quasi-maximum likelihood approach discussed in the previous chapter.

The two other alternatives in this chapter are dynamic panel models. The first one is a spatial version of the linear feedback model for counts of Blundell et al. (2002). The model contains multiplicative fixed effects and an additive serial autocorrelation term. The spatial terms are also added linearly and the resulting model is estimated using quasi-differenced GMM. The small Monte Carlo analysis of the model indicates problems in estimating the parameters and the empirical application leads to unrealistic results. Therefore, it must be assumed that the spatial version of the model cannot be estimated with quasi-differenced GMM. Blundell et al. (2002) propose a pre-sample mean estimator instead for the non-spatial LFM. But its applicability is limited since a long time series of the dependent variable previous to the actual estimation sample must be available in order to use this estimator. The third model is developed from a different starting point. It is based on a spatial panel model for continuous data which includes additive fixed effects. In order to apply it to count data, a version with and without an exponential function for the explanatory variables is considered. Estimation is carried out using system GMM as proposed by Kukenova and Monteiro (2009), which leads to good results regarding bias and RMSE.

All three models are applied to a data set of crime counts for Pittsburgh which shows serial and spatial correlation. As mentioned before, estimation of the dynamic model with multiplicative fixed effects does not lead to satisfactory results. Estimation of the other two models works well, though. The Broken Windows theory that the more severe Part I crimes are triggered by Part II crimes can be confirmed for this data. The best forecasts for Pittsburgh's crime counts are obtained from the spatial linear feedback panel model estimated with pseudo maximum likelihood. But nevertheless, the accuracy of the forecasts would have to be improved before a use in practice could be considered. If dynamic structures in the data shall be considered, the panel model with additive fixed effects introduced in this chapter is a good choice, even though it was not originally intended for count data.

6 Conclusion

This thesis introduces new spatial count data models. The models are observation-driven, referring to the concept known from the time series of counts literature. Further, the models fulfill two additional conditions: First, they measure a global spatial effect in the dependent variable. Second, they allow the inclusion of additional explanatory variables into the model. The first condition is motivated by the idea to estimate a parameter of spatial correlation which is comparable to the parameter of serial autocorrelation in time series. The second one allows to study the effect of covariates on the dependent variable, which is a major concern in econometric analysis. Therefore, spatial *econometric* models should be able to do so while dealing with the spatial process as well.

The literature on spatially correlated counts is limited. A spatial correlation in the error term is modelled more often, especially in connection with Bayesian hierarchical models and using a conditional autoregressive structure for the error term. But spatial autoregression is rarely used and none of the modelling approaches introduced so far received broad reception. Most probably, the main reason for the lack of literature on this topic lies in the special structure of count data models compared to the linear models for continuous data. Count data models typically do not establish a direct connection between the regressors and the dependent variable. Instead, the regressors are included in an equation modelling the intensity parameter, which equals the conditional expectation of the dependent variable. Therefore, a direct transfer of the spatial autoregressive structure from these models is not possible. The starting point for this thesis has been the investigation of the spatial autoregressive (SAR) Poisson model for counts, introduced by Lambert et al. (2010). It uses the spatially lagged intensity parameter to introduce a spatial association in the intensity equation. For this model, a limited information maximum likelihood estimation had been proposed by Lambert et al. (2010). But in a Monte Carlo study reported in this thesis it is shown that the model can be estimated using full information maximum likelihood which leads to better results than the limited information maximum likelihood approach. In a second step the spatial structure of the original Poisson model is transferred to other count data models, i.e. negative binomial, zero-inflated Poisson and hurdle Poisson.

In an empirical illustration, these SAR count models are applied to start-up firm births data from the manufacturing industry in the U.S. The cross-sectional data set contains the number of firm births between 2004 to 2007 for each county as well as several characteristics of the counties and their population. The estimations show a significant spatial correlation in the conditional expectations of the firm births. To evaluate the fit of the different count data models employed, the use of scoring rules, known from the non-spatial literature, is proposed. They indicate a better fit of the negative binomial and zero-inflated Poisson model compared to the Poisson model. The data has at the same time many small values (smaller than 4) and some very high values (up to 6938) which is untypical for count data and not well captured by standard count data models. To improve the fit, a next step would be to transfer the spatial structure to more advanced count data models, which are more flexible.

Examples could be a mixture of Poisson, a hurdle model whose hurdle is larger than 0, for example 4, or inflated models, which also inflate for other small outcome values aside from 0.

The SAR models for counts have two important drawbacks: First, the interpretation of their spatial correlation parameter is not intuitive since it measures the spatial dependence in an unobserved process. Second, this measured spatial dependence is only driven by the explanatory variables, since the intensity equation does not contain any random error term. To overcome these issues, a spatial linear feedback model (SLFM) is proposed that includes the spatially lagged dependent variable as a regressor. Because the incorporation of such a term into the exponential function could lead to explosive behaviour if the spatial correlation is positive, it is added linearly into the intensity equation. A reduced form and full maximum likelihood cannot be obtained for this model because an operational multivariate count distribution for a large number of related units does not exist. Therefore, a pseudo maximum likelihood approach is proposed, whose likelihood function consists of the conditional density functions of each outcome conditioned on the respective neighboring outcomes. The appropriateness of this pseudo maximum likelihood estimation is shown in a Monte Carlo study. As an empirical example, the start-up firm births data set is revisited. Nonrandomized PIT residuals, which are already used in the times series analysis of counts, are introduced to the spatial econometrics literature as an additional diagnostic tool. They are in clear favour of the negative binomial spatial linear feedback model for the application at hand.

In the last part of this thesis three spatial panel models for forecasting counts are proposed. Two of them prove to be reliable: the spatial linear feedback panel model (SLFPM), which is an extension of the cross-sectional spatial linear feedback model with multiplicative fixed effects, and a dynamic linear model with additive fixed effects. The former one can be estimated using the aforementioned pseudo maximum likelihood, the latter is estimated using system GMM. Asymptotic properties of the GMM estimation are illustrated with a small Monte Carlo analysis. The models are applied to forecasting crime counts for the census tracts of Pittsburgh. For this, the SLFPM outperforms the dynamic model.

In this thesis spatial models for count data are developed, which incorporate spatially lagged dependent variables to estimate a global spatial effect. Estimation can be carried out using pseudo maximum likelihood in the cross-sectional case and pseudo maximum likelihood or GMM in the panel case. These straightforward and only moderately computation intensive estimation procedures, especially compared to previous propositions in the field, foster the applicability of the models by the empirically working economist. Also, due to the inclusion of lagged observations in the SLFM, rather than lagged intensities, the interpretation of the measured spatial correlation is closer to the one known from linear spatial models for continuous data, making these count data models more accessible. Further research could give attention to extensions of the existing models, e.g. by introducing SEM structures or random spatial weight matrices into the proposed models, or could consider the use of the SLFM structure in more complex count data models, for example Poisson mixture models.

Also, the cross-sectional and panel SLFM should be applied to further economic problems to confirm their general suitability. Nevertheless, a big obstacle for the propagation of spatial count data models is the lack of available count data on a small geographic scale. This is especially true for Europe and leads to few and often slightly abstruse empirical examples in the literature. How fast spatial methods will be established in count data econometrics will – aside from the availability of manageable models, to which this thesis contributes – mainly depend on the availability of suitable economic data.

A Further Results for the SAR Models

A.1 Monte Carlo Parameter Estimates

The first section of this appendix contains the complete results of the parameter estimation from the Monte Carlo study for the P-SAR, NB-SAR and non-spatial Poisson (see also Section 3.4.2). Each table compiles the RMSE and bias for all parameters of one combination of DGP, model and estimation method. For an overview of all employed combinations see Table 1 on Page 37.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.3083	0.1507	0.0553	0.0429	0.1031	0.0546	0.0205	0.0174
0.2	0.3056	0.1188	0.0518	0.0331	0.1263	0.0527	0.0201	0.0135
0.4	0.2049	0.0736	0.0438	0.0252	0.0999	0.0383	0.0217	0.0125
0.6	0.0976	0.2973	0.0191	0.0153	0.0713	0.0580	0.0138	0.0108
0.8	0.0285	0.0133	0.0058	0.0041	0.0384	0.0192	0.0078	0.0054
Bias								
0.0	−0.0035	−0.0077	0.0034	−0.0065	0.0093	0.0008	−0.0001	0.0017
0.2	−0.0509	−0.0256	0.0013	0.0030	0.0278	0.0102	0.0000	−0.0016
0.4	−0.0330	0.0087	−0.0005	0.0005	0.0108	−0.0036	0.0005	0.0002
0.6	−0.0116	−0.1492	−0.0025	0.0006	0.0108	−0.0204	0.0008	0.0001
0.8	0.0037	0.0025	0.0000	−0.0006	−0.0037	−0.0025	−0.0002	0.0007
Relative bias								
0.0	-	-	-	-	0.0926	0.0078	−0.0011	0.0172
0.2	−0.2545	−0.1278	0.0065	0.0150	0.2776	0.1021	0.0002	−0.0165
0.4	−0.0825	0.0218	−0.0012	0.0012	0.1076	−0.0358	0.0051	0.0017
0.6	−0.0194	−0.2487	−0.0042	0.0011	0.1079	−0.2041	0.0079	0.0007
0.8	0.0046	0.0032	0.0000	−0.0007	−0.0373	−0.0245	−0.0018	0.0067
	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.0458	0.0219	0.0095	0.0069	0.0140	0.0054	0.0029	0.0020
0.2	0.0513	0.0229	0.0082	0.0067	0.0127	0.0062	0.0024	0.0019
0.4	0.0377	0.0200	0.0076	0.0055	0.0145	0.0062	0.0027	0.0017
0.6	0.0303	0.0509	0.0065	0.0048	0.0112	0.0492	0.0023	0.0015
0.8	0.0167	0.0081	0.0040	0.0026	0.0084	0.0037	0.0014	0.0011
Bias								
0.0	−0.0116	0.0008	−0.0009	0.0008	−0.0009	−0.0006	−0.0002	−0.0003
0.2	−0.0038	0.0003	−0.0002	0.0007	−0.0021	−0.0001	−0.0001	−0.0001
0.4	0.0014	−0.0014	−0.0003	−0.0003	0.0025	0.0001	0.0000	−0.0002
0.6	−0.0012	−0.0239	0.0009	−0.0005	−0.0004	−0.0235	0.0000	−0.0002
0.8	−0.0007	−0.0005	0.0000	0.0001	−0.0014	−0.0009	0.0001	0.0001
Relative bias								
0.0	−0.1163	0.0078	−0.0090	0.0075	−0.0091	−0.0062	−0.0023	−0.0029
0.2	−0.0376	0.0033	−0.0018	0.0072	−0.0214	−0.0014	−0.0013	−0.0014
0.4	0.0137	−0.0143	−0.0032	−0.0027	0.0250	0.0014	0.0005	−0.0015
0.6	−0.0123	−0.2389	0.0090	−0.0046	−0.0037	−0.2350	0.0000	−0.0019
0.8	−0.0074	−0.0051	0.0000	0.0006	−0.0136	−0.0087	0.0006	0.0008

Table 30: Monte Carlo results of parameter estimates for FIML estimations of P-SAR data and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.4311	0.1890	0.0734	0.0565	0.0901	0.0450	0.0180	0.0152
0.2	0.3246	0.1528	0.0877	0.0726	0.0892	0.0486	0.0224	0.0170
0.4	0.2469	0.1293	0.0888	0.0772	0.0876	0.0501	0.0435	0.0402
0.6	0.1152	0.0647	0.0366	0.0352	0.1211	0.1000	0.0919	0.0927
0.8	0.0900	0.0767	0.0778	0.0775	0.2237	0.2089	0.2121	0.2123
Bias								
0.0	0.0972	0.0115	0.0092	−0.0061	−0.0082	−0.0033	−0.0008	0.0009
0.2	0.0669	0.0344	0.0566	0.0592	0.0182	0.0195	0.0141	0.0124
0.4	0.0696	0.0857	0.0703	0.0714	0.0311	0.0342	0.0389	0.0387
0.6	0.0142	0.0214	0.0271	0.0282	0.1023	0.0947	0.0907	0.0919
0.8	−0.0770	−0.0733	−0.0772	−0.0771	0.2141	0.2069	0.2117	0.2121
Relative bias								
0.0	-	-	-	-	−0.0820	−0.0333	−0.0077	0.0088
0.2	0.3344	0.1721	0.2828	0.2961	0.1824	0.1948	0.1407	0.1242
0.4	0.1740	0.2142	0.1756	0.1785	0.3114	0.3423	0.3894	0.3868
0.6	0.0236	0.0357	0.0452	0.0470	1.0232	0.9467	0.9067	0.9191
0.8	−0.0962	−0.0917	−0.0965	−0.0964	2.1407	2.0693	2.1167	2.1210
	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.0467	0.0216	0.0095	0.0069	0.0140	0.0054	0.0029	0.0020
0.2	0.0515	0.0231	0.0083	0.0066	0.0131	0.0062	0.0025	0.0019
0.4	0.0411	0.0205	0.0086	0.0060	0.0149	0.0061	0.0028	0.0018
0.6	0.0361	0.0166	0.0080	0.0057	0.0111	0.0055	0.0025	0.0016
0.8	0.0260	0.0131	0.0060	0.0036	0.0089	0.0042	0.0017	0.0012
Bias								
0.0	−0.0133	0.0006	−0.0009	0.0007	−0.0009	−0.0006	−0.0002	−0.0003
0.2	−0.0032	0.0004	−0.0003	0.0007	−0.0019	−0.0002	−0.0001	−0.0001
0.4	0.0029	−0.0019	−0.0005	−0.0004	0.0021	0.0000	0.0001	−0.0002
0.6	−0.0035	−0.0005	0.0007	−0.0009	−0.0009	0.0003	−0.0001	−0.0002
0.8	−0.0026	0.0009	0.0008	0.0002	−0.0008	−0.0010	0.0000	0.0000
Relative bias								
0.0	−0.1329	0.0057	−0.0091	0.0074	−0.0086	−0.0056	−0.0023	−0.0029
0.2	−0.0320	0.0038	−0.0034	0.0066	−0.0192	−0.0022	−0.0014	−0.0014
0.4	0.0288	−0.0187	−0.0049	−0.0040	0.0209	0.0001	0.0006	−0.0016
0.6	−0.0348	−0.0053	0.0068	−0.0095	−0.0088	0.0032	−0.0009	−0.0021
0.8	−0.0257	0.0092	0.0076	0.0017	−0.0076	−0.0099	0.0002	0.0001

Table 31: Monte Carlo results of parameter estimates for LIML estimations of P-SAR data and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.3080	0.1507	0.0553	0.0429	0.1030	0.0546	0.0205	0.0174
0.2	0.3057	0.1188	0.0518	0.0331	0.1263	0.0527	0.0201	0.0135
0.4	0.2050	0.0736	0.0437	0.0252	0.0999	0.0383	0.0217	0.0125
0.6	0.0975	0.0483	0.0191	0.0153	0.0714	0.0334	0.0138	0.0108
0.8	0.0285	0.0133	0.0058	0.0041	0.0384	0.0192	0.0078	0.0054
Bias								
0.0	−0.0033	−0.0077	0.0034	−0.0065	0.0092	0.0008	−0.0001	0.0017
0.2	−0.0511	−0.0256	0.0013	0.0030	0.0278	0.0102	0.0000	−0.0016
0.4	−0.0328	0.0087	−0.0005	0.0005	0.0107	−0.0036	0.0005	0.0002
0.6	−0.0119	−0.0040	−0.0025	0.0006	0.0109	0.0031	0.0008	0.0001
0.8	0.0037	0.0026	0.0000	−0.0006	−0.0038	−0.0025	−0.0002	0.0007
Relative bias								
0.0	-	-	-	-	0.0925	0.0078	−0.0010	0.0171
0.2	−0.2553	−0.1281	0.0066	0.0150	0.2777	0.1023	0.0001	−0.0164
0.4	−0.0821	0.0217	−0.0012	0.0012	0.1069	−0.0358	0.0051	0.0017
0.6	−0.0198	−0.0067	−0.0041	0.0011	0.1085	0.0307	0.0079	0.0006
0.8	0.0046	0.0032	0.0000	−0.0007	−0.0376	−0.0247	−0.0018	0.0068

	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.0458	0.0219	0.0095	0.0069	0.0140	0.0054	0.0029	0.0020
0.2	0.0513	0.0229	0.0082	0.0067	0.0127	0.0062	0.0024	0.0019
0.4	0.0377	0.0200	0.0076	0.0055	0.0144	0.0062	0.0027	0.0017
0.6	0.0302	0.0152	0.0065	0.0048	0.0112	0.0052	0.0023	0.0015
0.8	0.0167	0.0081	0.0040	0.0026	0.0084	0.0037	0.0014	0.0011
Bias								
0.0	−0.0117	0.0008	−0.0009	0.0008	−0.0009	−0.0006	−0.0002	−0.0003
0.2	−0.0037	0.0003	−0.0002	0.0007	−0.0021	−0.0001	−0.0001	−0.0001
0.4	0.0014	−0.0014	−0.0003	−0.0003	0.0025	0.0001	0.0000	−0.0002
0.6	−0.0011	−0.0006	0.0009	−0.0005	−0.0003	0.0005	0.0000	−0.0002
0.8	−0.0007	−0.0005	0.0000	0.0001	−0.0014	−0.0009	0.0001	0.0001
Relative bias								
0.0	−0.1168	0.0079	−0.0090	0.0075	−0.0090	−0.0062	−0.0023	−0.0029
0.2	−0.0372	0.0033	−0.0018	0.0072	−0.0212	−0.0014	−0.0013	−0.0014
0.4	0.0138	−0.0143	−0.0032	−0.0027	0.0248	0.0014	0.0005	−0.0015
0.6	−0.0114	−0.0061	0.0090	−0.0046	−0.0034	0.0049	0.0000	−0.0019
0.8	−0.0075	−0.0051	0.0000	0.0006	−0.0136	−0.0087	0.0006	0.0008

	$\hat{\alpha}$			
$\lambda \backslash n$	1000	5000	25000	50000
RMSE				
0.0	0.0206	0.0115	0.0041	0.0026
0.2	0.0143	0.0091	0.0039	0.0029
0.4	0.0168	0.0082	0.0038	0.0025
0.6	0.0122	0.0052	0.0027	0.0020
0.8	0.0050	0.0031	0.0013	0.0009
Bias				
0.0	0.0104	0.0064	0.0025	0.0015
0.2	0.0069	0.0052	0.0021	0.0015
0.4	0.0069	0.0043	0.0021	0.0014
0.6	0.0064	0.0028	0.0014	0.0012
0.8	0.0023	0.0018	0.0008	0.0005

Table 32: Monte Carlo results of parameter estimates for FIML estimations of P-SAR data and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.4306	0.1889	0.0734	0.0564	0.0899	0.0450	0.0180	0.0152
0.2	0.3245	0.1527	0.0877	0.0726	0.0892	0.0486	0.0223	0.0170
0.4	0.2470	0.1292	0.0888	0.0772	0.0875	0.0501	0.0435	0.0402
0.6	0.1154	0.0647	0.0366	0.0352	0.1210	0.1000	0.0919	0.0927
0.8	0.0898	0.0767	0.0778	0.0775	0.2235	0.2089	0.2121	0.2123
Bias								
0.0	0.0975	0.0113	0.0092	−0.0061	−0.0083	−0.0033	−0.0008	0.0009
0.2	0.0669	0.0344	0.0567	0.0592	0.0182	0.0195	0.0140	0.0124
0.4	0.0700	0.0856	0.0702	0.0714	0.0310	0.0343	0.0389	0.0387
0.6	0.0144	0.0214	0.0271	0.0282	0.1022	0.0947	0.0907	0.0919
0.8	−0.0769	−0.0733	−0.0772	−0.0771	0.2139	0.2069	0.2117	0.2121
Relative bias								
0.0	-	-	-	-	−0.0827	−0.0329	−0.0077	0.0088
0.2	0.3345	0.1720	0.2834	0.2961	0.1820	0.1950	0.1401	0.1242
0.4	0.1749	0.2139	0.1756	0.1785	0.3102	0.3426	0.3894	0.3870
0.6	0.0240	0.0357	0.0452	0.0470	1.0217	0.9466	0.9067	0.9189
0.8	−0.0961	−0.0916	−0.0965	−0.0964	2.1394	2.0693	2.1167	2.1210

	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.0467	0.0216	0.0095	0.0069	0.0140	0.0054	0.0029	0.0020
0.2	0.0515	0.0231	0.0083	0.0066	0.0131	0.0062	0.0025	0.0019
0.4	0.0411	0.0205	0.0086	0.0060	0.0149	0.0061	0.0028	0.0018
0.6	0.0360	0.0166	0.0080	0.0057	0.0111	0.0055	0.0025	0.0016
0.8	0.0260	0.0131	0.0060	0.0036	0.0089	0.0042	0.0017	0.0012
Bias								
0.0	−0.0133	0.0006	−0.0009	0.0007	−0.0008	−0.0006	−0.0002	−0.0003
0.2	−0.0032	0.0004	−0.0003	0.0007	−0.0019	−0.0002	−0.0001	−0.0001
0.4	0.0029	−0.0019	−0.0005	−0.0004	0.0021	0.0000	0.0001	−0.0002
0.6	−0.0035	−0.0005	0.0007	−0.0009	−0.0009	0.0003	−0.0001	−0.0002
0.8	−0.0026	0.0009	0.0008	0.0002	−0.0008	−0.0010	0.0000	0.0000
Relative bias								
0.0	−0.1332	0.0056	−0.0091	0.0074	−0.0084	−0.0057	−0.0023	−0.0029
0.2	−0.0317	0.0037	−0.0032	0.0066	−0.0191	−0.0022	−0.0014	−0.0014
0.4	0.0289	−0.0187	−0.0050	−0.0040	0.0207	0.0001	0.0006	−0.0016
0.6	−0.0347	−0.0053	0.0068	−0.0095	−0.0088	0.0033	−0.0009	−0.0021
0.8	−0.0256	0.0091	0.0076	0.0017	−0.0076	−0.0099	0.0002	0.0001

	$\hat{\alpha}$			
$\lambda \backslash n$	1000	5000	25000	50000
RMSE				
0.0	0.0206	0.0114	0.0041	0.0026
0.2	0.0142	0.0091	0.0039	0.0029
0.4	0.0166	0.0081	0.0038	0.0025
0.6	0.0121	0.0052	0.0027	0.0020
0.8	0.0057	0.0038	0.0020	0.0017
Bias				
0.0	0.0104	0.0064	0.0025	0.0015
0.2	0.0068	0.0052	0.0021	0.0015
0.4	0.0069	0.0043	0.0021	0.0014
0.6	0.0064	0.0028	0.0015	0.0012
0.8	0.0028	0.0025	0.0016	0.0013

Table 33: Monte Carlo results of parameter estimates for LIML estimations of P-SAR data and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\beta}_0$				$\hat{\beta}_1$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.0550	0.0271	0.0128	0.0093	0.0460	0.0217	0.0095	0.0069
0.2	0.1062	0.0795	0.0750	0.0733	0.0512	0.0232	0.0082	0.0068
0.4	0.1939	0.1924	0.1956	0.1961	0.0405	0.0207	0.0086	0.0060
0.6	0.4603	0.4372	0.4335	0.4378	0.0344	0.0192	0.0127	0.0100
0.8	1.1535	1.1718	1.1620	1.1686	0.0355	0.0232	0.0231	0.0217
Bias								
0.0	0.0063	−0.0021	0.0009	−0.0003	−0.0106	0.0010	−0.0009	0.0008
0.2	0.0838	0.0745	0.0741	0.0729	−0.0022	0.0012	0.0004	0.0016
0.4	0.1866	0.1909	0.1952	0.1960	0.0062	0.0034	0.0026	0.0019
0.6	0.4581	0.4367	0.4334	0.4378	0.0047	0.0105	0.0103	0.0084
0.8	1.1532	1.1717	1.1620	1.1686	0.0243	0.0198	0.0224	0.0214
Relative bias								
0.0	0.0628	−0.0212	0.0089	−0.0032	−0.1058	0.0098	−0.0087	0.0078
0.2	0.8378	0.7452	0.7413	0.7286	−0.0224	0.0116	0.0043	0.0155
0.4	1.8657	1.9088	1.9523	1.9596	0.0619	0.0336	0.0262	0.0188
0.6	4.5815	4.3667	4.3342	4.3777	0.0468	0.1046	0.1027	0.0838
0.8	11.5315	11.7173	11.6201	11.6857	0.2427	0.1976	0.2244	0.2143

	$\hat{\beta}_2$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**
RMSE				
0.0	0.0135	0.0053	0.0028	0.0020
0.2	0.0119	0.0061	0.0025	0.0020
0.4	0.0152	0.0073	0.0041	0.0036
0.6	0.0132	0.0104	0.0089	0.0086
0.8	0.0194	0.0211	0.0219	0.0212
Bias				
0.0	0.0007	−0.0002	−0.0002	−0.0003
0.2	0.0002	0.0005	0.0006	0.0006
0.4	0.0061	0.0043	0.0033	0.0031
0.6	0.0081	0.0092	0.0086	0.0085
0.8	0.0176	0.0208	0.0219	0.0211
Relative bias				
0.0	0.0068	−0.0023	−0.0018	−0.0026
0.2	0.0019	0.0048	0.0063	0.0059
0.4	0.0613	0.0425	0.0326	0.0315
0.6	0.0812	0.0925	0.0863	0.0851
0.8	0.1761	0.2083	0.2187	0.2114

Table 34: Monte Carlo results of parameter estimates for QML estimations of P-SAR data and non-spatial Poisson model.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.4379	0.5026	0.0649	0.0462	0.1597	0.0548	0.0240	0.0154
0.2	0.3086	0.1176	0.0511	0.0417	0.1366	0.0465	0.0207	0.0166
0.4	0.2259	0.0960	0.0422	0.0263	0.1100	0.0481	0.0216	0.0139
0.6	0.1058	0.0518	0.0238	0.0164	0.0774	0.0370	0.0166	0.0117
0.8	0.0409	0.0159	0.0078	0.0051	0.0538	0.0224	0.0092	0.0070
Bias								
0.0	−0.1021	0.3038	0.0063	0.0053	0.0285	−0.0016	0.0014	−0.0016
0.2	−0.0229	−0.0252	0.0009	−0.0071	0.0170	0.0068	0.0002	0.0037
0.4	−0.0262	−0.0172	−0.0002	−0.0014	0.0064	0.0118	−0.0001	0.0006
0.6	−0.0094	−0.0015	−0.0007	0.0006	0.0110	0.0026	−0.0001	−0.0004
0.8	−0.0022	−0.0016	0.0003	−0.0006	0.0029	0.0042	−0.0006	0.0009
Relative bias								
0.0	-	-	-	-	0.2847	−0.0158	0.0141	−0.0162
0.2	−0.1145	−0.1260	0.0047	−0.0356	0.1700	0.0681	0.0020	0.0365
0.4	−0.0655	−0.0429	−0.0006	−0.0034	0.0638	0.1183	−0.0015	0.0056
0.6	−0.0157	−0.0026	−0.0012	0.0009	0.1101	0.0257	−0.0006	−0.0044
0.8	−0.0027	−0.0020	0.0004	−0.0007	0.0288	0.0416	−0.0057	0.0091

	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.0498	0.0208	0.0097	0.0073	0.0175	0.0082	0.0031	0.0020
0.2	0.0493	0.0206	0.0085	0.0070	0.0144	0.0066	0.0032	0.0020
0.4	0.0410	0.0181	0.0085	0.0060	0.0123	0.0061	0.0029	0.0019
0.6	0.0399	0.0157	0.0076	0.0059	0.0117	0.0064	0.0027	0.0018
0.8	0.0227	0.0104	0.0047	0.0027	0.0089	0.0043	0.0020	0.0013
Bias								
0.0	0.0017	−0.0014	−0.0022	0.0008	−0.0020	0.0004	−0.0006	−0.0002
0.2	−0.0054	0.0024	0.0004	−0.0001	−0.0024	0.0009	−0.0002	−0.0002
0.4	0.0050	−0.0033	0.0012	0.0000	−0.0002	0.0001	−0.0004	0.0000
0.6	−0.0020	−0.0007	0.0008	−0.0001	−0.0001	−0.0003	−0.0002	0.0001
0.8	−0.0009	−0.0020	0.0000	−0.0001	0.0006	0.0003	0.0001	0.0001
Relative bias								
0.0	0.0175	−0.0139	−0.0224	0.0077	−0.0198	0.0037	−0.0059	−0.0018
0.2	−0.0536	0.0238	0.0044	−0.0006	−0.0243	0.0088	−0.0018	−0.0024
0.4	0.0498	−0.0327	0.0116	−0.0001	−0.0016	0.0006	−0.0042	−0.0003
0.6	−0.0203	−0.0074	0.0081	−0.0006	−0.0013	−0.0030	−0.0021	0.0006
0.8	−0.0089	−0.0205	−0.0003	−0.0011	0.0061	0.0028	0.0011	0.0011

Table 35: Monte Carlo results of parameter estimates for FIML estimations of NB-SAR data ($\alpha = 1/8$) and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.4848	0.1936	0.0894	0.0648	0.1029	0.0400	0.0200	0.0129
0.2	0.3938	0.1604	0.0946	0.0779	0.0958	0.0416	0.0230	0.0236
0.4	0.2935	0.1498	0.1076	0.0972	0.0871	0.0642	0.0457	0.0445
0.6	0.1513	0.0951	0.0656	0.0619	0.1292	0.1075	0.0974	0.0964
0.8	0.0890	0.0661	0.0568	0.0592	0.2432	0.2398	0.2310	0.2352
Bias								
0.0	0.0307	0.0219	0.0137	0.0096	−0.0113	−0.0032	0.0010	−0.0017
0.2	0.1712	0.0473	0.0683	0.0559	−0.0002	0.0177	0.0159	0.0193
0.4	0.1256	0.0813	0.0932	0.0911	0.0266	0.0498	0.0418	0.0427
0.6	0.0480	0.0544	0.0561	0.0578	0.1072	0.1001	0.0959	0.0957
0.8	−0.0572	−0.0598	−0.0553	−0.0585	0.2292	0.2370	0.2304	0.2349
Relative bias								
0.0	-	-	-	-	−0.1128	−0.0324	0.0099	−0.0166
0.2	0.8560	0.2365	0.3413	0.2794	−0.0023	0.1768	0.1586	0.1931
0.4	0.3141	0.2032	0.2329	0.2279	0.2664	0.4979	0.4175	0.4272
0.6	0.0801	0.0906	0.0935	0.0963	1.0717	1.0006	0.9594	0.9575
0.8	−0.0715	−0.0748	−0.0691	−0.0731	2.2918	2.3695	2.3041	2.3492
	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.0495	0.0213	0.0096	0.0073	0.0164	0.0073	0.0031	0.0020
0.2	0.0475	0.0209	0.0090	0.0073	0.0141	0.0066	0.0032	0.0020
0.4	0.0459	0.0194	0.0093	0.0067	0.0135	0.0064	0.0030	0.0019
0.6	0.0488	0.0168	0.0085	0.0071	0.0123	0.0069	0.0028	0.0019
0.8	0.0327	0.0144	0.0067	0.0041	0.0108	0.0045	0.0022	0.0016
Bias								
0.0	0.0037	0.0004	−0.0022	0.0008	−0.0014	0.0001	−0.0006	−0.0002
0.2	−0.0049	0.0023	0.0007	0.0000	−0.0024	0.0009	−0.0002	−0.0002
0.4	0.0025	−0.0046	0.0012	−0.0003	−0.0005	0.0000	−0.0004	0.0000
0.6	−0.0047	−0.0020	0.0012	0.0001	−0.0007	−0.0004	−0.0002	0.0001
0.8	0.0026	−0.0002	0.0004	−0.0001	0.0005	−0.0001	−0.0001	0.0001
Relative bias								
0.0	0.0374	0.0039	−0.0223	0.0080	−0.0142	0.0008	−0.0058	−0.0018
0.2	−0.0495	0.0233	0.0065	0.0005	−0.0238	0.0091	−0.0021	−0.0025
0.4	0.0253	−0.0458	0.0120	−0.0031	−0.0055	−0.0003	−0.0041	−0.0002
0.6	−0.0470	−0.0203	0.0123	0.0013	−0.0068	−0.0042	−0.0020	0.0009
0.8	0.0261	−0.0017	0.0037	−0.0006	0.0055	−0.0010	−0.0006	0.0015

Table 36: Monte Carlo results of parameter estimates for LIML estimations of NB-SAR data ($\alpha = 1/8$) and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.4352	0.1457	0.0649	0.0462	0.1586	0.0501	0.0240	0.0155
0.2	0.3090	0.1172	0.0510	0.0417	0.1368	0.0462	0.0207	0.0165
0.4	0.2257	0.0964	0.0422	0.0261	0.1096	0.0482	0.0216	0.0138
0.6	0.1057	0.0520	0.0238	0.0163	0.0772	0.0371	0.0167	0.0116
0.8	0.0411	0.0156	0.0076	0.0050	0.0533	0.0220	0.0091	0.0069
Bias								
0.0	−0.0994	−0.0021	0.0063	0.0053	0.0276	0.0016	0.0014	−0.0016
0.2	−0.0220	−0.0244	0.0010	−0.0072	0.0166	0.0065	0.0002	0.0037
0.4	−0.0272	−0.0176	−0.0003	−0.0014	0.0070	0.0121	−0.0002	0.0006
0.6	−0.0088	−0.0016	−0.0007	0.0005	0.0106	0.0028	−0.0001	−0.0004
0.8	−0.0025	−0.0013	0.0003	−0.0005	0.0028	0.0036	−0.0005	0.0008
Relative bias								
0.0	-	-	-	-	0.2759	0.0157	0.0142	−0.0164
0.2	−0.1099	−0.1220	0.0051	−0.0360	0.1662	0.0652	0.0020	0.0371
0.4	−0.0680	−0.0440	−0.0006	−0.0035	0.0700	0.1212	−0.0016	0.0057
0.6	−0.0146	−0.0027	−0.0012	0.0008	0.1057	0.0279	−0.0007	−0.0036
0.8	−0.0031	−0.0016	0.0004	−0.0006	0.0285	0.0363	−0.0054	0.0082
	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.0497	0.0216	0.0097	0.0073	0.0174	0.0072	0.0031	0.0020
0.2	0.0490	0.0205	0.0085	0.0070	0.0144	0.0066	0.0032	0.0020
0.4	0.0411	0.0180	0.0085	0.0059	0.0123	0.0060	0.0029	0.0019
0.6	0.0397	0.0155	0.0076	0.0059	0.0117	0.0064	0.0027	0.0018
0.8	0.0224	0.0103	0.0047	0.0027	0.0087	0.0043	0.0020	0.0013
Bias								
0.0	0.0018	−0.0001	−0.0023	0.0008	−0.0020	0.0000	−0.0006	−0.0002
0.2	−0.0053	0.0024	0.0004	−0.0001	−0.0024	0.0008	−0.0002	−0.0002
0.4	0.0050	−0.0034	0.0012	0.0000	−0.0003	0.0001	−0.0004	0.0000
0.6	−0.0021	−0.0008	0.0008	−0.0001	−0.0002	−0.0004	−0.0002	0.0001
0.8	−0.0007	−0.0021	−0.0001	−0.0001	0.0009	0.0003	0.0001	0.0001
Relative bias								
0.0	0.0175	−0.0011	−0.0226	0.0080	−0.0195	0.0005	−0.0058	−0.0018
0.2	−0.0535	0.0243	0.0040	−0.0010	−0.0238	0.0083	−0.0017	−0.0023
0.4	0.0502	−0.0340	0.0117	0.0002	−0.0028	0.0012	−0.0041	−0.0004
0.6	−0.0206	−0.0083	0.0082	−0.0006	−0.0016	−0.0037	−0.0020	0.0005
0.8	−0.0074	−0.0206	−0.0010	−0.0014	0.0086	0.0029	0.0011	0.0009
	$\hat{\alpha}$				$\hat{\alpha}$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE					**Relative bias**			
0.0	0.0430	0.0175	0.0081	0.0062	−0.1137	−0.0005	−0.0107	−0.0017
0.2	0.0351	0.0154	0.0084	0.0052	−0.0178	0.0166	−0.0097	−0.0017
0.4	0.0315	0.0153	0.0073	0.0048	−0.0792	−0.0323	0.0033	−0.0052
0.6	0.0261	0.0113	0.0054	0.0038	−0.0373	−0.0224	0.0009	0.0082
0.8	0.0139	0.0067	0.0033	0.0022	−0.0073	−0.0023	0.0050	0.0021
Bias								
0.0	−0.0142	−0.0001	−0.0013	−0.0002				
0.2	−0.0022	0.0021	−0.0012	−0.0002				
0.4	−0.0099	−0.0040	0.0004	−0.0006				
0.6	−0.0047	−0.0028	0.0001	0.0010				
0.8	−0.0009	−0.0003	0.0006	0.0003				

Table 37: Monte Carlo results of parameter estimates for FIML estimations of NB-SAR data ($\alpha = 1/8$) and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.4839	0.1939	0.0893	0.0649	0.1024	0.0403	0.0199	0.0129
0.2	0.3947	0.1605	0.0947	0.0777	0.0956	0.0413	0.0231	0.0236
0.4	0.2916	0.1500	0.1076	0.0970	0.0872	0.0644	0.0457	0.0445
0.6	0.1502	0.0947	0.0656	0.0617	0.1282	0.1079	0.0974	0.0965
0.8	0.0881	0.0651	0.0568	0.0591	0.2409	0.2391	0.2313	0.2351
Bias								
0.0	0.0330	0.0213	0.0137	0.0096	−0.0117	−0.0030	0.0010	−0.0017
0.2	0.1736	0.0483	0.0684	0.0557	−0.0008	0.0175	0.0158	0.0194
0.4	0.1231	0.0809	0.0932	0.0910	0.0275	0.0500	0.0417	0.0427
0.6	0.0491	0.0536	0.0561	0.0576	0.1066	0.1006	0.0959	0.0959
0.8	−0.0562	−0.0595	−0.0554	−0.0585	0.2278	0.2366	0.2308	0.2349
Relative bias								
0.0	-	-	-	-	−0.1175	−0.0296	0.0100	−0.0169
0.2	0.8682	0.2417	0.3422	0.2787	−0.0084	0.1745	0.1584	0.1936
0.4	0.3078	0.2022	0.2329	0.2276	0.2754	0.5002	0.4174	0.4274
0.6	0.0818	0.0894	0.0936	0.0961	1.0656	1.0059	0.9588	0.9586
0.8	−0.0703	−0.0743	−0.0693	−0.0731	2.2779	2.3657	2.3082	2.3489

	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.0495	0.0214	0.0096	0.0073	0.0163	0.0073	0.0031	0.0020
0.2	0.0472	0.0208	0.0089	0.0073	0.0141	0.0066	0.0032	0.0020
0.4	0.0461	0.0193	0.0092	0.0067	0.0135	0.0063	0.0030	0.0019
0.6	0.0486	0.0166	0.0085	0.0071	0.0122	0.0069	0.0028	0.0019
0.8	0.0326	0.0142	0.0067	0.0041	0.0106	0.0046	0.0022	0.0015
Bias								
0.0	0.0038	0.0002	−0.0023	0.0008	−0.0014	0.0001	−0.0006	−0.0002
0.2	−0.0049	0.0024	0.0006	0.0000	−0.0023	0.0009	−0.0002	−0.0002
0.4	0.0026	−0.0047	0.0012	−0.0003	−0.0007	0.0000	−0.0004	0.0000
0.6	−0.0046	−0.0020	0.0013	0.0001	−0.0007	−0.0005	−0.0002	0.0001
0.8	0.0025	−0.0003	0.0002	−0.0001	0.0008	−0.0001	−0.0001	0.0001
Relative bias								
0.0	0.0378	0.0025	−0.0226	0.0083	−0.0141	0.0006	−0.0057	−0.0018
0.2	−0.0491	0.0239	0.0062	0.0001	−0.0232	0.0086	−0.0020	−0.0024
0.4	0.0256	−0.0474	0.0120	−0.0028	−0.0067	0.0003	−0.0040	−0.0003
0.6	−0.0463	−0.0204	0.0126	0.0011	−0.0069	−0.0049	−0.0020	0.0008
0.8	0.0251	−0.0026	0.0018	−0.0006	0.0079	−0.0012	−0.0008	0.0012

	$\hat{\alpha}$				$\hat{\alpha}$			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE					**Relative bias**			
0.0	0.0432	0.0175	0.0081	0.0062	−0.1124	−0.0005	−0.0107	−0.0017
0.2	0.0351	0.0153	0.0084	0.0052	−0.0200	0.0163	−0.0098	−0.0017
0.4	0.0317	0.0154	0.0073	0.0048	−0.0812	−0.0329	0.0032	−0.0052
0.6	0.0262	0.0112	0.0054	0.0038	−0.0396	−0.0224	0.0012	0.0086
0.8	0.0137	0.0068	0.0037	0.0026	0.0033	0.0063	0.0144	0.0115
Bias								
0.0	−0.0140	−0.0001	−0.0013	−0.0002				
0.2	−0.0025	0.0020	−0.0012	−0.0002				
0.4	−0.0102	−0.0041	0.0004	−0.0007				
0.6	−0.0049	−0.0028	0.0002	0.0011				
0.8	0.0004	0.0008	0.0018	0.0014				

Table 38: Monte Carlo results of parameter estimates for LIML estimations of NB-SAR data ($\alpha = 1/8$) and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\beta}_0$				$\hat{\beta}_1$			
$\lambda\backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.0625	0.0274	0.0133	0.0084	0.0499	0.0214	0.0096	0.0073
0.2	0.1026	0.0761	0.0749	0.0751	0.0487	0.0208	0.0088	0.0072
0.4	0.1943	0.1966	0.1952	0.1956	0.0447	0.0186	0.0099	0.0068
0.6	0.4662	0.4402	0.4338	0.4370	0.0452	0.0189	0.0133	0.0114
0.8	1.1472	1.1723	1.1626	1.1694	0.0418	0.0229	0.0229	0.0213
Bias								
0.0	−0.0072	−0.0001	0.0032	−0.0001	0.0053	0.0005	−0.0022	0.0008
0.2	0.0818	0.0716	0.0742	0.0746	−0.0042	0.0030	0.0012	0.0008
0.4	0.1865	0.1952	0.1948	0.1955	0.0078	0.0006	0.0043	0.0020
0.6	0.4633	0.4396	0.4337	0.4369	0.0024	0.0094	0.0106	0.0092
0.8	1.1464	1.1722	1.1626	1.1694	0.0276	0.0181	0.0222	0.0210
Relative bias								
0.0	−0.0723	−0.0009	0.0315	−0.0008	0.0529	0.0054	−0.0218	0.0082
0.2	0.8180	0.7159	0.7416	0.7457	−0.0422	0.0304	0.0120	0.0077
0.4	1.8647	1.9524	1.9483	1.9546	0.0781	0.0056	0.0429	0.0200
0.6	4.6334	4.3959	4.3367	4.3689	0.0243	0.0938	0.1059	0.0918
0.8	11.4642	11.7220	11.6261	11.6936	0.2764	0.1808	0.2215	0.2099

	$\hat{\beta}_2$			
$\lambda\backslash n$	1000	5000	25000	50000
RMSE				
0.0	0.0160	0.0072	0.0031	0.0020
0.2	0.0140	0.0067	0.0033	0.0019
0.4	0.0127	0.0069	0.0039	0.0037
0.6	0.0140	0.0104	0.0089	0.0089
0.8	0.0210	0.0221	0.0220	0.0213
Bias				
0.0	0.0005	0.0004	−0.0005	−0.0001
0.2	0.0002	0.0015	0.0006	0.0004
0.4	0.0038	0.0038	0.0028	0.0033
0.6	0.0086	0.0086	0.0085	0.0088
0.8	0.0190	0.0217	0.0219	0.0212
Relative bias				
0.0	0.0054	0.0042	−0.0052	−0.0014
0.2	0.0025	0.0152	0.0057	0.0042
0.4	0.0385	0.0376	0.0277	0.0325
0.6	0.0864	0.0857	0.0853	0.0878
0.8	0.1905	0.2173	0.2193	0.2124

Table 39: Monte Carlo results of parameter estimates for QML estimations of NB-SAR data ($\alpha = 1/8$) and non-spatial Poisson model.

$\lambda\backslash n$	$\hat{\lambda}$ 1000	5000	25000	50000	$\hat{\beta}_0$ 1000	5000	25000	50000
RMSE								
0.0	0.4796	0.1689	0.0723	0.0543	0.1649	0.0631	0.0254	0.0188
0.2	0.3956	0.1635	0.0593	0.0468	0.1626	0.0745	0.0253	0.0185
0.4	0.3197	0.1100	0.0537	0.0328	0.1546	0.0576	0.0280	0.0173
0.6	0.1471	0.0741	0.0322	0.0197	0.0971	0.0540	0.0235	0.0139
0.8	0.0560	0.0232	0.0114	0.0083	0.0696	0.0310	0.0141	0.0110
Bias								
0.0	−0.0254	0.0114	0.0032	0.0016	0.0147	0.0004	−0.0004	0.0009
0.2	−0.0785	−0.0484	−0.0112	0.0005	0.0393	0.0205	0.0057	−0.0005
0.4	−0.1054	−0.0028	0.0009	−0.0008	0.0397	0.0039	−0.0012	0.0005
0.6	−0.0205	−0.0086	0.0041	0.0027	0.0025	0.0055	−0.0016	−0.0016
0.8	−0.0098	−0.0009	−0.0019	−0.0016	0.0165	0.0026	0.0013	0.0025
Relative bias								
0.0	-	-	-	-	0.1470	0.0041	−0.0042	0.0088
0.2	−0.3926	−0.2420	−0.0560	0.0026	0.3929	0.2051	0.0567	−0.0055
0.4	−0.2634	−0.0069	0.0023	−0.0019	0.3975	0.0394	−0.0122	0.0051
0.6	−0.0342	−0.0143	0.0068	0.0045	0.0251	0.0551	−0.0158	−0.0156
0.8	−0.0123	−0.0011	−0.0023	−0.0020	0.1649	0.0257	0.0127	0.0245

$\lambda\backslash n$	$\hat{\beta}_1$ 1000	5000	25000	50000	$\hat{\beta}_2$ 1000	5000	25000	50000
RMSE								
0.0	0.0643	0.0250	0.0102	0.0074	0.0205	0.0076	0.0038	0.0024
0.2	0.0625	0.0282	0.0116	0.0091	0.0188	0.0078	0.0036	0.0028
0.4	0.0547	0.0219	0.0106	0.0077	0.0175	0.0068	0.0033	0.0026
0.6	0.0450	0.0200	0.0083	0.0066	0.0188	0.0074	0.0036	0.0022
0.8	0.0388	0.0158	0.0070	0.0047	0.0148	0.0067	0.0028	0.0020
Bias								
0.0	−0.0049	−0.0016	−0.0006	−0.0011	−0.0045	−0.0010	−0.0006	−0.0003
0.2	−0.0041	−0.0013	−0.0002	0.0009	−0.0069	−0.0009	−0.0002	−0.0002
0.4	0.0048	−0.0018	0.0003	0.0004	0.0013	−0.0009	0.0000	−0.0001
0.6	0.0084	0.0008	−0.0016	−0.0007	0.0019	−0.0003	−0.0002	0.0001
0.8	−0.0026	−0.0011	0.0009	−0.0001	0.0003	−0.0001	0.0006	0.0000
Relative bias								
0.0	−0.0491	−0.0160	−0.0058	−0.0106	−0.0449	−0.0099	−0.0061	−0.0027
0.2	−0.0410	−0.0126	−0.0018	0.0088	−0.0686	−0.0095	−0.0021	−0.0024
0.4	0.0483	−0.0184	0.0027	0.0041	0.0132	−0.0091	0.0000	−0.0008
0.6	0.0842	0.0077	−0.0155	−0.0069	0.0193	−0.0026	−0.0022	0.0009
0.8	−0.0255	−0.0107	0.0092	−0.0012	0.0030	−0.0012	0.0057	−0.0002

Table 40: Monte Carlo results of parameter estimates for FIML estimations of NB-SAR data ($\alpha = 1/2$) and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\boldsymbol{\lambda}}$				$\hat{\boldsymbol{\beta}}_0$			
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.5971	0.2440	0.1133	0.0829	0.0944	0.0464	0.0191	0.0139
0.2	0.5621	0.2257	0.1233	0.1227	0.1093	0.0640	0.0334	0.0272
0.4	0.3901	0.2273	0.1745	0.1577	0.1210	0.0739	0.0591	0.0578
0.6	0.2587	0.1664	0.1483	0.1493	0.1474	0.1261	0.1167	0.1110
0.8	0.1158	0.0504	0.0243	0.0211	0.2946	0.2900	0.2752	0.2761
Bias								
0.0	0.2185	0.0422	0.0165	0.0059	−0.0197	−0.0014	−0.0016	0.0006
0.2	0.1751	0.0511	0.0867	0.1002	0.0180	0.0339	0.0269	0.0227
0.4	0.0908	0.1621	0.1563	0.1494	0.0582	0.0556	0.0543	0.0559
0.6	0.1170	0.1277	0.1413	0.1462	0.1141	0.1161	0.1148	0.1102
0.8	0.0182	0.0043	0.0120	0.0137	0.2684	0.2851	0.2743	0.2756
Relative bias								
0.0	-	-	-	-	−0.1968	−0.0141	−0.0156	0.0064
0.2	0.8757	0.2553	0.4335	0.5011	0.1805	0.3387	0.2686	0.2265
0.4	0.2269	0.4051	0.3908	0.3736	0.5819	0.5561	0.5427	0.5592
0.6	0.1950	0.2128	0.2355	0.2437	1.1407	1.1609	1.1484	1.1019
0.8	0.0227	0.0054	0.0150	0.0171	2.6843	2.8514	2.7425	2.7562
	$\hat{\boldsymbol{\beta}}_1$				$\hat{\boldsymbol{\beta}}_2$			
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.0618	0.0252	0.0103	0.0074	0.0187	0.0077	0.0037	0.0024
0.2	0.0639	0.0293	0.0119	0.0091	0.0189	0.0078	0.0036	0.0028
0.4	0.0544	0.0228	0.0123	0.0082	0.0179	0.0068	0.0033	0.0027
0.6	0.0536	0.0229	0.0094	0.0079	0.0193	0.0080	0.0038	0.0022
0.8	0.0534	0.0246	0.0095	0.0076	0.0176	0.0078	0.0033	0.0023
Bias								
0.0	−0.0035	−0.0015	−0.0005	−0.0011	−0.0034	−0.0009	−0.0006	−0.0003
0.2	−0.0027	−0.0013	−0.0003	0.0009	−0.0072	−0.0010	−0.0002	−0.0003
0.4	0.0063	−0.0024	−0.0001	0.0008	0.0013	−0.0011	0.0000	−0.0002
0.6	0.0054	0.0011	−0.0027	0.0000	0.0011	−0.0001	0.0001	0.0000
0.8	0.0010	−0.0003	0.0009	−0.0014	−0.0016	0.0007	0.0007	0.0001
Relative bias								
0.0	−0.0347	−0.0150	−0.0048	−0.0108	−0.0338	−0.0092	−0.0061	−0.0027
0.2	−0.0268	−0.0134	−0.0031	0.0095	−0.0716	−0.0096	−0.0023	−0.0026
0.4	0.0629	−0.0241	−0.0008	0.0083	0.0135	−0.0109	0.0001	−0.0016
0.6	0.0543	0.0111	−0.0269	−0.0002	0.0106	−0.0010	0.0012	−0.0002
0.8	0.0101	−0.0031	0.0093	−0.0144	−0.0157	0.0067	0.0068	0.0010

Table 41: Monte Carlo results of parameter estimates for LIML estimations of NB-SAR data ($\alpha = 1/2$) and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.4837	0.1696	0.0708	0.0542	0.1654	0.0636	0.0251	0.0189
0.2	0.4015	0.1641	0.0596	0.0471	0.1651	0.0756	0.0255	0.0184
0.4	0.3241	0.1082	0.0534	0.0330	0.1573	0.0562	0.0278	0.0174
0.6	0.1507	0.0739	0.0324	0.0198	0.1024	0.0536	0.0237	0.0139
0.8	0.0579	0.0230	0.0111	0.0080	0.0704	0.0308	0.0138	0.0106
Bias								
0.0	−0.0235	0.0127	0.0041	0.0008	0.0138	0.0001	−0.0007	0.0010
0.2	−0.0854	−0.0481	−0.0116	0.0003	0.0427	0.0202	0.0057	−0.0004
0.4	−0.1063	−0.0011	0.0012	−0.0008	0.0410	0.0026	−0.0014	0.0005
0.6	−0.0209	−0.0087	0.0043	0.0026	0.0038	0.0054	−0.0015	−0.0017
0.8	−0.0094	0.0005	−0.0016	−0.0016	0.0138	0.0011	0.0012	0.0023
Relative bias								
0.0	-	-	-	-	0.1381	0.0014	−0.0075	0.0095
0.2	−0.4268	−0.2404	−0.0581	0.0016	0.4265	0.2022	0.0567	−0.0035
0.4	−0.2657	−0.0027	0.0031	−0.0020	0.4104	0.0264	−0.0135	0.0045
0.6	−0.0349	−0.0146	0.0071	0.0043	0.0381	0.0544	−0.0153	−0.0168
0.8	−0.0117	0.0006	−0.0021	−0.0019	0.1383	0.0107	0.0119	0.0227

	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.0628	0.0251	0.0103	0.0073	0.0206	0.0076	0.0037	0.0025
0.2	0.0637	0.0282	0.0117	0.0088	0.0185	0.0079	0.0035	0.0027
0.4	0.0553	0.0214	0.0107	0.0077	0.0174	0.0067	0.0033	0.0026
0.6	0.0448	0.0201	0.0083	0.0066	0.0182	0.0073	0.0035	0.0022
0.8	0.0376	0.0152	0.0068	0.0046	0.0145	0.0065	0.0027	0.0019
Bias								
0.0	−0.0043	−0.0016	−0.0005	−0.0010	−0.0049	−0.0011	−0.0006	−0.0002
0.2	−0.0054	−0.0011	0.0000	0.0007	−0.0066	−0.0009	−0.0002	−0.0002
0.4	0.0046	−0.0014	0.0003	0.0005	0.0008	−0.0008	−0.0001	−0.0001
0.6	0.0076	0.0009	−0.0017	−0.0006	0.0018	−0.0003	−0.0003	0.0001
0.8	−0.0006	−0.0016	0.0007	0.0000	0.0003	−0.0002	0.0005	0.0000
Relative bias								
0.0	−0.0428	−0.0156	−0.0053	−0.0098	−0.0486	−0.0113	−0.0060	−0.0021
0.2	−0.0536	−0.0112	0.0000	0.0075	−0.0660	−0.0090	−0.0023	−0.0023
0.4	0.0462	−0.0145	0.0033	0.0048	0.0083	−0.0085	−0.0007	−0.0006
0.6	0.0761	0.0094	−0.0169	−0.0055	0.0183	−0.0025	−0.0026	0.0014
0.8	−0.0063	−0.0156	0.0074	−0.0003	0.0029	−0.0023	0.0052	0.0004

	$\hat{\alpha}$				$\hat{\alpha}$			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE					**Relative bias**			
0.0	0.0607	0.0281	0.0123	0.0084	−0.0138	0.0062	0.0040	−0.0004
0.2	0.0581	0.0267	0.0110	0.0083	−0.0405	−0.0086	−0.0032	0.0013
0.4	0.0610	0.0221	0.0125	0.0083	0.0069	−0.0047	0.0049	−0.0011
0.6	0.0485	0.0212	0.0089	0.0069	−0.0013	−0.0128	−0.0021	−0.0008
0.8	0.0341	0.0140	0.0065	0.0050	−0.0077	0.0013	−0.0006	0.0006
Bias								
0.0	−0.0069	0.0031	0.0020	−0.0002				
0.2	−0.0203	−0.0043	−0.0016	0.0007				
0.4	0.0035	−0.0024	0.0024	−0.0006				
0.6	−0.0006	−0.0064	−0.0010	−0.0004				
0.8	−0.0039	0.0006	−0.0003	0.0003				

Table 42: Monte Carlo results of parameter estimates for FIML estimations of NB-SAR data ($\alpha = 1/2$) and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.6003	0.2452	0.1111	0.0828	0.0943	0.0470	0.0189	0.0140
0.2	0.5678	0.2238	0.1227	0.1227	0.1111	0.0645	0.0333	0.0272
0.4	0.3897	0.2288	0.1746	0.1577	0.1220	0.0723	0.0590	0.0578
0.6	0.2616	0.1664	0.1490	0.1496	0.1501	0.1260	0.1168	0.1106
0.8	0.1130	0.0472	0.0232	0.0208	0.2878	0.2889	0.2755	0.2753
Bias								
0.0	0.2205	0.0441	0.0180	0.0049	−0.0205	−0.0015	−0.0018	0.0006
0.2	0.1671	0.0510	0.0861	0.1000	0.0201	0.0337	0.0268	0.0228
0.4	0.0915	0.1658	0.1566	0.1494	0.0587	0.0542	0.0542	0.0558
0.6	0.1179	0.1275	0.1418	0.1466	0.1144	0.1161	0.1149	0.1098
0.8	0.0204	0.0057	0.0120	0.0142	0.2642	0.2847	0.2746	0.2749
Relative bias								
0.0	-	-	-	-	−0.2046	−0.0148	−0.0180	0.0061
0.2	0.8357	0.2548	0.4306	0.5001	0.2005	0.3373	0.2680	0.2281
0.4	0.2287	0.4145	0.3915	0.3736	0.5873	0.5419	0.5421	0.5582
0.6	0.1966	0.2125	0.2363	0.2443	1.1437	1.1605	1.1486	1.0983
0.8	0.0255	0.0072	0.0151	0.0177	2.6418	2.8470	2.7464	2.7490

	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE								
0.0	0.0609	0.0254	0.0104	0.0073	0.0186	0.0077	0.0037	0.0025
0.2	0.0648	0.0293	0.0120	0.0089	0.0186	0.0079	0.0035	0.0027
0.4	0.0547	0.0225	0.0124	0.0082	0.0179	0.0066	0.0033	0.0027
0.6	0.0522	0.0231	0.0094	0.0078	0.0187	0.0080	0.0038	0.0022
0.8	0.0522	0.0238	0.0095	0.0075	0.0176	0.0075	0.0032	0.0022
Bias								
0.0	−0.0028	−0.0015	−0.0004	−0.0010	−0.0037	−0.0011	−0.0006	−0.0002
0.2	−0.0038	−0.0012	−0.0001	0.0008	−0.0070	−0.0009	−0.0003	−0.0003
0.4	0.0062	−0.0020	0.0000	0.0009	0.0009	−0.0010	−0.0001	−0.0001
0.6	0.0050	0.0012	−0.0029	0.0001	0.0009	−0.0001	0.0001	0.0000
0.8	0.0028	−0.0013	0.0006	−0.0013	−0.0018	0.0006	0.0006	0.0001
Relative bias								
0.0	−0.0275	−0.0147	−0.0042	−0.0101	−0.0370	−0.0106	−0.0061	−0.0021
0.2	−0.0378	−0.0122	−0.0012	0.0081	−0.0696	−0.0092	−0.0026	−0.0025
0.4	0.0623	−0.0204	−0.0001	0.0090	0.0089	−0.0103	−0.0006	−0.0014
0.6	0.0496	0.0124	−0.0288	0.0013	0.0088	−0.0010	0.0007	0.0003
0.8	0.0279	−0.0135	0.0062	−0.0134	−0.0179	0.0059	0.0058	0.0014

	$\hat{\alpha}$				$\hat{\alpha}$			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
RMSE					**Relative bias**			
0.0	0.0606	0.0281	0.0123	0.0084	−0.0135	0.0062	0.0040	−0.0004
0.2	0.0584	0.0267	0.0110	0.0083	−0.0414	−0.0087	−0.0033	0.0013
0.4	0.0609	0.0221	0.0125	0.0083	0.0063	−0.0049	0.0048	−0.0012
0.6	0.0487	0.0213	0.0089	0.0069	−0.0018	−0.0127	−0.0020	−0.0007
0.8	0.0338	0.0142	0.0066	0.0052	−0.0057	0.0038	0.0018	0.0030
Bias								
0.0	−0.0067	0.0031	0.0020	−0.0002				
0.2	−0.0207	−0.0043	−0.0016	0.0006				
0.4	0.0031	−0.0025	0.0024	−0.0006				
0.6	−0.0009	−0.0064	−0.0010	−0.0003				
0.8	−0.0028	0.0019	0.0009	0.0015				

Table 43: Monte Carlo results of parameter estimates for LIML estimations of NB-SAR data ($\alpha = 1/2$) and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\beta}_0$				$\hat{\beta}_1$			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
RMSE								
0.0	0.0711	0.0340	0.0130	0.0097	0.0629	0.0251	0.0102	0.0074
0.2	0.1089	0.0844	0.0768	0.0739	0.0645	0.0291	0.0119	0.0092
0.4	0.1919	0.1955	0.1945	0.1959	0.0560	0.0229	0.0123	0.0086
0.6	0.4446	0.4359	0.4362	0.4370	0.0512	0.0251	0.0114	0.0117
0.8	1.1548	1.1726	1.1615	1.1697	0.0582	0.0289	0.0246	0.0214
Bias								
0.0	0.0024	0.0029	0.0003	0.0013	−0.0018	−0.0011	−0.0005	−0.0010
0.2	0.0807	0.0763	0.0754	0.0730	−0.0021	−0.0004	0.0004	0.0018
0.4	0.1782	0.1932	0.1939	0.1957	0.0102	0.0029	0.0032	0.0029
0.6	0.4398	0.4351	0.4361	0.4369	0.0133	0.0118	0.0072	0.0089
0.8	1.1531	1.1723	1.1615	1.1697	0.0249	0.0184	0.0229	0.0202
Relative bias								
0.0	0.0235	0.0288	0.0033	0.0131	−0.0183	−0.0110	−0.0046	−0.0102
0.2	0.8070	0.7627	0.7544	0.7304	−0.0214	−0.0041	0.0036	0.0176
0.4	1.7822	1.9323	1.9389	1.9569	0.1020	0.0286	0.0325	0.0293
0.6	4.3984	4.3508	4.3608	4.3691	0.1327	0.1179	0.0716	0.0892
0.8	11.5310	11.7229	11.6148	11.6967	0.2489	0.1844	0.2294	0.2023

	$\hat{\beta}_2$			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**
RMSE				
0.0	0.0179	0.0075	0.0037	0.0024
0.2	0.0176	0.0077	0.0036	0.0028
0.4	0.0174	0.0075	0.0046	0.0040
0.6	0.0203	0.0108	0.0094	0.0091
0.8	0.0232	0.0228	0.0225	0.0211
Bias				
0.0	−0.0008	−0.0005	−0.0005	−0.0002
0.2	−0.0040	−0.0002	0.0005	0.0005
0.4	0.0052	0.0032	0.0033	0.0032
0.6	0.0107	0.0085	0.0089	0.0089
0.8	0.0174	0.0217	0.0224	0.0210
Relative bias				
0.0	−0.0084	−0.0048	−0.0053	−0.0022
0.2	−0.0399	−0.0023	0.0047	0.0048
0.4	0.0521	0.0316	0.0327	0.0318
0.6	0.1069	0.0854	0.0887	0.0885
0.8	0.1745	0.2175	0.2237	0.2102

Table 44: Monte Carlo results of parameter estimates for QML estimations of NB-SAR data ($\alpha = 1/2$) and non-spatial Poisson model.

A.2 Monte Carlo Marginal Effects Estimates

Here, the complete results of the marginal effects estimations from the Monte Carlo study for the P-SAR, NB-SAR and non-spatial Poisson models are reported (see also Section 3.4.3). Each table compiles the RMSE and bias for all parameters of one combination of DGP, model and estimation method. For an overview of all employed combinations see Table 1 on Page 37 in the main text.

Total Marginal Effects								
RMSE	$\boldsymbol{X_1}$				$\boldsymbol{X_2}$			
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0870	0.0388	0.0152	0.0120	0.0502	0.0211	0.0086	0.0065
0.2	0.1174	0.0512	0.0182	0.0143	0.0692	0.0262	0.0124	0.0085
0.4	0.1290	0.0638	0.0273	0.0166	0.0841	0.0336	0.0179	0.0114
0.6	0.1986	0.2802	0.0385	0.0301	0.1079	0.2724	0.0220	0.0173
0.8	0.4816	0.2448	0.1008	0.0669	0.2430	0.1053	0.0497	0.0343
Bias								
0	−0.0039	0.0032	−0.0004	0.0005	0.0110	0.0009	0.0005	−0.0010
0.2	0.0049	−0.0011	0.0007	0.0023	0.0073	−0.0025	0.0009	0.0008
0.4	0.0098	0.0033	0.0001	−0.0003	0.0140	0.0078	0.0011	0.0002
0.6	−0.0023	−0.1334	0.0021	−0.0012	0.0041	−0.1307	−0.0027	0.0003
0.8	0.0605	0.0270	0.0001	−0.0049	0.0359	0.0144	0.0017	−0.0044
Relative bias								
0	−0.0282	0.0233	−0.0027	0.0034	0.0801	0.0062	0.0040	−0.0075
0.2	0.0261	−0.0057	0.0035	0.0125	0.0393	−0.0136	0.0046	0.0042
0.4	0.0347	0.0119	0.0005	−0.0009	0.0497	0.0279	0.0040	0.0007
0.6	−0.0041	−0.2436	0.0038	−0.0023	0.0073	−0.2386	−0.0050	0.0005
0.8	0.0264	0.0115	0.0000	−0.0021	0.0157	0.0062	0.0007	−0.0019
Direct Marginal Effects								
RMSE	$\boldsymbol{X_1}$				$\boldsymbol{X_2}$			
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0635	0.0299	0.0132	0.0095	0.0186	0.0076	0.0041	0.0028
0.2	0.0779	0.0346	0.0120	0.0101	0.0180	0.0092	0.0036	0.0030
0.4	0.0657	0.0346	0.0129	0.0093	0.0248	0.0104	0.0042	0.0030
0.6	0.0732	0.1211	0.0153	0.0113	0.0240	0.1172	0.0052	0.0031
0.8	0.0929	0.0469	0.0224	0.0147	0.0405	0.0166	0.0060	0.0050
Bias								
0	−0.0143	0.0014	−0.0012	0.0011	0.0005	−0.0005	−0.0003	−0.0003
0.2	−0.0031	0.0009	−0.0001	0.0012	−0.0002	0.0001	0.0000	−0.0001
0.4	0.0033	−0.0020	−0.0005	−0.0005	0.0052	0.0007	0.0002	−0.0002
0.6	−0.0029	−0.0570	0.0018	−0.0011	−0.0008	−0.0560	−0.0003	−0.0004
0.8	−0.0024	−0.0016	−0.0003	0.0000	−0.0063	−0.0039	0.0001	0.0001
Relative bias								
0	−0.1039	0.0100	−0.0087	0.0080	0.0035	−0.0037	−0.0020	−0.0025
0.2	−0.0204	0.0058	−0.0010	0.0078	−0.0012	0.0006	−0.0003	−0.0008
0.4	0.0189	−0.0115	−0.0027	−0.0026	0.0300	0.0042	0.0010	−0.0013
0.6	−0.0120	−0.2391	0.0078	−0.0046	−0.0034	−0.2350	−0.0011	−0.0018
0.8	−0.0043	−0.0028	−0.0005	0.0000	−0.0112	−0.0068	0.0002	0.0002

Table 45: Monte Carlo results of marginal effects for FIML estimations of P-SAR data and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

Total Marginal Effects

RMSE	X_1				X_2			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
0	0.3050	0.0486	0.0176	0.0139	0.3170	0.0330	0.0122	0.0090
0.2	17.3133	0.0800	0.0374	0.0330	23.2387	0.0592	0.0349	0.0293
0.4	2.2432	0.1642	0.1055	0.0920	2.6324	0.1519	0.1027	0.0912
0.6	0.5801	0.3416	0.2713	0.2573	0.9400	0.3193	0.2591	0.2591
0.8	0.9867	0.4086	0.1768	0.1268	0.6276	0.2664	0.1144	0.0848
Bias								
0	0.0663	0.0094	0.0011	0.0007	0.0907	0.0073	0.0020	−0.0008
0.2	1.7920	0.0293	0.0253	0.0269	2.3968	0.0275	0.0258	0.0253
0.4	0.3393	0.1119	0.0895	0.0865	0.1569	0.1195	0.0914	0.0875
0.6	0.2996	0.2600	0.2555	0.2460	0.4226	0.2676	0.2496	0.2518
0.8	0.1341	0.1275	0.0598	0.0404	0.1465	0.0826	0.0427	0.0362
Relative bias								
0	0.4818	0.0683	0.0078	0.0054	0.6596	0.0527	0.0146	−0.0057
0.2	9.5906	0.1577	0.1360	0.1448	12.8275	0.1481	0.1391	0.1362
0.4	1.2061	0.3994	0.3179	0.3071	0.5579	0.4265	0.3247	0.3108
0.6	0.5333	0.4747	0.4693	0.4503	0.7523	0.4886	0.4584	0.4610
0.8	0.0586	0.0545	0.0257	0.0173	0.0640	0.0353	0.0184	0.0155

Direct Marginal Effects

RMSE	X_1				X_2			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
0	0.0737	0.0304	0.0133	0.0097	0.0366	0.0081	0.0042	0.0029
0.2	0.9858	0.0393	0.0153	0.0140	1.3228	0.0146	0.0096	0.0090
0.4	0.2642	0.0542	0.0345	0.0312	0.3602	0.0388	0.0313	0.0297
0.6	0.1506	0.1047	0.0934	0.0874	0.1472	0.0943	0.0877	0.0879
0.8	0.2895	0.2272	0.1999	0.1924	0.2115	0.1929	0.1899	0.1892
Bias								
0	−0.0061	0.0023	−0.0009	0.0011	0.0130	0.0007	0.0000	−0.0004
0.2	0.1099	0.0098	0.0079	0.0093	0.1476	0.0088	0.0082	0.0081
0.4	0.0873	0.0308	0.0292	0.0288	0.0979	0.0347	0.0303	0.0293
0.6	0.0863	0.0875	0.0894	0.0850	0.1075	0.0905	0.0870	0.0874
0.8	0.1849	0.2043	0.1950	0.1901	0.1943	0.1899	0.1895	0.1889
Relative bias								
0	−0.0442	0.0169	−0.0066	0.0078	0.0948	0.0053	0.0002	−0.0025
0.2	0.7299	0.0658	0.0526	0.0624	0.9799	0.0586	0.0549	0.0541
0.4	0.5012	0.1775	0.1672	0.1651	0.5616	0.2001	0.1736	0.1680
0.6	0.3529	0.3672	0.3774	0.3576	0.4398	0.3798	0.3670	0.3677
0.8	0.3316	0.3586	0.3449	0.3343	0.3485	0.3334	0.3351	0.3321

Table 46: Monte Carlo results of marginal effects for LIML estimations of P-SAR data and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

Total Marginal Effects

RMSE	X_1				X_2			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
0	0.0869	0.0388	0.0152	0.0120	0.0502	0.0211	0.0086	0.0065
0.2	0.1173	0.0511	0.0182	0.0143	0.0692	0.0262	0.0124	0.0085
0.4	0.1289	0.0638	0.0273	0.0166	0.0841	0.0336	0.0178	0.0114
0.6	0.1986	0.0878	0.0385	0.0301	0.1082	0.0503	0.0220	0.0173
0.8	0.4816	0.2449	0.1008	0.0670	0.2430	0.1052	0.0497	0.0342
Bias								
0	−0.0039	0.0032	−0.0004	0.0005	0.0110	0.0009	0.0005	−0.0010
0.2	0.0049	−0.0011	0.0007	0.0023	0.0073	−0.0025	0.0009	0.0008
0.4	0.0099	0.0033	0.0001	−0.0003	0.0140	0.0078	0.0011	0.0002
0.6	−0.0020	−0.0044	0.0021	−0.0012	0.0040	0.0026	−0.0027	0.0003
0.8	0.0607	0.0272	0.0001	−0.0050	0.0362	0.0145	0.0016	−0.0045
Relative bias								
0	−0.0286	0.0233	−0.0028	0.0034	0.0803	0.0062	0.0040	−0.0075
0.2	0.0262	−0.0059	0.0036	0.0125	0.0392	−0.0136	0.0046	0.0042
0.4	0.0350	0.0119	0.0005	−0.0009	0.0499	0.0279	0.0040	0.0007
0.6	−0.0036	−0.0081	0.0038	−0.0022	0.0071	0.0048	−0.0050	0.0006
0.8	0.0265	0.0116	0.0000	−0.0021	0.0158	0.0062	0.0007	−0.0019

Direct Marginal Effects

RMSE	X_1				X_2			
$\lambda \backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
0	0.0635	0.0299	0.0132	0.0095	0.0186	0.0076	0.0041	0.0028
0.2	0.0779	0.0346	0.0120	0.0101	0.0180	0.0092	0.0036	0.0030
0.4	0.0657	0.0346	0.0129	0.0093	0.0248	0.0104	0.0042	0.0030
0.6	0.0730	0.0353	0.0153	0.0113	0.0239	0.0111	0.0052	0.0031
0.8	0.0930	0.0469	0.0224	0.0147	0.0405	0.0166	0.0060	0.0050
Bias								
0	−0.0144	0.0014	−0.0012	0.0011	0.0005	−0.0005	−0.0003	−0.0003
0.2	−0.0030	0.0009	−0.0001	0.0012	−0.0002	0.0001	0.0000	−0.0001
0.4	0.0033	−0.0020	−0.0005	−0.0005	0.0052	0.0007	0.0002	−0.0002
0.6	−0.0027	−0.0016	0.0018	−0.0011	−0.0008	0.0011	−0.0003	−0.0004
0.8	−0.0024	−0.0016	−0.0003	0.0000	−0.0062	−0.0039	0.0001	0.0001
Relative bias								
0	−0.1043	0.0101	−0.0087	0.0080	0.0035	−0.0037	−0.0020	−0.0025
0.2	−0.0200	0.0057	−0.0010	0.0078	−0.0010	0.0006	−0.0003	−0.0008
0.4	0.0191	−0.0115	−0.0028	−0.0026	0.0298	0.0042	0.0010	−0.0013
0.6	−0.0112	−0.0068	0.0078	−0.0046	−0.0032	0.0047	−0.0011	−0.0018
0.8	−0.0043	−0.0028	−0.0005	0.0000	−0.0112	−0.0068	0.0002	0.0002

Table 47: Monte Carlo results of marginal effects for FIML estimations of P-SAR data and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

Total Marginal Effects								
RMSE		$\boldsymbol{X_1}$				$\boldsymbol{X_2}$		
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.3043	0.0487	0.0176	0.0139	0.3165	0.0330	0.0122	0.0090
0.2	16.5297	0.0799	0.0374	0.0330	22.2452	0.0591	0.0349	0.0293
0.4	2.0760	0.1640	0.1055	0.0920	2.9451	0.1519	0.1027	0.0912
0.6	0.5876	0.3417	0.2713	0.2574	0.9482	0.3193	0.2590	0.2592
0.8	0.9867	0.4083	0.1767	0.1268	0.6237	0.2658	0.1142	0.0847
Bias								
0	0.0660	0.0094	0.0011	0.0007	0.0907	0.0072	0.0020	−0.0008
0.2	1.6611	0.0293	0.0253	0.0269	2.2652	0.0275	0.0259	0.0253
0.4	0.2844	0.1117	0.0895	0.0864	0.0797	0.1194	0.0914	0.0875
0.6	0.3022	0.2601	0.2555	0.2461	0.4257	0.2676	0.2495	0.2519
0.8	0.1346	0.1277	0.0598	0.0404	0.1467	0.0827	0.0427	0.0362
Relative bias								
0	0.4802	0.0680	0.0078	0.0054	0.6592	0.0525	0.0146	−0.0057
0.2	8.8899	0.1574	0.1363	0.1448	12.1229	0.1479	0.1392	0.1362
0.4	1.0109	0.3988	0.3178	0.3071	0.2832	0.4261	0.3246	0.3107
0.6	0.5380	0.4748	0.4692	0.4504	0.7578	0.4886	0.4583	0.4612
0.8	0.0588	0.0545	0.0257	0.0173	0.0641	0.0353	0.0184	0.0155

Direct Marginal Effects								
RMSE		$\boldsymbol{X_1}$				$\boldsymbol{X_2}$		
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0736	0.0304	0.0133	0.0097	0.0365	0.0081	0.0042	0.0029
0.2	1.0292	0.0393	0.0153	0.0140	1.2995	0.0146	0.0096	0.0090
0.4	0.3243	0.0541	0.0345	0.0312	0.6157	0.0388	0.0313	0.0297
0.6	0.1512	0.1047	0.0934	0.0874	0.1482	0.0943	0.0877	0.0879
0.8	0.2896	0.2272	0.1999	0.1924	0.2113	0.1929	0.1899	0.1892
Bias								
0	−0.0062	0.0023	−0.0009	0.0011	0.0130	0.0007	0.0000	−0.0004
0.2	0.0677	0.0098	0.0079	0.0093	0.1189	0.0088	0.0082	0.0081
0.4	0.0968	0.0308	0.0292	0.0288	0.1237	0.0347	0.0303	0.0293
0.6	0.0867	0.0876	0.0894	0.0850	0.1079	0.0905	0.0870	0.0874
0.8	0.1849	0.2043	0.1950	0.1902	0.1943	0.1899	0.1894	0.1889
Relative bias								
0	−0.0447	0.0168	−0.0066	0.0078	0.0947	0.0053	0.0002	−0.0025
0.2	0.4495	0.0656	0.0529	0.0624	0.7894	0.0585	0.0549	0.0541
0.4	0.5556	0.1773	0.1672	0.1650	0.7098	0.2000	0.1736	0.1680
0.6	0.3545	0.3673	0.3774	0.3576	0.4416	0.3798	0.3670	0.3677
0.8	0.3318	0.3587	0.3449	0.3343	0.3486	0.3334	0.3351	0.3321

Table 48: Monte Carlo results of marginal effects for LIML estimations of P-SAR data and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

Total Marginal Effects

RMSE	X_1				X_2			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0633	0.0296	0.0132	0.0095	0.0185	0.0076	0.0041	0.0028
0.2	0.0863	0.0494	0.0384	0.0362	0.0404	0.0374	0.0363	0.0364
0.4	0.1228	0.1119	0.1090	0.1099	0.1051	0.1055	0.1072	0.1073
0.6	0.3355	0.3078	0.3048	0.3097	0.3198	0.3086	0.3079	0.3092
0.8	1.7262	1.7810	1.7561	1.7709	1.7530	1.7752	1.7586	1.7722
Bias								
0	−0.0147	0.0011	−0.0012	0.0011	0.0005	−0.0005	−0.0003	−0.0003
0.2	−0.0400	−0.0352	−0.0365	−0.0348	−0.0361	−0.0363	−0.0362	−0.0362
0.4	−0.1020	−0.1065	−0.1082	−0.1094	−0.1022	−0.1050	−0.1071	−0.1073
0.6	−0.3266	−0.3058	−0.3043	−0.3095	−0.3189	−0.3084	−0.3079	−0.3092
0.8	−1.7221	−1.7801	−1.7559	−1.7708	−1.7526	−1.7751	−1.7586	−1.7722
Relative bias								
0	−0.1072	0.0083	−0.0089	0.0079	0.0036	−0.0034	−0.0020	−0.0024
0.2	−0.2142	−0.1896	−0.1962	−0.1875	−0.1933	−0.1953	−0.1946	−0.1952
0.4	−0.3626	−0.3801	−0.3843	−0.3888	−0.3633	−0.3748	−0.3804	−0.3812
0.6	−0.5814	−0.5583	−0.5589	−0.5666	−0.5677	−0.5630	−0.5655	−0.5661
0.8	−0.7521	−0.7604	−0.7552	−0.7570	−0.7654	−0.7583	−0.7563	−0.7576

Direct Marginal Effects

RMSE	X_1				X_2			
$\lambda \backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0633	0.0296	0.0132	0.0095	0.0185	0.0076	0.0041	0.0028
0.2	0.0766	0.0347	0.0121	0.0100	0.0182	0.0092	0.0036	0.0030
0.4	0.0685	0.0345	0.0136	0.0099	0.0249	0.0104	0.0042	0.0030
0.6	0.0773	0.0352	0.0168	0.0119	0.0242	0.0111	0.0053	0.0031
0.8	0.1200	0.0582	0.0269	0.0161	0.0444	0.0172	0.0064	0.0054
Bias								
0	−0.0147	0.0011	−0.0012	0.0011	0.0005	−0.0005	−0.0003	−0.0003
0.2	−0.0037	0.0009	−0.0004	0.0013	0.0002	−0.0002	−0.0001	−0.0002
0.4	0.0050	0.0001	−0.0011	−0.0023	0.0048	0.0016	0.0000	−0.0002
0.6	−0.0093	0.0036	0.0031	−0.0010	−0.0016	0.0010	−0.0005	−0.0007
0.8	0.0103	−0.0088	0.0039	−0.0005	−0.0202	−0.0038	0.0012	−0.0019
Relative bias								
0	−0.1072	0.0083	−0.0089	0.0079	0.0036	−0.0034	−0.0020	−0.0024
0.2	−0.0249	0.0058	−0.0026	0.0084	0.0010	−0.0013	−0.0006	−0.0012
0.4	0.0288	0.0008	−0.0060	−0.0133	0.0278	0.0094	0.0002	−0.0010
0.6	−0.0379	0.0150	0.0131	−0.0041	−0.0066	0.0042	−0.0019	−0.0028
0.8	0.0185	−0.0155	0.0069	−0.0009	−0.0362	−0.0067	0.0022	−0.0033

Table 49: Monte Carlo results of marginal effects for QML estimations of P-SAR data and non-spatial Poisson model. Values reported as 0.0000 are smaller than 0.00005.

	Total Marginal Effects							
RMSE	X_1				X_2			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0998	1.4383	0.0156	0.0116	0.0570	1.4091	0.0108	0.0074
0.2	0.1265	0.0437	0.0195	0.0153	0.0737	0.0265	0.0137	0.0090
0.4	0.1442	0.0642	0.0304	0.0199	0.0880	0.0391	0.0171	0.0109
0.6	0.2411	0.1012	0.0475	0.0347	0.1299	0.0522	0.0249	0.0193
0.8	0.5233	0.2764	0.1088	0.0765	0.3269	0.1303	0.0692	0.0462
Bias								
0	0.0133	0.8463	−0.0017	0.0021	0.0040	0.8476	0.0007	0.0009
0.2	0.0104	0.0020	0.0019	−0.0012	0.0141	−0.0006	0.0008	−0.0015
0.4	0.0250	−0.0107	0.0046	−0.0003	0.0146	−0.0020	0.0000	−0.0004
0.6	0.0023	0.0011	0.0048	0.0010	0.0171	0.0025	−0.0008	0.0018
0.8	−0.0018	−0.0563	0.0041	−0.0074	0.0408	−0.0044	0.0080	−0.0022
Relative bias								
0	0.0964	6.1458	−0.0121	0.0155	0.0290	6.1554	0.0053	0.0064
0.2	0.0555	0.0105	0.0103	−0.0064	0.0756	−0.0030	0.0046	−0.0082
0.4	0.0889	−0.0380	0.0164	−0.0009	0.0520	−0.0073	−0.0001	−0.0014
0.6	0.0042	0.0019	0.0088	0.0019	0.0304	0.0046	−0.0015	0.0033
0.8	−0.0008	−0.0241	0.0018	−0.0031	0.0178	−0.0019	0.0034	−0.0010

	Direct Marginal Effects							
RMSE	X_1				X_2			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0697	0.2687	0.0134	0.0101	0.0225	0.2602	0.0044	0.0028
0.2	0.0750	0.0309	0.0129	0.0105	0.0214	0.0099	0.0051	0.0029
0.4	0.0704	0.0313	0.0149	0.0104	0.0210	0.0098	0.0048	0.0033
0.6	0.0956	0.0369	0.0178	0.0138	0.0268	0.0131	0.0055	0.0038
0.8	0.1188	0.0590	0.0249	0.0149	0.0407	0.0202	0.0086	0.0065
Bias								
0	0.0064	0.1557	−0.0030	0.0012	0.0011	0.1584	−0.0007	−0.0001
0.2	−0.0053	0.0039	0.0009	0.0000	−0.0008	0.0017	0.0000	−0.0003
0.4	0.0095	−0.0056	0.0022	−0.0001	0.0010	0.0001	−0.0005	−0.0001
0.6	−0.0036	−0.0014	0.0019	−0.0001	0.0013	−0.0004	−0.0005	0.0002
0.8	−0.0070	−0.0123	−0.0001	−0.0007	0.0020	0.0008	0.0007	0.0006
Relative bias								
0	0.0464	1.1308	−0.0215	0.0087	0.0077	1.1505	−0.0049	−0.0008
0.2	−0.0355	0.0261	0.0062	0.0000	−0.0050	0.0113	−0.0001	−0.0018
0.4	0.0546	−0.0325	0.0129	−0.0004	0.0057	0.0006	−0.0030	−0.0006
0.6	−0.0146	−0.0058	0.0081	−0.0002	0.0053	−0.0016	−0.0021	0.0010
0.8	−0.0125	−0.0216	−0.0001	−0.0013	0.0036	0.0014	0.0013	0.0010

Table 50: Monte Carlo results of marginal effects for FIML estimations of NB-SAR data ($\alpha = 1/8$) and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

Total Marginal Effects								
RMSE		$\boldsymbol{X_1}$				$\boldsymbol{X_2}$		
$\boldsymbol{\lambda\backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.2359	0.0428	0.0190	0.0143	0.2071	0.0383	0.0161	0.0114
0.2	7.54E+29	0.0800	0.0446	0.0354	1.18E+30	0.0727	0.0403	0.0315
0.4	1.23E+79	0.2446	0.1473	0.1227	1.16E+79	0.2273	0.1345	0.1212
0.6	528.2818	0.7548	0.4668	0.4390	1.86E+03	0.7535	0.4486	0.4368
0.8	25.5919	1.2148	0.9936	0.8767	22.3771	1.0870	0.9651	0.8709
Bias								
0	0.0909	0.0106	0.0004	0.0034	0.0815	0.0114	0.0029	0.0022
0.2	7.54E+28	0.0395	0.0331	0.0266	1.18E+29	0.0380	0.0315	0.0260
0.4	1.23E+78	0.1332	0.1293	0.1153	1.16E+78	0.1498	0.1222	0.1165
0.6	61.6330	0.5249	0.4383	0.4255	197.1658	0.5413	0.4252	0.4260
0.8	4.5189	0.9066	0.9403	0.8470	4.0827	0.9017	0.9277	0.8528
Relative bias								
0	0.6610	0.0769	0.0031	0.0247	0.5928	0.0826	0.0213	0.0156
0.2	4.04E+29	0.2124	0.1782	0.1434	6.30E+29	0.2045	0.1693	0.1400
0.4	4.38E+78	0.4754	0.4591	0.4098	4.14E+78	0.5347	0.4342	0.4138
0.6	109.7132	0.9583	0.8050	0.7789	350.9760	0.9883	0.7810	0.7799
0.8	1.9735	0.3873	0.4044	0.3621	1.7829	0.3852	0.3990	0.3646

Direct Marginal Effects								
RMSE		$\boldsymbol{X_1}$				$\boldsymbol{X_2}$		
$\boldsymbol{\lambda\backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0827	0.0290	0.0132	0.0102	0.0367	0.0114	0.0050	0.0032
0.2	3.66E+30	0.0362	0.0186	0.0151	5.71E+30	0.0189	0.0122	0.0097
0.4	4.11E+76	0.0601	0.0476	0.0401	3.88E+76	0.0512	0.0398	0.0383
0.6	25.0430	0.1718	0.1397	0.1337	87.3700	0.1662	0.1318	0.1318
0.8	2.3022	0.4307	0.4102	0.3919	2.0051	0.4053	0.4022	0.3918
Bias								
0	0.0215	0.0027	−0.0025	0.0015	0.0136	0.0025	−0.0002	0.0002
0.2	3.657E+29	0.0148	0.0114	0.0095	5.706E+29	0.0129	0.0100	0.0090
0.4	4.106E+75	0.0334	0.0420	0.0372	3.877E+75	0.0428	0.0385	0.0378
0.6	3.2036	0.1403	0.1347	0.1308	9.6043	0.1462	0.1296	0.1308
0.8	0.7478	0.3958	0.4043	0.3890	0.6945	0.3953	0.4004	0.3909
Relative bias								
0	0.156	0.0195	−0.0179	0.0112	0.099	0.0181	−0.0011	0.0015
0.2	2.43E+30	0.0991	0.0760	0.0632	3.79E+30	0.0860	0.0670	0.0600
0.4	2.36E+76	0.1925	0.2409	0.2132	2.22E+76	0.2468	0.2206	0.2167
0.6	13.1057	0.5886	0.5684	0.5502	39.2900	0.6135	0.5469	0.5503
0.8	1.3415	0.6948	0.7153	0.6839	1.2459	0.6939	0.7082	0.6872

Table 51: Monte Carlo results of marginal effects for LIML estimations of NB-SAR data ($\alpha = 1/8$) and P-SAR model.

Total Marginal Effects

RMSE	X_1				X_2			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0996	0.0330	0.0156	0.0117	0.0569	0.0229	0.0108	0.0074
0.2	0.1260	0.0435	0.0194	0.0153	0.0739	0.0265	0.0137	0.0090
0.4	0.1432	0.0641	0.0304	0.0198	0.0874	0.0391	0.0170	0.0108
0.6	0.2384	0.1006	0.0479	0.0343	0.1303	0.0522	0.0248	0.0192
0.8	0.5148	0.2733	0.1089	0.0772	0.3344	0.1286	0.0683	0.0453
Bias								
0	0.0134	0.0020	−0.0017	0.0022	0.0042	0.0029	0.0007	0.0009
0.2	0.0107	0.0022	0.0019	−0.0013	0.0145	−0.0005	0.0009	−0.0015
0.4	0.0242	−0.0112	0.0046	−0.0002	0.0137	−0.0020	0.0000	−0.0004
0.6	0.0029	0.0005	0.0049	0.0009	0.0180	0.0021	−0.0008	0.0016
0.8	−0.0018	−0.0529	0.0027	−0.0070	0.0456	−0.0003	0.0083	−0.0016
Relative bias								
0	0.0977	0.0147	−0.0123	0.0158	0.0307	0.0212	0.0054	0.0064
0.2	0.0573	0.0119	0.0100	−0.0069	0.0775	−0.0026	0.0048	−0.0082
0.4	0.0859	−0.0399	0.0164	−0.0008	0.0486	−0.0073	0.0000	−0.0016
0.6	0.0051	0.0010	0.0090	0.0016	0.0320	0.0037	−0.0015	0.0030
0.8	−0.0008	−0.0226	0.0011	−0.0030	0.0199	−0.0001	0.0036	−0.0007

Direct Marginal Effects

RMSE	X_1				X_2			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0697	0.0297	0.0134	0.0101	0.0223	0.0103	0.0044	0.0028
0.2	0.0745	0.0306	0.0128	0.0104	0.0214	0.0098	0.0051	0.0029
0.4	0.0705	0.0312	0.0149	0.0103	0.0210	0.0096	0.0047	0.0032
0.6	0.0949	0.0365	0.0178	0.0137	0.0267	0.0131	0.0054	0.0038
0.8	0.1166	0.0584	0.0251	0.0151	0.0398	0.0202	0.0088	0.0064
Bias								
0	0.0063	0.0005	−0.0030	0.0012	0.0011	0.0007	−0.0007	−0.0001
0.2	−0.0053	0.0040	0.0009	−0.0001	−0.0007	0.0016	0.0000	−0.0003
0.4	0.0095	−0.0059	0.0022	0.0000	0.0008	0.0002	−0.0005	−0.0001
0.6	−0.0036	−0.0016	0.0019	−0.0001	0.0013	−0.0005	−0.0005	0.0002
0.8	−0.0063	−0.0122	−0.0005	−0.0008	0.0034	0.0011	0.0008	0.0005
Relative bias								
0	0.0461	0.0034	−0.0217	0.0090	0.0077	0.0054	−0.0049	−0.0008
0.2	−0.0351	0.0267	0.0057	−0.0004	−0.0044	0.0109	0.0000	−0.0017
0.4	0.0548	−0.0339	0.0129	−0.0002	0.0044	0.0011	−0.0029	−0.0007
0.6	−0.0147	−0.0068	0.0082	−0.0003	0.0053	−0.0023	−0.0021	0.0009
0.8	−0.0113	−0.0214	−0.0009	−0.0015	0.0062	0.0019	0.0013	0.0009

Table 52: Monte Carlo results of marginal effects for FIML estimations of NB-SAR data ($\alpha = 1/8$) and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

Total Marginal Effects								
RMSE	$\boldsymbol{X_1}$				$\boldsymbol{X_2}$			
$\boldsymbol{\lambda\backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.2365	0.0429	0.0189	0.0143	0.2146	0.0384	0.0161	0.0114
0.2	8.0769	0.0802	0.0446	0.0353	10.9556	0.0729	0.0403	0.0315
0.4	7.42E+03	0.2448	0.1472	0.1225	6.94E+03	0.2282	0.1346	0.1209
0.6	4.00E+04	0.7511	0.4672	0.4374	1.41E+05	0.7530	0.4484	0.4353
0.8	31.7100	1.1757	0.9775	0.8768	27.2540	1.0487	0.9523	0.8693
Bias								
0	0.0920	0.0103	0.0004	0.0035	0.0837	0.0113	0.0029	0.0021
0.2	1.2972	0.0400	0.0331	0.0265	1.5735	0.0383	0.0315	0.0260
0.4	743.0076	0.1323	0.1293	0.1153	695.0297	0.1500	0.1223	0.1163
0.6	4.00E+03	0.5204	0.4387	0.4243	1.41E+04	0.5370	0.4253	0.4250
0.8	5.1326	0.8987	0.9283	0.8478	4.7729	0.8951	0.9209	0.8525
Relative bias								
0	0.6687	0.0750	0.0028	0.0251	0.6086	0.0818	0.0214	0.0156
0.2	6.9423	0.2151	0.1782	0.1428	8.4213	0.2058	0.1697	0.1398
0.4	2.64E+03	0.4724	0.4591	0.4095	2.47E+03	0.5353	0.4344	0.4130
0.6	7.13E+03	0.9500	0.8058	0.7767	2.51E+04	0.9804	0.7811	0.7779
0.8	2.2414	0.3839	0.3992	0.3624	2.0844	0.3824	0.3961	0.3644
Direct Marginal Effects								
RMSE	$\boldsymbol{X_1}$				$\boldsymbol{X_2}$			
$\boldsymbol{\lambda\backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0826	0.0291	0.0132	0.0102	0.0374	0.0114	0.0049	0.0032
0.2	0.5649	0.0360	0.0185	0.0150	0.7417	0.0188	0.0122	0.0097
0.4	265.10	0.0598	0.0476	0.0401	247.95	0.0513	0.0398	0.0383
0.6	1363.14	0.1706	0.1398	0.1334	4802.51	0.1655	0.1318	0.1316
0.8	2.7147	0.4269	0.4074	0.3921	2.3502	0.4026	0.4008	0.3916
Bias								
0	0.0217	0.0025	−0.0025	0.0016	0.0139	0.0139	−0.0001	0.0002
0.2	0.1183	0.0150	0.0113	0.0094	0.1403	0.1403	0.0101	0.0090
0.4	26.6734	0.0330	0.0420	0.0372	24.9681	24.9681	0.0385	0.0377
0.6	136.6723	0.1395	0.1349	0.1306	480.6801	480.6801	0.1296	0.1306
0.8	0.7876	0.3945	0.4016	0.3892	0.7487	0.7487	0.3993	0.3907
Relative bias								
0	0.1575	0.0180	−0.0181	0.0116	0.1012	0.0178	−0.0010	0.0015
0.2	0.7856	0.1002	0.0756	0.0628	0.9318	0.0859	0.0671	0.0601
0.4	153.0662	0.1902	0.2409	0.2134	143.2804	0.2473	0.2208	0.2164
0.6	559.1086	0.5852	0.5690	0.5492	1966.3993	0.6096	0.5470	0.5494
0.8	1.4129	0.6925	0.7105	0.6842	1.3430	0.6925	0.7063	0.6869

Table 53: Monte Carlo results of marginal effects for LIML estimations of NB-SAR data ($\alpha = 1/8$) and NB-SAR model.

Total Marginal Effects								
RMSE		$\boldsymbol{X_1}$				$\boldsymbol{X_2}$		
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0692	0.0293	0.0133	0.0102	0.0226	0.0103	0.0044	0.0028
0.2	0.0847	0.0446	0.0376	0.0374	0.0424	0.0360	0.0365	0.0365
0.4	0.1242	0.1154	0.1063	0.1098	0.1085	0.1062	0.1079	0.1072
0.6	0.3459	0.3100	0.3040	0.3080	0.3178	0.3099	0.3081	0.3086
0.8	1.7136	1.7891	1.7573	1.7727	1.7471	1.7710	1.7581	1.7714
Bias								
0	0.0074	0.0009	−0.0030	0.0012	0.0008	0.0008	−0.0007	−0.0001
0.2	−0.0434	−0.0324	−0.0352	−0.0359	−0.0366	−0.0347	−0.0361	−0.0364
0.4	−0.0998	−0.1111	−0.1053	−0.1093	−0.1063	−0.1057	−0.1078	−0.1072
0.6	−0.3306	−0.3079	−0.3035	−0.3077	−0.3167	−0.3097	−0.3080	−0.3085
0.8	−1.7077	−1.7879	−1.7571	−1.7726	−1.7466	−1.7709	−1.7581	−1.7714
Relative bias								
0	0.0539	0.0065	−0.0216	0.0088	0.0060	0.0056	−0.0050	−0.0008
0.2	−0.2323	−0.1745	−0.1894	−0.1932	−0.1958	−0.1866	−0.1945	−0.1960
0.4	−0.3548	−0.3965	−0.3739	−0.3882	−0.3779	−0.3773	−0.3830	−0.3807
0.6	−0.5886	−0.5621	−0.5575	−0.5633	−0.5637	−0.5654	−0.5657	−0.5648
0.8	−0.7458	−0.7638	−0.7557	−0.7578	−0.7627	−0.7565	−0.7561	−0.7573
Direct Marginal Effects								
RMSE		$\boldsymbol{X_1}$				$\boldsymbol{X_2}$		
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0692	0.0293	0.0133	0.0102	0.0226	0.0103	0.0044	0.0028
0.2	0.0731	0.0308	0.0132	0.0106	0.0215	0.0099	0.0051	0.0029
0.4	0.0743	0.0316	0.0153	0.0111	0.0216	0.0099	0.0048	0.0032
0.6	0.1026	0.0361	0.0182	0.0149	0.0268	0.0132	0.0055	0.0037
0.8	0.1443	0.0681	0.0278	0.0177	0.0448	0.0203	0.0090	0.0068
Bias								
0	0.0074	0.0009	−0.0030	0.0012	0.0008	0.0008	−0.0007	−0.0001
0.2	−0.0071	0.0037	0.0009	0.0002	−0.0003	0.0014	−0.0001	−0.0003
0.4	0.0072	−0.0045	0.0019	−0.0021	0.0007	0.0009	−0.0007	0.0000
0.6	−0.0133	0.0015	0.0039	0.0009	0.0007	−0.0003	−0.0006	0.0000
0.8	0.0247	−0.0166	0.0028	−0.0023	−0.0142	0.0004	0.0017	−0.0011
Relative bias								
0	−0.0249	0.0058	−0.0026	0.0084	0.0010	−0.0013	−0.0006	−0.0012
0.2	0.0288	0.0008	−0.0060	−0.0133	0.0278	0.0094	0.0002	−0.0010
0.4	−0.0379	0.0150	0.0131	−0.0041	−0.0066	0.0042	−0.0019	−0.0028
0.6	0.0185	−0.0155	0.0069	−0.0009	−0.0362	−0.0067	0.0022	−0.0033
0.8	−0.1072	0.0083	−0.0089	0.0079	0.0036	−0.0034	−0.0020	−0.0024

Table 54: Monte Carlo results of marginal effects for QML estimations of NB-SAR data ($\alpha = 1/8$) and non-spatial Poisson model. Values reported as 0.0000 are smaller than 0.00005.

Total Marginal Effects								
RMSE	$\boldsymbol{X_1}$				$\boldsymbol{X_2}$			
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.1195	0.0452	0.0175	0.0125	0.0777	0.0270	0.0108	0.0090
0.2	0.1482	0.0697	0.0253	0.0182	0.0763	0.0371	0.0144	0.0113
0.4	0.1752	0.0752	0.0390	0.0268	0.1114	0.0508	0.0237	0.0136
0.6	0.2917	0.1369	0.0637	0.0412	0.1498	0.0854	0.0355	0.0241
0.8	0.8377	0.4265	0.1637	0.1302	0.4772	0.1971	0.0931	0.0693
Bias								
0	0.0168	0.0042	0.0003	−0.0009	0.0202	0.0045	0.0002	0.0003
0.2	0.0030	−0.0042	−0.0017	0.0022	−0.0012	−0.0060	−0.0018	0.0003
0.4	0.0107	0.0013	0.0032	0.0019	0.0014	0.0047	0.0022	0.0003
0.6	0.0577	0.0077	0.0003	0.0009	0.0233	0.0003	0.0066	0.0052
0.8	−0.0968	−0.0102	0.0043	−0.0194	0.0017	0.0063	−0.0031	−0.0176
Relative bias								
0	0.1221	0.0308	0.0018	−0.0062	0.1465	0.0324	0.0013	0.0021
0.2	0.0161	−0.0228	−0.0091	0.0121	−0.0066	−0.0325	−0.0097	0.0018
0.4	0.0381	0.0045	0.0114	0.0067	0.0049	0.0166	0.0078	0.0010
0.6	0.1028	0.0141	0.0005	0.0016	0.0414	0.0006	0.0122	0.0095
0.8	−0.0423	−0.0044	0.0018	−0.0083	0.0008	0.0027	−0.0013	−0.0075

Direct Marginal Effects								
RMSE	$\boldsymbol{X_1}$				$\boldsymbol{X_2}$			
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0911	0.0341	0.0142	0.0102	0.0260	0.0107	0.0054	0.0035
0.2	0.0960	0.0426	0.0174	0.0135	0.0277	0.0121	0.0054	0.0042
0.4	0.0939	0.0376	0.0186	0.0136	0.0290	0.0118	0.0055	0.0042
0.6	0.1071	0.0469	0.0201	0.0159	0.0411	0.0149	0.0075	0.0048
0.8	0.2079	0.0892	0.0382	0.0264	0.0712	0.0322	0.0132	0.0094
Bias								
0	−0.0025	−0.0015	−0.0008	−0.0014	−0.0015	−0.0005	−0.0008	−0.0003
0.2	−0.0029	−0.0015	−0.0001	0.0014	−0.0073	−0.0011	−0.0001	−0.0002
0.4	0.0085	−0.0027	0.0006	0.0009	0.0025	−0.0010	0.0001	0.0000
0.6	0.0196	0.0015	−0.0033	−0.0014	0.0040	−0.0010	−0.0002	0.0005
0.8	−0.0177	−0.0057	0.0045	−0.0016	−0.0012	−0.0006	0.0026	−0.0011
Relative bias								
0	−0.0181	−0.0107	−0.0057	−0.0102	−0.0106	−0.0037	−0.0060	−0.0022
0.2	−0.0193	−0.0101	−0.0006	0.0096	−0.0484	−0.0072	−0.0009	−0.0016
0.4	0.0489	−0.0154	0.0032	0.0052	0.0143	−0.0057	0.0004	0.0001
0.6	0.0802	0.0064	−0.0139	−0.0058	0.0162	−0.0041	−0.0008	0.0020
0.8	−0.0317	−0.0101	0.0080	−0.0029	−0.0021	−0.0010	0.0045	−0.0019

Table 55: Monte Carlo results of marginal effects for FIML estimations of NB-SAR data ($\alpha = 1/2$) and P-SAR model. Values reported as 0.0000 are smaller than 0.00005.

			Total Marginal Effects					
RMSE		$\boldsymbol{X_1}$				$\boldsymbol{X_2}$		
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	1.12E+66	0.0760	0.0249	0.0169	1.23E+66	0.0601	0.0197	0.0154
0.2	Inf	0.2139	0.0696	0.0619	Inf	0.1527	0.0639	0.0577
0.4	1.95E+152	0.7295	0.3153	0.2620	1.98E+152	0.7530	0.3047	0.2494
0.6	1.06E+53	446.5421	1.9878	1.7635	1.66E+53	532.2618	1.9755	1.7797
0.8	8.77E+14	44.7439	11.9960	10.9190	1.08E+15	53.4690	11.9124	11.1559
Bias								
0	1.12E+65	0.0217	0.0045	0.0008	1.23E+65	0.0210	0.0042	0.0020
0.2	Inf	0.0762	0.0464	0.0515	Inf	0.0661	0.0464	0.0491
0.4	1.95E+151	0.4564	0.2688	0.2408	1.98E+151	0.4739	0.2681	0.2343
0.6	1.06E+52	57.6102	1.6736	1.6475	1.66E+52	65.5708	1.7101	1.6558
0.8	8.80E+13	19.9051	10.7386	10.3423	1.09E+14	21.6886	10.7113	10.5680
Relative bias								
0	8.13E+65	0.1574	0.0323	0.0057	8.94E+65	0.1522	0.0302	0.0147
0.2	Inf	0.4101	0.2495	0.2776	Inf	0.3558	0.2495	0.2647
0.4	6.94E+151	1.6292	0.9549	0.8556	7.02E+151	1.6916	0.9524	0.8324
0.6	1.88E+52	105.1779	3.0739	3.0159	2.96E+52	119.7115	3.1409	3.0312
0.8	3.84E+13	8.5031	4.6185	4.4215	4.75E+13	9.2650	4.6068	4.5180

			Direct Marginal Effects					
RMSE		$\boldsymbol{X_1}$				$\boldsymbol{X_2}$		
$\boldsymbol{\lambda \backslash n}$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	1.49E+66	0.0375	0.0148	0.0103	1.64E+66	0.0144	0.0058	0.0045
0.2	-Inf	0.0681	0.0265	0.0233	-Inf	0.0337	0.0184	0.0173
0.4	4.30E+152	0.1534	0.0886	0.0783	4.35E+152	0.1501	0.0824	0.0723
0.6	7.30E+50	29.2160	0.4308	0.4106	1.15E+51	34.6744	0.4329	0.4119
0.8	1.97562E+13	5.8314	2.4899	2.3525	2.44E+13	6.7321	2.4715	2.3993
Bias								
0	1.49E+65	0.0030	0.0003	−0.0010	1.64E+65	0.0037	0.0001	0.0001
0.2	-Inf	0.0197	0.0151	0.0176	-Inf	0.0183	0.0152	0.0157
0.4	-4.30E+151	0.1043	0.0770	0.0730	-4.35E+151	0.1093	0.0770	0.0703
0.6	7.30E+49	4.3426	0.3876	0.3949	1.15E+50	4.8244	0.4009	0.3962
0.8	1.98E+12	3.3874	2.3497	2.2845	2.45E+12	3.6008	2.3433	2.3334
Relative bias								
0	1.08E+66	0.0218	0.0022	−0.0073	1.19E+66	0.0266	0.0007	0.0010
0.2	-Inf	0.1317	0.1012	0.1175	-Inf	0.1222	0.1018	0.1047
0.4	-2.47E+152	0.6011	0.4417	0.4189	-2.49E+152	0.6297	0.4417	0.4032
0.6	2.99E+50	18.2173	1.6352	1.6610	4.69E+50	20.2385	1.6913	1.6668
0.8	3.56E+12	5.9464	4.1566	4.0162	4.40E+12	6.3211	4.1453	4.1021

Table 56: Monte Carlo results of marginal effects for LIML estimations of NB-SAR data ($\alpha = 1/2$) and P-SAR model.

Total Marginal Effects

RMSE	X_1				X_2			
$\lambda\backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
0	0.1211	0.0456	0.0177	0.0125	0.0776	0.0269	0.0107	0.0090
0.2	0.1496	0.0698	0.0255	0.0179	0.0770	0.0373	0.0143	0.0113
0.4	0.1743	0.0732	0.0390	0.0271	0.1152	0.0506	0.0236	0.0137
0.6	0.2958	0.1351	0.0639	0.0406	0.1527	0.0836	0.0355	0.0246
0.8	0.8173	0.4204	0.1616	0.1264	0.4847	0.2034	0.0918	0.0669
Bias								
0	0.0190	0.0045	0.0004	−0.0008	0.0199	0.0045	0.0003	0.0003
0.2	−0.0005	−0.0037	−0.0014	0.0020	−0.0015	−0.0058	−0.0019	0.0003
0.4	0.0093	0.0026	0.0035	0.0021	0.0012	0.0055	0.0021	0.0003
0.6	0.0557	0.0079	−0.0002	0.0015	0.0243	−0.0001	0.0067	0.0054
0.8	−0.0430	−0.0050	0.0024	−0.0173	0.0127	0.0207	−0.0019	−0.0160
Relative bias								
0	0.1380	0.0330	0.0032	−0.0061	0.1449	0.0324	0.0021	0.0020
0.2	−0.0026	−0.0198	−0.0077	0.0106	−0.0081	−0.0313	−0.0104	0.0016
0.4	0.0331	0.0094	0.0123	0.0074	0.0042	0.0195	0.0076	0.0012
0.6	0.0991	0.0144	−0.0003	0.0027	0.0432	−0.0002	0.0123	0.0098
0.8	−0.0188	−0.0021	0.0010	−0.0074	0.0055	0.0088	−0.0008	−0.0068

Direct Marginal Effects

RMSE	X_1				X_2			
$\lambda\backslash n$	1000	5000	25000	50000	1000	5000	25000	50000
0	0.0895	0.0343	0.0143	0.0102	0.0260	0.0107	0.0053	0.0035
0.2	0.0979	0.0426	0.0176	0.0132	0.0275	0.0122	0.0053	0.0041
0.4	0.0946	0.0368	0.0187	0.0137	0.0286	0.0118	0.0055	0.0042
0.6	0.1070	0.0469	0.0201	0.0157	0.0396	0.0148	0.0074	0.0048
0.8	0.2004	0.0862	0.0373	0.0259	0.0686	0.0316	0.0123	0.0090
Bias								
0	−0.0016	−0.0014	−0.0007	−0.0013	−0.0019	−0.0007	−0.0008	−0.0002
0.2	−0.0047	−0.0013	0.0002	0.0012	−0.0068	−0.0010	−0.0002	−0.0002
0.4	0.0081	−0.0020	0.0007	0.0010	0.0017	−0.0008	0.0000	0.0001
0.6	0.0178	0.0019	−0.0036	−0.0010	0.0038	−0.0010	−0.0003	0.0006
0.8	−0.0067	−0.0078	0.0036	−0.0011	−0.0008	−0.0005	0.0024	−0.0007
Relative bias								
0	−0.0115	−0.0103	−0.0052	−0.0094	−0.0137	−0.0051	−0.0060	−0.0016
0.2	−0.0315	−0.0086	0.0013	0.0083	−0.0454	−0.0066	−0.0011	−0.0015
0.4	0.0462	−0.0114	0.0038	0.0059	0.0097	−0.0049	−0.0002	0.0003
0.6	0.0728	0.0080	−0.0152	−0.0044	0.0154	−0.0041	−0.0012	0.0025
0.8	−0.0120	−0.0137	0.0063	−0.0020	−0.0014	−0.0009	0.0042	−0.0013

Table 57: Monte Carlo results of marginal effects for FIML estimations of NB-SAR data ($\alpha = 1/2$) and NB-SAR model. Values reported as 0.0000 are smaller than 0.00005.

Total Marginal Effects

RMSE		X_1				X_2		
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	1.06E+16	0.0778	0.0251	0.0170	8.96E+15	0.0602	0.0196	0.0154
0.2	Inf	0.1967	0.0687	0.0616	Inf	0.1381	0.0620	0.0577
0.4	4.57E+26	0.7879	0.3152	0.2627	7.26E+26	0.7906	0.3044	0.2496
0.6	5.46E+63	82.5983	2.0728	1.7762	3.02E+64	94.8308	2.0387	1.7886
0.8	3.37E+11	30.5315	11.3337	10.7979	3.81E+11	27.5187	11.2170	11.0595
Bias								
0	-1.06E+15	0.0226	0.0047	0.0007	-8.96E+14	0.0213	0.0044	0.0019
0.2	-2.84E+181	0.0730	0.0464	0.0512	-2.78E+181	0.0630	0.0458	0.0491
0.4	-4.57E+25	0.4720	0.2694	0.2413	-7.26E+25	0.4872	0.2681	0.2345
0.6	-5.46E+62	15.2580	1.7019	1.6578	3.02E+63	16.1389	1.7371	1.6635
0.8	4.77E+10	16.1096	10.4972	10.3492	5.13E+10	16.1813	10.4732	10.5835
Relative bias								
0	-7.67E+15	0.1639	0.0343	0.0054	-6.52E+15	0.1544	0.0316	0.0141
0.2	-1.52E+182	0.3926	0.2495	0.2757	-1.49E+182	0.3387	0.2465	0.2645
0.4	-1.62E+26	1.6848	0.9568	0.8573	-2.58E+26	1.7390	0.9522	0.8330
0.6	-9.71E+62	27.8563	3.1258	3.0349	5.37E+63	29.4644	3.1905	3.0453
0.8	2.08E+10	6.8818	4.5147	4.4244	2.24E+10	6.9124	4.5044	4.5246

Direct Marginal Effects

RMSE		X_1				X_2		
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	2.27E+16	0.0380	0.0150	0.0103	1.92E+16	0.0142	0.0058	0.0045
0.2	Inf	0.0661	0.0266	0.0229	Inf	0.0317	0.0180	0.0173
0.4	1.60E+28	0.1595	0.0887	0.0785	2.55E+28	0.1544	0.0823	0.0723
0.6	3.58E+64	6.7833	0.4405	0.4129	1.98E+65	7.6738	0.4403	0.4135
0.8	9.38E+09	4.5338	2.4166	2.3411	1.07E+10	4.2062	2.3944	2.3921
Bias								
0	2.27E+15	0.0032	0.0004	−0.0009	1.92E+15	0.0035	0.0001	0.0002
0.2	-3.62E+181	0.0193	0.0154	0.0173	-3.55E+181	0.0178	0.0151	0.0157
0.4	1.60E+27	0.1073	0.0772	0.0732	2.55E+27	0.1116	0.0769	0.0704
0.6	-3.58E+63	1.6205	0.3912	0.3971	1.98E+64	1.6677	0.4048	0.3977
0.8	1.33E+09	2.9783	2.3145	2.2885	1.43E+09	3.0194	2.3106	2.3384
Relative bias								
0	1.65E+16	0.0232	0.0029	−0.0067	1.40E+16	0.0257	0.0009	0.0015
0.2	-2.41E+182	0.1290	0.1027	0.1159	-2.36E+182	0.1190	0.1008	0.1048
0.4	9.19E+27	0.6185	0.4429	0.4199	1.46E+28	0.6430	0.4411	0.4036
0.6	-1.46E+64	6.7980	1.6506	1.6703	8.10E+64	6.9961	1.7079	1.6730
0.8	2.38E+09	5.2283	4.0944	4.0232	2.57E+09	5.3005	4.0874	4.1109

Table 58: Monte Carlo results of marginal effects for LIML estimations of NB-SAR data ($\alpha = 1/2$) and NB-SAR model.

Total Marginal Effects

RMSE	X_1				X_2			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0871	0.0343	0.0142	0.0102	0.0254	0.0108	0.0054	0.0035
0.2	0.1053	0.0574	0.0405	0.0370	0.0510	0.0393	0.0366	0.0366
0.4	0.1338	0.1138	0.1091	0.1084	0.1085	0.1074	0.1073	0.1072
0.6	0.3281	0.3070	0.3117	0.3087	0.3167	0.3106	0.3075	0.3085
0.8	1.7348	1.7890	1.7537	1.7769	1.7539	1.7707	1.7559	1.7729
Bias								
0	−0.0019	−0.0014	−0.0007	−0.0014	−0.0009	−0.0005	−0.0008	−0.0003
0.2	−0.0404	−0.0378	−0.0364	−0.0345	−0.0435	−0.0374	−0.0362	−0.0364
0.4	−0.0963	−0.1073	−0.1072	−0.1075	−0.1046	−0.1068	−0.1072	−0.1071
0.6	−0.3084	−0.3031	−0.3111	−0.3083	−0.3140	−0.3102	−0.3074	−0.3084
0.8	−1.7179	−1.7859	−1.7532	−1.7766	−1.7524	−1.7704	−1.7558	−1.7729
Relative bias								
0	−0.0135	−0.0104	−0.0053	−0.0103	−0.0067	−0.0034	−0.0060	−0.0022
0.2	−0.2162	−0.2034	−0.1959	−0.1856	−0.2327	−0.2014	−0.1950	−0.1958
0.4	−0.3422	−0.3830	−0.3809	−0.3819	−0.3718	−0.3811	−0.3808	−0.3804
0.6	−0.5490	−0.5535	−0.5715	−0.5643	−0.5590	−0.5663	−0.5646	−0.5646
0.8	−0.7502	−0.7629	−0.7540	−0.7595	−0.7653	−0.7563	−0.7552	−0.7579

Direct Marginal Effects

RMSE	X_1				X_2			
$\lambda\backslash n$	**1000**	**5000**	**25000**	**50000**	**1000**	**5000**	**25000**	**50000**
0	0.0871	0.0343	0.0142	0.0102	0.0254	0.0108	0.0054	0.0035
0.2	0.0973	0.0432	0.0177	0.0135	0.0277	0.0120	0.0054	0.0042
0.4	0.0936	0.0380	0.0199	0.0138	0.0291	0.0117	0.0055	0.0043
0.6	0.1123	0.0486	0.0199	0.0168	0.0411	0.0151	0.0076	0.0048
0.8	0.2425	0.1059	0.0426	0.0331	0.0759	0.0332	0.0142	0.0099
Bias								
0	−0.0019	−0.0014	−0.0007	−0.0014	−0.0009	−0.0005	−0.0008	−0.0003
0.2	−0.0041	−0.0017	−0.0003	0.0016	−0.0072	−0.0013	−0.0002	−0.0003
0.4	0.0108	−0.0007	−0.0001	−0.0004	0.0024	−0.0001	−0.0001	0.0000
0.6	0.0089	0.0062	−0.0037	0.0003	0.0033	−0.0008	0.0000	0.0001
0.8	0.0145	−0.0146	0.0067	−0.0063	−0.0200	0.0009	0.0040	−0.0026
Relative bias								
0	−0.0135	−0.0104	−0.0053	−0.0103	−0.0067	−0.0034	−0.0060	−0.0022
0.2	−0.0274	−0.0113	−0.0022	0.0107	−0.0478	−0.0089	−0.0011	−0.0020
0.4	0.0618	−0.0038	−0.0006	−0.0022	0.0140	−0.0008	−0.0003	0.0002
0.6	0.0365	0.0261	−0.0156	0.0011	0.0134	−0.0035	0.0002	0.0005
0.8	0.0261	−0.0256	0.0118	−0.0112	−0.0359	0.0016	0.0071	−0.0046

Table 59: Monte Carlo results of marginal effects for QML estimations of NB-SAR data ($\alpha = 1/2$) and non-spatial Poisson model. Values reported as 0.0000 are smaller than 0.00005.

A.3 Descriptives of the Start-Up Firm Birth Data Set

The following table contains descriptive statistics for the start-up firm births data set described in Section 3.6.1.

Variable	Mean	Median	Std. Dev.	Min.	Max.
subirths	32.8363	8.0000	157.5482	0	6938
msemp	15.1887	13.4459	10.3547	0	63.6628
tfdens	0.0518	0.0084	0.5997	0.0000	31.9162
pel10emp	52.1149	50.0000	19.9908	0	100
pem100emp	11.0494	9.7571	9.9253	0	100
mhhi	35.2150	33.6920	8.7417	12.6920	82.9290
pop	9.1035	2.5339	29.5729	0.0065	954.5829
cclass	17.1841	15.6987	5.9427	3.8793	54.0700
uer	4.3191	4.0000	1.6419	1.4000	17.5000
pedas	5.7015	5.6925	1.9857	0.3831	15.6037
awage	24.6862	23.6085	5.5921	13.6730	74.3810
netflow	7.5757	3.8629	16.8373	−461.6669	102.2310
proad	1.8341	1.6830	1.5096	0.0279	20.8855
interst	14.6878	0	25.2290	0	398.3130
hwypc	1.7746	1.2796	2.5008	0	76.0398
avland	31.3027	23.1203	25.9624	0	98.2397
educpc	11.8356	10.8124	11.6976	0	561.5168
bci	5.9071	5.8300	0.9869	3.9700	8.3000
metro	0.3454	-	-	-	-
micro	0.2160	-	-	-	-

Table 60: Descriptives of start-up firm births data set, $n = 3078$. Values reported as 0.0000 are smaller than 0.00005.

A.4 Empirical Results for SAR Models

In this section the full results of the parameter estimates for the start-up firm births data set are reported. The tables contain the results for each of the 4 employed spatial weight matrices (see Section 3.6.1) and the non-spatial variant for the P-SAR, NB-SAR, ZIP-SAR and HP-SAR models.

Variable	W_{dnn}	W_{con}	W_{nn}	W_d	Poisson
rho		0.2533***	0.2774***	0.2306***	−0.0353
	(0.0634)	(0.0655)	(0.0537)	(0.5407)	
const	−1.1407***	−1.1344***	−1.1636***	−0.8438	−0.9337***
	(0.2416)	(0.2392)	(0.2417)	(1.4942)	(0.2818)
msemp	0.0284***	0.0282***	0.0288***	0.0306***	0.0306***
	(0.0039)	(0.0038)	(0.0039)	(0.0042)	(0.0041)
pelt10	−0.0007	−0.0003	−0.0008	−0.0021	−0.0021
	(0.0020)	(0.0020)	(0.0020)	(0.0022)	(0.0019)
pemt100	−0.0270***	−0.0276***	−0.0269***	−0.0293***	−0.0294***
	(0.0035)	(0.0033)	(0.0034)	(0.0041)	(0.0036)
tfdens	0.0047	0.0056	0.0019	0.0067	0.0062
	(0.0103)	(0.0106)	(0.01)	(0.0115)	(0.0102)
mhhi	−0.0122	−0.0104	−0.0113	0.0006	0.0002
	(0.0072)	(0.0075)	(0.0075)	(0.0106)	(0.0089)
pop	0.0022***	0.0021***	0.0023***	0.0023***	0.0023***
	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0004)
cclass	0.0498***	0.0484***	0.0521***	0.0481***	0.0484***
	(0.0121)	(0.012)	(0.0119)	(0.0137)	(0.0129)
uer	0.0603***	0.0667***	0.0649***	0.0725***	0.0730***
	(0.0192)	(0.0187)	(0.0196)	(0.0244)	(0.0223)
pedas	0.1051***	0.1036***	0.1082***	0.1306***	0.1300***
	(0.0151)	(0.0161)	(0.0164)	(0.0254)	(0.0208)
awage	0.0191***	0.0166**	0.0179**	0.0187***	0.0186***
	(0.0074)	(0.0081)	(0.0076)	(0.0068)	(0.0072)
netflow	0.0021	0.0018	0.0018	0.0019	0.0019
	(0.0022)	(0.0022)	(0.0022)	(0.0023)	(0.0023)
proad	0.067***	0.0661***	0.0741***	0.1038***	0.1028***
	(0.0175)	(0.018)	(0.017)	(0.0249)	(0.0185)
interst	0.0062***	0.0066***	0.0063***	0.0073***	0.0073***
	(0.0011)	(0.0009)	(0.0011)	(0.0011)	(0.0010)
avland	−0.0051***	−0.0051***	−0.0055***	−0.0088***	−0.0086***
	(0.0011)	(0.0012)	(0.0012)	(0.0034)	(0.0015)
bci	0.0878***	0.0767**	0.0869***	0.0791	0.0801**
	(0.0269)	(0.0273)	(0.0283)	(0.0465)	(0.0371)
educpc	0.0027	0.0034	0.0027	0.0038	0.0036**
	(0.0020)	(0.0022)	(0.0018)	(0.0024)	(0.0019)
hwypc	−0.0252	−0.0258	−0.0228	−0.0313	−0.0300
	(0.0192)	(0.0190)	(0.0172)	(0.0254)	(0.0193)
metro	1.2405***	1.2085***	1.2461***	1.2660***	1.2651***
	(0.0906)	(0.0841)	(0.0883)	(0.0894)	(0.0922)
micro	0.6430***	0.6171***	0.6335***	0.5723***	0.5726***
	(0.0634)	(0.0597)	(0.0618)	(0.0633)	(0.0630)

Table 61: P-SAR estimates for the start-up firm births data set with 4 different weighting matrices and the non-spatial Poisson model, robust standard errors in brackets. ** and *** denote a 5% and 1% significance.

Variable	$\boldsymbol{W}_{dnn}$	$\boldsymbol{W}_{con}$	$\boldsymbol{W}_{nn}$	$\boldsymbol{W}_{d}$	**NB**
rho	0.2803***	0.2902***	0.2932***	0.5416***	
	(0.0297)	(0.0321)	(0.0257)	(0.0534)	
const	−0.8673***	−0.8857***	−0.8464***	−2.0638***	−1.0664***
	(0.1703)	(0.1491)	(0.1513)	(0.2104)	(0.1950)
msemp	0.0438***	0.0431***	0.0437***	0.0475***	0.0500***
	(0.0022)	(0.0022)	(0.0021)	(0.0021)	(0.0023)
pelt10	0.0051***	0.0053***	0.0051***	0.0056***	0.0051***
	(0.0011)	(0.0012)	(0.0011)	(0.0012)	(0.0012)
pemt100	−0.0165***	−0.0163***	−0.0166***	−0.0185***	−0.0183***
	(0.0024)	(0.0021)	(0.0021)	(0.0023)	(0.0023)
tfdens	−0.0599***	−0.0619***	−0.0624***	−0.0567***	−0.0527***
	(0.0160)	(0.0163)	(0.0171)	(0.0142)	(0.0134)
mhhi	0.0089	0.0102	0.0075**	0.0207***	0.0265***
	(0.0063)	(0.0053)	(0.0037)	(0.0055)	(0.0053)
pop	0.0151***	0.0147***	0.0153***	0.0183***	0.0182
	(0.0026)	(0.0026)	(0.0025)	(0.0030)	(0.0031)***
cclass	0.0828***	0.0812***	0.0842***	0.0858***	0.0817***
	(0.0046)	(0.0046)	(0.0044)	(0.0049)	(0.0049)
uer	0.0548***	0.0544***	0.0529***	0.0716***	0.0802***
	(0.0133)	(0.0097)	(0.0087)	(0.0128)	(0.0128)
pedas	0.0336***	0.0338***	0.0326***	0.0344***	0.0444***
	(0.0082)	(0.008)	(0.0076)	(0.0094)	(0.0090)
awage	−0.0373***	−0.0373***	−0.0378***	−0.0419***	−0.0380***
	(0.0079)	(0.0065)	(0.0052)	(0.0066)	(0.0065)
netflow	−0.0166***	−0.0169***	−0.0168***	−0.0163***	−0.0156***
	(0.0028)	(0.0026)	(0.0023)	(0.0027)	(0.0027)
proad	0.0511***	0.0557***	0.0518***	0.0564***	0.0830***
	(0.0190)	(0.0187)	(0.0168)	(0.0198)	(0.0225)
interst	0.0041***	0.0042***	0.0040***	0.0049***	0.0047***
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0008)
avland	−0.0042***	−0.0043***	−0.0040***	−0.0045***	−0.0066***
	(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
bci	0.0292**	0.0264**	0.0284***	0.0365**	0.0338**
	(0.0146)	(0.0113)	(0.0107)	(0.0143)	(0.0146)
educpc	0.0036	0.0037	0.0037**	0.0033	0.0038
	(0.0026)	(0.0022)	(0.0018)	(0.0022)	(0.0027)
hwypc	−0.0277	−0.0266	−0.0271	−0.0226	−0.0276
	(0.0203)	(0.0183)	(0.0143)	(0.0163)	(0.0210)
metro	0.8314***	0.8236***	0.8397***	0.8600***	0.8452***
	(0.0491)	(0.0463)	(0.0477)	(0.0520)	(0.0536)
micro	0.5368***	0.5288***	0.5360***	0.5381***	0.5462***
	(0.0349)	(0.0344)	(.0349)	(0.0359)	(0.0377)
α	0.4131***	0.4120***	0.4104***	0.4267***	0.4365***
	(0.0228)	(0.0225)	(0.0225)	(0.0413)	(0.0236)

Table 62: NB-SAR estimates for the start-up firm births data set with 4 different weighting matrices and the non-spatial NB model, robust standard errors in brackets. ** and *** denote a 5% and 1% significance.

Variable	W_{dnn}	W_{con}	W_{nn}	W_d	ZIP
rho	0.1685**	0.2000**	0.1878***	0.0030	
	(0.0728)	(0.0784)	(0.0559)		
const	−0.6154**	−0.6130**	−0.6523**	−0.3650	−0.3576
	(0.2810)	(0.2800)	(0.2787)		(0.3154)
msemp	0.0286***	0.0282***	0.0290***	0.0300	0.0301***
	(0.0037)	(0.0036)	(0.0036)		(0.0039)
pelt10	−0.0022	−0.0019	−0.0015	−0.0020	−0.0025
	(0.0028)	(0.0029)	(0.0028)		(0.0028)
pemt100	−0.0360***	−0.0362***	−0.0357***	0.0380	−0.0390***
	(0.0041)	(0.0039)	(0.0039)		(0.0041)
tfdens	0.0081	0.0090	0.0045	0.0070	0.0072
	(0.0118)	(0.0123)	(0.0115)		(0.0111)
mhhi	0.0119	0.0134	0.0131	0.0260	0.0260**
	(0.0133)	(0.0129)	(0.0122)		(0.0125)
pop	0.0025***	0.0024***	0.0025***	0.0030	0.0026***
	(0.0004)	(0.0004)	(0.0004)		(0.0005)
cclass	0.0349***	0.0338***	0.0365***	0.0330	0.0327**
	(0.0133)	(0.0130)	(0.0129)		(0.0135)
uer	0.0680***	0.0716***	0.0694***	0.0760	0.0759***
	(0.0204)	(0.0201)	(0.0207)		(0.0230)
pedas	0.10297***	0.1000***	0.0992***	0.1110	0.1111***
	(0.0177)	(0.0189)	(0.0180)		(0.0224)
awage	−0.0015	−0.0045	−0.0057	−0.0090	−0.0090
	(0.0114)	(0.0118)	(0.0113)		(0.0118)
netflow	−0.0117**	−0.0125**	−0.0138***	−0.0150	−0.0153***
	(0.0057)	(0.0055)	(0.0051)		(0.0053)
proad	0.0748***	0.0733***	0.0774***	0.1000	0.1000***
	(0.0197)	(0.0204)	(0.0195)		(0.0215)
interst	0.0059***	0.0061***	0.0057***	0.0060	0.0065***
	(0.0010)	(0.0009)	(0.0010)		(0.0010)
avland	−0.0073***	−0.0072***	−0.0072***	−0.0100	−0.0097***
	(0.0016)	(0.0016)	(0.0015)		(0.0015)
bci	0.0772***	0.0694**	0.0759***	0.0640	0.0643**
	(0.0273)	(0.0284)	(0.0279)		(0.0354)
educpc	0.0022	0.0029	0.0021	0.0030	0.0029
	(0.0022)	(0.0023)	(0.0021)		(0.0021)
hwypc	−0.0242	−0.0262	−0.0226	−0.0290	−0.0288
	(0.0222)	(0.0218)	(0.0216)		(0.0233)
metro	1.3394***	1.3221***	1.3496***	1.3520	1.3520***
	(0.0996)	(0.0948)	(0.0967)		(0.0975)
micro	0.5953***	0.5857***	0.5958***	0.5230	0.5233***
	(0.0625)	(0.0627)	(0.0628)		(0.0684)

Table 63: ZIP-SAR estimates for the start-up firm births data set with 4 different weighting matrices and the non-spatial ZIP model, robust standard errors in brackets. ** and *** denote a 5% and 1% significance.

Variable	W_{dnn}	W_{con}	W_{nn}	W_d	HP
rho	0.1686**	0.2001**	0.1879***	0.0019	
	(0.0727)	(0.0786)	(0.0559)	(0.3201)	
const	−0.6169**	−0.6143**	−0.6534**	−0.3631	−0.3582
	(0.2812)	(0.2815)	(0.2786)	(0.8040)	(17.5158)
msemp	0.0286***	0.0283***	0.0291***	0.0302***	0.0302
	(0.0037)	(0.0036)	(0.0037)	(0.0039)	(0.2195)
pelt10	−0.0022	−0.0019	−0.0015	−0.0025	−0.0025
	(0.0028)	(0.0029)	(0.0028)	(0.0028)	(0.1537)
pemt100	−0.0362***	−0.0363***	−0.0358***	−0.0391***	−0.0391
	(0.0041)	(0.0040)	(0.0040)	(0.0042)	(0.2506)
tfdens	0.0081	0.0090	0.0045	0.0072	0.0072
	(0.0118)	(0.0124)	(0.0116)	(0.0130)	(0.6192)
mhhi	0.0119	0.0134	0.0131	0.0256	0.0257
	(0.0133)	(0.0129)	(0.0123)	(0.0139)	(0.6962)
pop	0.0025***	0.0024***	0.0025***	0.0026***	0.0026
	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0250)
cclass	0.0349***	0.0338***	0.0365***	0.0327**	0.0327
	(0.0133)	(0.0131)	(0.0129)	(0.0140)	(0.7487)
uer	0.0680***	0.0717***	0.0695***	0.0759***	0.0759
	(0.0204)	(0.0202)	(0.0207)	(0.0230)	(1.2780)
pedas	0.1031***	0.1002***	0.0994***	0.1112***	0.1112
	(0.0177)	(0.0189)	(0.0180)	(0.0251)	(1.2468)
awage	−0.0015	−0.0045	−0.0057	−0.0090	−0.0090
	(0.0113)	(0.0117)	(0.0115)	(0.0116)	(0.6583)
netflow	−0.0117**	−0.0125**	−0.0138***	−0.0153***	−0.0153
	(0.0057)	(0.0055)	(0.0051)	(0.0052)	(0.2924)
proad	0.0749***	0.0733***	0.0774***	0.1000***	0.1001
	(0.0197)	(0.0204)	(0.0195)	(0.0215)	(1.1925)
interst	0.0059***	0.0061***	0.0057***	0.0065***	0.0065
	(0.0010)	(0.0009)	(0.0010)	(0.0010)	(0.0544)
avland	−0.0073***	−0.0072***	−0.0072***	−0.0097***	−0.0097
	(0.0016)	(0.0016)	(0.0015)	(0.0022)	(0.0842)
bci	0.0775***	0.0696**	0.0761***	0.0646	0.0645
	(0.0274)	(0.0288)	(0.0279)	(0.0359)	(1.9631)
educpc	0.0023	0.0030	0.0021	0.0030	0.0030
	(0.0022)	(0.0023)	(0.0022)	(0.0027)	(0.1213)
hwypc	−0.0251	−0.0271	−0.0234	−0.0298	−0.0299
	(0.0229)	(0.0224)	(0.0224)	(0.0294)	(1.4328)
metro	1.3392***	1.3219***	1.3497***	1.3514***	1.3513
	(0.1001)	(0.0960)	(0.0974)	(0.0981)	(5.4671)
micro	0.5956***	0.5859***	0.5963***	0.5234***	0.5233
	(0.0632)	(0.0636)	(0.0634)	(0.0689)	(3.8623)

Table 64: HP-SAR estimates for the start-up firm births data set with 4 different weighting matrices and the non-spatial HP model, robust standard errors in brackets. ** and *** denote a 5% and 1% significance.

	P-SAR			NB-SAR			ZIP-SAR		
Variable	**Total M.E.**	**Direct M.E.**	**Indirect M.E.**	**Total M.E.**	**Direct M.E.**	**Indirect M.E.**	**Total M.E.**	**Direct M.E.**	**Indirect M.E.**
msemp	0.4854***	0.2969***	0.1651***	0.7533***	0.4097***	0.2575***	0.1654***	0.0079**	0.1078***
	(0.0780)	(0.0423)	(0.0530)	(0.0552)	(0.0226)	(0.0335)	(0.0447)	(0.0032)	(0.0382)
pelt10	−0.0052	−0.0032	−0.0018	0.0926***	0.0504***	0.0317***	−0.0051	0.0009	−0.0073
	(0.0360)	(0.0217)	(0.0130)	(0.0204)	(0.0107)	(0.0082)	(0.0116)	(0.0005)	(0.0113)
pemt100	−0.4751***	−0.2906***	−0.1615***	−0.2849***	−0.1549***	−0.0974***	−0.1285***	0.0023	−0.1384***
	(0.0773)	(0.0364)	(0.0541)	(0.0384)	(0.0201)	(0.0168)	(0.0442)	(0.0023)	(0.0449)
tfdens	0.0964	0.0590	0.0328	−1.0819***	−0.5884***	−0.3698***	−0.0293	−0.0694	0.0344
	(0.2105)	(0.1213)	(0.0806)	(0.3091)	(0.1593)	(0.1167)	(0.0757)	(0.0409)	(0.0495)
mhhi	−0.1790	−0.1095	−0.0609	0.1783	0.0970	0.0609	0.0506	−0.0001	0.0512
	(0.1393)	(0.0830)	(0.0534)	(0.0968)	(0.0538)	(0.0322)	(0.0482)	(0.0018)	(0.0469)
pop	0.0361***	0.0221***	0.0123**	0.2569***	0.1397***	0.0878***	0.0387***	0.0145	0.0092**
	(0.0094)	(0.0044)	(0.0052)	(0.0487)	(0.0245)	(0.0171)	(0.0125)	(0.0085)	(0.0037)
cclass	0.8331***	0.5096***	0.2833***	1.4193***	0.7719***	0.4851***	0.1931**	0.0087	0.1292
	(0.2116)	(0.1306)	(0.1037)	(0.1233)	(0.0492)	(0.0743)	(0.0792)	(0.0045)	(0.0717)
uer	1.1481***	0.7022***	0.3904**	0.9508***	0.5171***	0.3250***	0.2871**	0.0017	0.2737**
	(0.3546)	(0.2030)	(0.1688)	(0.1931)	(0.1088)	(0.0725)	(0.1270)	(0.0037)	(0.1277)
pedas	1.7833***	1.0907***	0.6064***	0.5908***	0.3213***	0.2019***	0.3693**	−0.0020	0.3822***
	(0.2978)	(0.1727)	(0.1899)	(0.1283)	(0.0697)	(0.0479)	(0.1472)	(0.0034)	(0.1478)
awage	0.2857**	0.1748**	0.0972	−0.6520***	−0.3546***	−0.2228***	−0.0147	0.0008	−0.0172
	(0.1351)	(0.0875)	(0.0496)	(0.1201)	(0.0620)	(0.0465)	(0.0481)	(0.0022)	(0.0461)
netflow	0.0310	0.0190	0.0105	−0.2954***	−0.1607***	−0.1010***	−0.0419**	0.0016	−0.0478**
	(0.0406)	(0.0244)	(0.0149)	(0.0474)	(0.0244)	(0.0201)	(0.0210)	(0.0010)	(0.0223)
proad	1.1378***	0.6959***	0.3869***	0.9736***	0.5295***	0.3327***	0.2525**	−0.0069	0.2802**
	(0.2900)	(0.1926)	(0.1313)	(0.3092)	(0.1716)	(0.1158)	(0.1127)	(0.0082)	(0.1168)
interst	0.1136***	0.0695***	0.0386***	0.0734***	0.0399***	0.0251***	0.0301***	0.0008	0.0233***
	(0.0169)	(0.0094)	(0.0119)	(0.0131)	(0.0067)	(0.0056)	(0.0080)	(0.0005)	(0.0075)
avland	−0.0878***	−0.0537***	−0.0298***	−0.0752***	−0.0409***	−0.0257***	−0.0258***	0.0004	−0.0275***
	(0.0200)	(0.0125)	(0.0100)	(0.0097)	(0.0056)	(0.0039)	(0.0085)	(0.0003)	(0.0087)
bci	1.3202***	0.8075***	0.4489**	0.4614***	0.2510***	0.1577***	0.2630	−0.0004	0.2653**
	(0.4886)	(0.2902)	(0.2115)	(0.1687)	(0.0918)	(0.0599)	(0.1349)	(0.0034)	(0.1346)
educpc	0.0585	0.0358	0.0199	0.0647	0.0352	0.0221	0.0105	−0.0001	0.0111
	(0.0345)	(0.0212)	(0.0131)	(0.0390)	(0.0214)	(0.0132)	(0.0087)	(0.0004)	(0.0082)
hwypc	−0.4441	−0.2716	−0.1510	−0.4649	−0.2529	−0.1589	−0.1063	−0.0008	−0.1001
	(0.2973)	(0.1816)	(0.1118)	(0.3022)	(0.1662)	(0.1028)	(0.0861)	(0.0014)	(0.0848)
metro	28.2409***	9.5402***	18.5824***	14.6455***	5.1013***	9.4061***	27.1320***	12.3288***	14.7558***
	(5.3945)	(1.7572)	(6.0259)	(1.1690)	(0.4884)	(0.7344)	(4.2862)	(2.0028)	(4.6865)
micro	10.1053***	3.5118***	6.5528***	8.0310***	2.8809***	5.1485***	−0.1041	3.9093***	−2.1495
	(1.9977)	(0.6095)	(2.1706)	(1.4793)	(0.4368)	(1.3724)	(1.4684)	(0.6390)	(1.3431)

Table 65: Median marginal effects of P-SAR, NB-SAR, ZIP-SAR for the start-up firm births data set with weight matrix W_{con}. For the dummy variables *metro* and *micro* the effect of a change from 0 to 1 is given. ** and *** denote a 5% and 1% significance, respectively, Standard errors (in brackets) are estimated using their sample counterparts of 2000 draws of the asymptotic joint distribution of the coefficients.

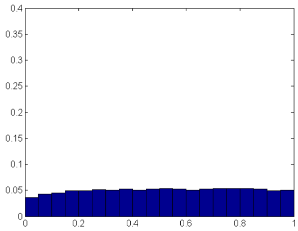
Variable	P-SAR Total M.E.	P-SAR Direct M.E.	P-SAR Indirect M.E.	NB-SAR Total M.E.	NB-SAR Direct M.E.	NB-SAR Indirect M.E.	ZIP-SAR Total M.E.	ZIP-SAR Direct M.E.	ZIP-SAR Indirect M.E.
msemp	0.4628***	0.3031***	0.1357***	0.7473***	0.4139***	0.2553***	0.1652***	0.0081**	0.1055***
	(0.0681)	(0.0410)	(0.0405)	(0.0475)	(0.0223)	(0.0324)	(0.0474)	(0.0032)	(0.0395)
pelt10	−0.0129	−0.0084	−0.0038	0.0872***	0.0483***	0.0298***	−0.0036	0.0010	−0.0055
	(0.0332)	(0.0216)	(0.0102)	(0.0198)	(0.0107)	(0.0078)	(0.0119)	(0.0005)	(0.0111)
pemt100	−0.4323***	−0.2831***	−0.1267***	−0.2839***	−0.1572***	−0.0970***	−0.1213***	0.0022	−0.1299***
	(0.0639)	(0.0373)	(0.0383)	(0.0373)	(0.0200)	(0.0168)	(0.0442)	(0.0023)	(0.0444)
tfdens	0.0305	0.0200	0.0090	−1.0671***	−0.5910***	−0.3645***	−0.0507	−0.0738	0.0164
	(0.1679)	(0.1065)	(0.0533)	(0.2970)	(0.1585)	(0.1135)	(0.0893)	(0.0439)	(0.0465)
mhhi	−0.1816	−0.1189	−0.0532	0.1283	0.0710	0.0438	0.0464	−0.0002	0.0477
	(0.1295)	(0.0836)	(0.0419)	(0.0950)	(0.0533)	(0.0320)	(0.0468)	(0.0018)	(0.0455)
pop	0.0370***	0.0242***	0.0108**	0.2617***	0.1449***	0.0894***	0.0400***	0.0153	0.0091**
	(0.0103)	(0.0055)	(0.0047)	(0.0425)	(0.0242)	(0.0153)	(0.0135)	(0.0091)	(0.0037)
cclass	0.8373***	0.5484***	0.2455***	1.4399***	0.7975***	0.4918***	0.2008**	0.0090	0.1328
	(0.1939)	(0.1300)	(0.0816)	(0.1218)	(0.0504)	(0.0762)	(0.0823)	(0.0047)	(0.0736)
uer	1.0430***	0.6831***	0.3058**	0.9047***	0.5011***	0.3090***	0.2659**	0.0017	0.2526**
	(0.3388)	(0.2066)	(0.1356)	(0.1943)	(0.1091)	(0.0730)	(0.1220)	(0.0039)	(0.1226)
pedas	1.7389***	1.1389***	0.5098***	0.5575***	0.3088***	0.1904***	0.3484**	−0.0022	0.3610**
	(0.2853)	(0.1729)	(0.1563)	(0.1197)	(0.0679)	(0.0442)	(0.1440)	(0.0033)	(0.1454)
awage	0.2877**	0.1884**	0.0843**	−0.6464***	−0.3580***	−0.2208***	−0.0181	0.0009	−0.0207
	(0.1229)	(0.0837)	(0.0398)	(0.1137)	(0.0608)	(0.0451)	(0.0469)	(0.0023)	(0.0447)
netflow	0.0289	0.0189	0.0085	−0.2873***	−0.1591***	−0.0981***	−0.0450**	0.0017	−0.0502**
	(0.0384)	(0.0248)	(0.0119)	(0.0474)	(0.0242)	(0.0204)	(0.0216)	(0.0010)	(0.0226)
proad	1.1908***	0.7800***	0.3491***	0.8858***	0.4906***	0.3026***	0.2577**	−0.0069	0.2817**
	(0.2560)	(0.1772)	(0.1082)	(0.3111)	(0.1692)	(0.1159)	(0.1185)	(0.0087)	(0.1206)
interst	0.1012***	0.0663***	0.0297***	0.0684***	0.0379***	0.0234***	0.0275***	0.0008	0.0207***
	(0.0160)	(0.0115)	(0.0081)	(0.0129)	(0.0067)	(0.0054)	(0.0079)	(0.0005)	(0.0071)
avland	−0.0884***	−0.0579***	−0.0259***	−0.0684***	−0.0379***	−0.0234***	−0.0247***	0.0004	−0.0262***
	(0.0189)	(0.0124)	(0.0084)	(0.0098)	(0.0056)	(0.0039)	(0.0087)	(0.0003)	(0.0088)
bci	1.3965***	0.9147***	0.4094**	0.4857***	0.2690***	0.1659***	0.2725**	−0.0005	0.2762**
	(0.4588)	(0.2938)	(0.1710)	(0.1634)	(0.0910)	(0.0590)	(0.1342)	(0.0035)	(0.1350)
educpc	0.0434	0.0284	0.0127	0.0633	0.0350	0.0216	0.0071	−0.0001	0.0076
	(0.0285)	(0.0189)	(0.0089)	(0.0388)	(0.0217)	(0.0132)	(0.0087)	(0.0004)	(0.0080)
hwypc	−0.3664	−0.2400	−0.1074	−0.4634	−0.2567	−0.1583	−0.0889	−0.0008	−0.0822
	(0.2670)	(0.1750)	(0.0842)	(0.2958)	(0.1654)	(0.1013)	(0.0840)	(0.0015)	(0.0822)
metro	26.9964***	10.8807***	16.0255***	15.0467***	5.0940***	9.8002***	27.3782***	12.8813***	14.3528***
	(4.7115)	(1.6549)	(5.0202)	(1.1850)	(0.4979)	(0.7386)	(4.3013)	(2.2206)	(4.8874)
micro	9.6429***	3.9303***	5.6273***	8.2422***	2.8435***	5.4109***	−0.3396	4.0780***	−2.4291
	(1.6890)	(0.5819)	(1.7795)	(1.5044)	(0.4568)	(1.3929)	(1.5238)	(0.6761)	(1.4565)

Table 66: Median marginal effects of P-SAR, NB-SAR, ZIP-SAR for the start-up firm births data set with weight matrix W_{nn}. For the dummy variables *metro* and *micro* the effect of a change from 0 to 1 is given. ** and *** denote a 5% and 1% significance, respectively, Standard errors (in brackets) are estimated using their sample counterparts of 2000 draws of the asymptotic joint distribution of the coefficients.

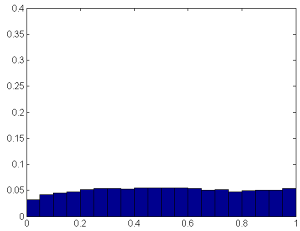
B Further Results for the SLF Models

B.1 Monte Carlo Results for PIT Histograms

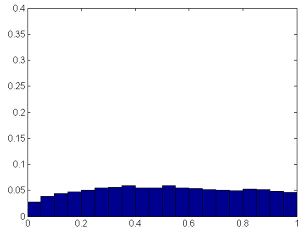
This section reports the PIT histograms for the Monte Carlo results from P-SLFM estimated with P-SLF data and NB-SLF data, respectively (see Section 4.4.2).



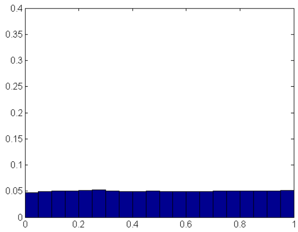
(a) $\lambda = 0$



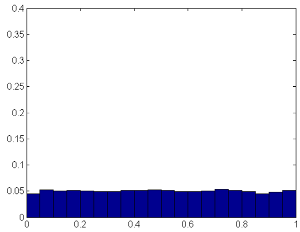
(b) $\lambda = 0$



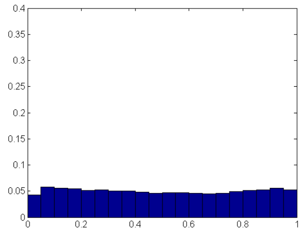
(c) $\lambda = 0$



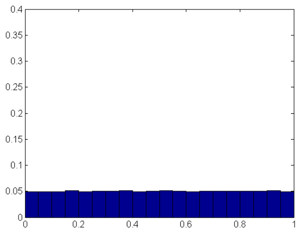
(d) $\lambda = 0.2$



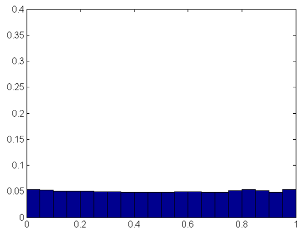
(e) $\lambda = 0.2$



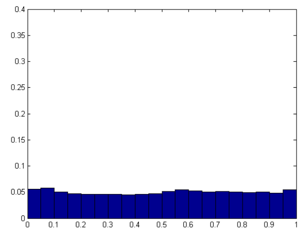
(f) $\lambda = 0.2$



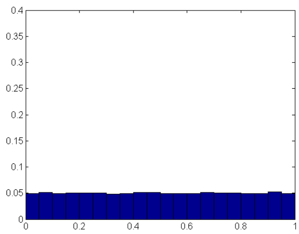
(g) $\lambda = 0.4$



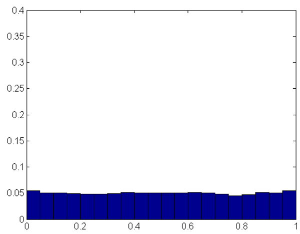
(h) $\lambda = 0.4$



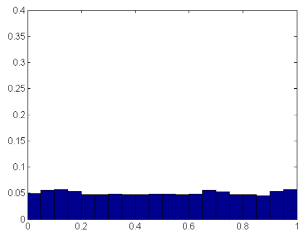
(i) $\lambda = 0.4$



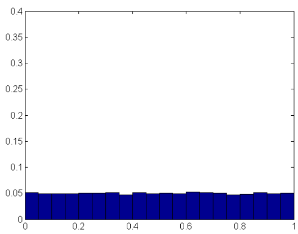
(j) $\lambda = 0.6$



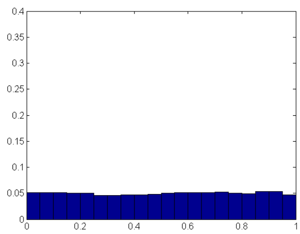
(k) $\lambda = 0.6$



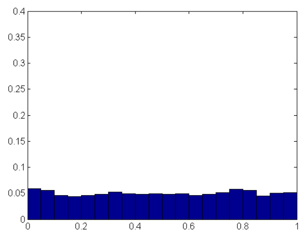
(l) $\lambda = 0.6$



(m) $\lambda = 0.8$

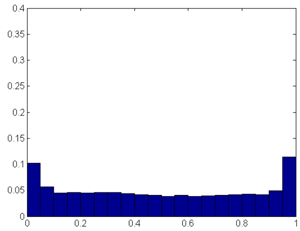


(n) $\lambda = 0.8$

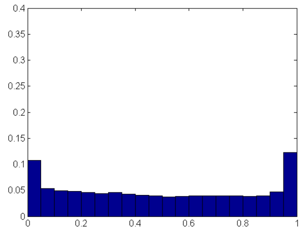


(o) $\lambda = 0.8$

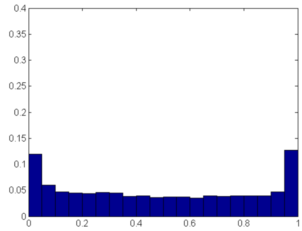
Figure 20: PIT histograms of Monte Carlo results for P-SLFM. n=5000.



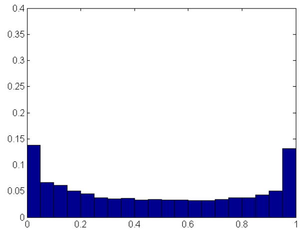
(a) $\lambda = 0$



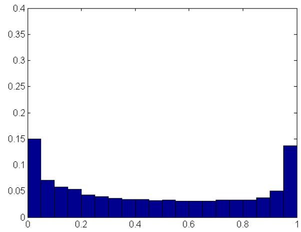
(b) $\lambda = 0$



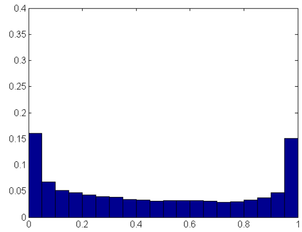
(c) $\lambda = 0$



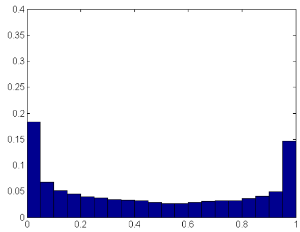
(d) $\lambda = 0.2$



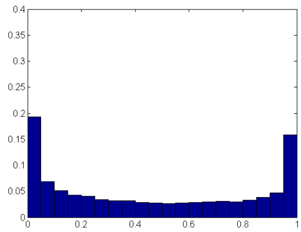
(e) $\lambda = 0.2$



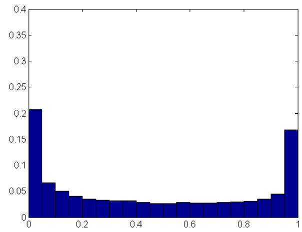
(f) $\lambda = 0.2$



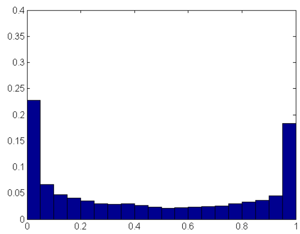
(g) $\lambda = 0.4$



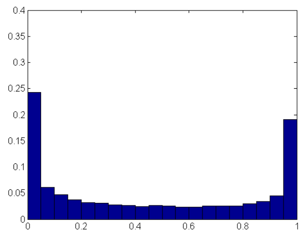
(h) $\lambda = 0.4$



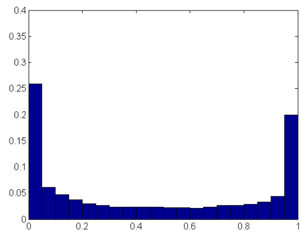
(i) $\lambda = 0.4$



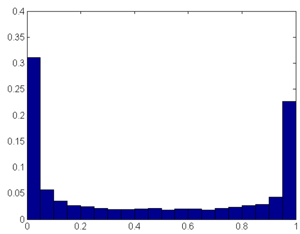
(j) $\lambda = 0.6$



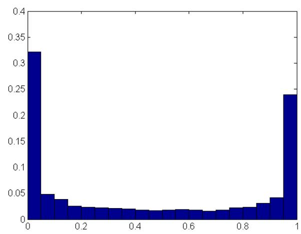
(k) $\lambda = 0.6$



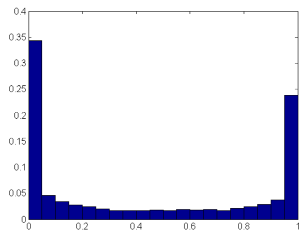
(l) $\lambda = 0.6$



(m) $\lambda = 0.8$



(n) $\lambda = 0.8$



(o) $\lambda = 0.8$

Figure 21: PIT histograms of Monte Carlo results for NB-SLFM estimated with P-SLFM. n=5000.

B.2 Monte Carlo Parameter Estimates

This section contains additional Monte Carlo results for the SLF models, for which the standard specifications in Section 4.4.2 are alternated. First, a different spatial weight matrix is employed, afterwards the parameter vector β is changed.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda \backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.0296	0.0091	0.0038	0.0028	0.1425	0.0433	0.0177	0.0131
0.2	0.0418	0.0125	0.0060	0.0045	0.1532	0.0505	0.0229	0.0172
0.4	0.0389	0.0143	0.0063	0.0044	0.1802	0.0650	0.0281	0.0194
0.6	0.0353	0.0109	0.0056	0.0041	0.2583	0.0674	0.0335	0.0248
0.8	0.0440	0.0090	0.0053	0.0042	0.6374	0.0924	0.0484	0.0356
Bias								
0	−0.0049	−0.0001	0.0002	0.0001	0.0051	0.0004	−0.0014	−0.0003
0.2	−0.0047	−0.0011	−0.0016	−0.0014	0.0081	0.0033	0.0051	0.0050
0.4	−0.0025	−0.0010	−0.0013	−0.0012	−0.0045	0.0047	0.0063	0.0057
0.6	−0.0039	0.0001	−0.0008	−0.0007	−0.0015	0.0003	0.0054	0.0049
0.8	−0.0078	−0.0028	−0.0029	−0.0028	−0.0747	−0.0002	0.0001	−0.0002
Relative bias								
0	-	-	-	-	0.0102	0.0008	−0.0028	−0.0007
0.2	−0.0236	−0.0056	−0.0079	−0.0071	0.0162	0.0067	0.0101	0.0100
0.4	−0.0064	−0.0025	−0.0032	−0.0031	−0.0089	0.0094	0.0125	0.0114
0.6	−0.0065	0.0001	−0.0013	−0.0011	−0.0031	0.0007	0.0107	0.0098
0.8	−0.0098	−0.0036	−0.0037	−0.0035	−0.1494	−0.0005	0.0002	−0.0003
	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda \backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.0557	0.0215	0.0092	0.0067	0.0242	0.0080	0.0032	0.0023
0.2	0.0638	0.0221	0.0103	0.0073	0.0231	0.0074	0.0039	0.0028
0.4	0.0812	0.0262	0.0117	0.0080	0.0251	0.0105	0.0043	0.0030
0.6	0.0923	0.0297	0.0134	0.0095	0.0378	0.0091	0.0051	0.0037
0.8	0.2369	0.0351	0.0172	0.0121	0.0881	0.0121	0.0067	0.0052
Bias								
0	0.0007	−0.0003	0.0007	0.0001	−0.0013	−0.0002	0.0001	0.0001
0.2	−0.0022	−0.0004	−0.0010	−0.0008	−0.0015	−0.0008	−0.0008	−0.0008
0.4	0.0017	−0.0015	−0.0007	−0.0008	−0.0003	−0.0009	−0.0011	−0.0009
0.6	0.0042	−0.0006	−0.0002	−0.0010	−0.0009	−0.0001	−0.0011	−0.0008
0.8	0.0384	−0.0009	0.0000	0.0000	0.0093	0.0001	0.0000	0.0000
Relative bias								
0	0.0015	−0.0007	0.0013	0.0002	−0.0026	−0.0004	0.0003	0.0002
0.2	−0.0044	−0.0007	−0.0020	−0.0016	−0.0030	−0.0017	−0.0015	−0.0015
0.4	0.0034	−0.0030	−0.0013	−0.0017	−0.0006	−0.0019	−0.0023	−0.0019
0.6	0.0084	−0.0012	−0.0004	−0.0019	−0.0018	−0.0002	−0.0023	−0.0016
0.8	0.0768	−0.0017	−0.0001	0.0001	0.0187	0.0002	0.0000	0.0000

Table 67: Monte Carlo results for P-SLFM data with a contiguity matrix as spatial weighting matrix. The bias is calculated as the average difference between estimates and true parameter value. The relative bias is the absolute value of the bias divided by the true parameter value. Values reported as 0.0000 are smaller than 0.00005.

	$\hat{\lambda}$				$\hat{\beta}_0$			
$\lambda\backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.0000	0.0000	0.0000	0.0000	0.0073	0.0000	0.0000	0.0000
0.2	0.0005	0.0002	0.0001	0.0000	0.0150	0.0025	0.0008	0.0006
0.4	0.0004	0.0002	0.0001	0.0001	0.0086	0.0021	0.0007	0.0006
0.6	0.0007	0.0002	0.0001	0.0001	0.0151	0.0025	0.0008	0.0007
0.8	0.0029	0.0029	0.0031	0.0028	0.0138	0.0059	0.0031	0.0043
Bias								
0	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000
0.2	0.0000	0.0000	0.0000	0.0000	−0.0019	0.0000	0.0000	0.0000
0.4	0.0000	0.0000	0.0000	0.0000	−0.0003	0.0000	0.0000	0.0000
0.6	0.0000	0.0000	0.0000	0.0000	0.0004	−0.0001	−0.0001	−0.0001
0.8	−0.0028	−0.0028	−0.0031	−0.0028	−0.0027	−0.0048	−0.0029	−0.0042
Relative bias								
0	-	-	-	-	0.0001	0.0000	0.0000	0.0000
0.2	0.0000	0.0000	0.0000	0.0000	−0.0012	0.0000	0.0000	0.0000
0.4	0.0000	0.0000	0.0000	0.0000	−0.0002	0.0000	0.0000	0.0000
0.6	−0.0001	−0.0001	−0.0001	0.0000	0.0003	−0.0001	0.0000	−0.0001
0.8	−0.0035	−0.0036	−0.0038	−0.0036	−0.0018	−0.0032	−0.0019	−0.0028

	$\hat{\beta}_1$				$\hat{\beta}_2$			
$\lambda\backslash n$	100	1000	5000	10000	100	1000	5000	10000
RMSE								
0	0.0018	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000
0.2	0.0064	0.0007	0.0003	0.0002	0.0017	0.0003	0.0001	0.0001
0.4	0.0035	0.0008	0.0003	0.0002	0.0013	0.0003	0.0001	0.0001
0.6	0.0045	0.0008	0.0003	0.0003	0.0024	0.0003	0.0001	0.0001
0.8	0.0028	0.0014	0.0004	0.0009	0.0021	0.0008	0.0005	0.0005
Bias								
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	0.0007	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000
0.4	−0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
0.6	−0.0001	0.0000	0.0000	0.0000	−0.0001	0.0000	0.0000	0.0000
0.8	0.0008	0.0005	0.0001	0.0009	0.0003	0.0007	0.0005	0.0005
Relative bias								
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	0.0005	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
0.4	−0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.6	0.0000	0.0000	0.0000	0.0000	−0.0001	0.0000	0.0000	0.0000
0.8	0.0005	0.0003	0.0001	0.0006	0.0002	0.0005	0.0003	0.0003

Table 68: Monte Carlo results for P-SLFM data with $\beta = [1.5, 1.5, 1.5]$. The bias is calculated as the average difference between estimates and true parameter value. The relative bias is the absolute value of the bias divided by the true parameter value. Values reported as 0.0000 are smaller than 0.00005.

B.3 Empirical Results for SLF Models

The following table displays estimation results of the SLF models for the start-up firm births data set using alternative spatial weight matrices (see Section 4.5).

	$\boldsymbol{W_{con}}$		$\boldsymbol{W_{nn}}$	
Variable	**P-SLFM**	**NB-SLFM**	**P-SLFM**	**NB-SLFM**
λ	0.2202***	0.1345***	0.2487***	0.1304***
	(0.0438)	(0.0188)	(0.0398)	(0.0160)
const	−1.4865***	−1.2060***	−1.5958***	−1.1837***
	(0.3686)	(0.4491)	(0.3861)	(0.2138)
msemp	0.0339***	0.0531***	0.0351***	0.0531***
	(0.0052)	(0.0026)	(0.0053)	(0.0026)
pelt10	−0.0049**	0.0054***	−0.0060**	0.0052***
	(0.0024)	(0.0013)	(0.0026)	(0.0013)
pemt100	−0.0322***	−0.0222***	−0.0339***	−0.0218***
	(0.0048)	(0.0029)	(0.0052)	(0.0030)
tfdens	−0.0075	−0.0517***	−0.0106	−0.0507***
	(0.0107)	(0.0143)	(0.0109)	(0.0132)
mhhi	−0.0220**	0.0226***	−0.0274***	0.0212***
	(0.0094)	(0.0085)	(0.0086)	(0.0072)
pop	0.0021***	0.0163***	0.0022***	0.0166***
	(0.0004)	(0.0031)	(0.0005)	(0.0032)
cclass	0.0754***	0.0983***	0.0823***	0.0998***
	(0.0120)	(0.0059)	(0.0109)	(0.0060)
uer	0.0468	0.0749***	0.0408	0.0721***
	(0.0321)	(0.0151)	(0.0350)	(0.0140)
pedas	0.1530***	0.0539***	0.1559***	0.0535***
	(0.0232)	(0.0130)	(0.0240)	(0.0104)
awage	0.0278***	−0.0520***	0.0287***	−0.0521***
	(0.0079)	(0.0122)	(0.0082)	(0.0089)
netflow	0.0030	−0.0247***	0.0032	−0.0240***
	(0.0024)	(0.0052)	(0.0024)	(0.0046)
proad	0.1023***	0.0905***	0.1020***	0.0885***
	(0.0222)	(0.0340)	(0.0230)	(0.0254)
interst	0.0083***	0.0053***	0.0081***	0.0052***
	(0.0011)	(0.0009)	(0.0013)	(0.0009)
avland	−0.0076***	−0.0060***	−0.0074***	−0.0059***
	(0.0017)	(0.0009)	(0.0017)	(0.0009)
bci	0.1008**	0.0300	0.1125***	0.0321
	(0.0414)	(0.0641)	(0.0427)	(0.0204)
educpc	0.0048**	0.0120***	0.0050**	0.0119***
	(0.0021)	(0.0043)	(0.0021)	(0.0031)
hwypc	−0.0332	−0.0957***	−0.0339	−0.0930***
	(0.0222)	(0.0341)	(0.0221)	(0.0232)
metro	1.4780***	0.9824***	1.5866***	1.0004***
	(0.1331)	(0.0754)	(0.1480)	(0.0745)
micro	0.7126***	0.6237***	0.7944***	0.6344***
	(0.0966)	(0.0505)	(0.1086)	(0.0499)
α		0.4071***		0.4037***
		(0.0248)		(0.0246)
Log L	−29494	−10309	−29248	−10303
LogS	9.2792	3.3502	9.1869	3.3482
QS	−0.0263	−0.0730	−0.0256	−0.0726
RPS	14.4995	23.7479	14.3704	24.6600

Table 69: Estimation results from P-SLF and NB-SLF for the start-up firm births data set with weight matrices W_{con} and W_{nn}. N=3078, Heteroscedasticity robust standard errors in brackets, calculated with the sandwich formula (White, 1980), ** and *** denote a 5% and 1% significance, respectively.

C Further Results for the Panel Models

C.1 Empirical Results for the P-SLFP Model

Here, results for the P-SLFPM (see Section 5.3) using a summer dummy are displayed first. Afterwards, the following tables contain detailed results for the model with monthly dummies which is discussed in the main text.

	(2)	(3) fixed effects	(4) fixed effects	(5) fixed effects
$WPartI_t$	0.1685***	0.1529***	0.1427***	0.1272
	(0.0087)	(0.0306)	(0.0529)	(0.0699)
$WPartI_{t-1}$		0.0446**		0.0401
		(0.0208)		(0.0517)
$log(PartII_{t-1})$	0.8126***	0.1061	0.1087	
	(0.0066)	(0.0585)	(0.0967)	
summer	0.0421**	0.1251***	0.1272***	0.1396***
	(0.0199)	(0.0334)	(0.0417)	(0.0391)
Log L	−26858	−223940	−223968	−223954

Table 70: Estimation results from the P-SLFPM with a summer dummy for the Pittsburgh crime data. The estimation sample is January 2008 to December 2012. Robust standard errors in parentheses; W is a queen contiguity spatial weighting matrix; ** and *** denote 5% and 1% statistical significance.

	(2)		(3)		(4)		(5)	
$WPartI_t$	x		x		x		x	
$WPartI_{t-1}$			x				x	
$log(PartII_{t-1})$	x		x		x			
fixed effects			x		x		x	
summer dummy	x		x		x		x	
	RMSFE	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**
Jan 13	4.9292	3.1812	3.1033	2.3551	3.1714	2.3768	3.1427	2.4130
Feb 13	4.5437	3.4275	3.4505	2.6449	3.7504	2.8333	3.5793	2.7536
Mar 13	4.7739	3.2971	3.4662	2.4928	3.5529	2.5942	3.4336	2.4855
Apr 13	6.1568	3.8043	3.7926	2.7609	3.7455	2.7536	3.7552	2.7246
May 13	5.1941	3.9638	4.0271	2.9855	4.4827	3.1377	3.9673	2.9710
Jun 13	5.6735	3.5652	4.3597	3.1087	4.6703	3.2754	4.4019	3.1739
Jul 13	5.6209	3.1739	3.4881	2.6739	3.6978	2.8043	3.5733	2.7391
Aug 13	7.0721	3.7681	4.0975	2.7754	4.1205	2.8333	4.2006	2.8478
Sep 13	6.7066	3.8478	3.9489	2.7826	4.0352	2.8768	3.9873	2.8551
Oct 13	6.0774	3.6594	3.3805	2.5435	3.6166	2.6014	3.3794	2.5362
Nov 13	5.5950	3.6522	3.2737	2.3986	3.3665	2.4783	3.3210	2.4638
Dec 13	5.5063	3.6377	3.7888	2.6304	3.7281	2.6377	3.8795	2.7174
Average	5.6541	3.5815	3.6814	2.6793	3.8282	2.7669	3.7184	2.7234

Table 71: Point forecast evaluation of P-SLFPM with summer dummy for the Pittsburgh crime data.

	(2)			**(3)**			**(4)**			**(5)**		
$WPartI_t$	x			x			x			x		
$WPartI_{t-1}$				x						x		
$log(PartII_{t-1})$	x			x			x					
fixed effects				x			x			x		
summer dummy	x			x			x			x		
	LogS	**QS**	**RPS**	**LogS**	**QS**	**RPS**	**LogS**	**QS**	**RPS**	**LogS**	**QS**	**RPS**
Jan 13	3.0496	−0.0805	2.4057	2.4310	−0.1091	1.6185	2.4598	−0.1051	1.6647	2.4504	−0.1070	1.6492
Feb 13	2.9485	−0.0664	2.4451	2.4953	−0.1061	1.8504	2.5480	−0.1039	1.9553	2.5234	−0.1033	1.8818
Mar 13	3.2974	−0.0739	2.4607	2.4990	−0.1151	1.7690	2.5184	−0.1122	1.8351	2.4927	−0.1154	1.7686
Apr 13	3.4554	−0.0540	2.9145	2.5987	−0.1034	1.9487	2.6080	−0.1026	1.9626	2.5957	−0.1043	1.9452
May 13	3.3551	−0.0442	2.8986	2.6515	−0.1064	2.1106	2.7288	−0.1033	2.2756	2.6422	−0.1075	2.0804
Jun 13	3.1610	−0.0743	2.6725	2.7977	−0.0908	2.2389	2.8571	−0.0888	2.3596	2.8064	−0.0896	2.2508
Jul 13	2.9655	−0.0744	2.3540	2.5020	−0.1079	1.8285	2.5519	−0.1046	1.9386	2.5250	−0.1070	1.8667
Aug 13	3.3551	−0.0572	2.8515	2.6245	−0.1160	2.0350	2.6586	−0.1132	2.0889	2.6597	−0.1142	2.0913
Sep 13	3.2466	−0.0630	2.8987	2.5571	−0.1150	1.9635	2.5991	−0.1117	2.0342	2.5794	−0.1124	2.0075
Oct 13	3.1734	−0.0578	2.7286	2.4644	−0.1193	1.7571	2.4822	−0.1190	1.8261	2.4755	−0.1184	1.7642
Nov 13	3.1855	−0.0627	2.7033	2.3332	−0.1289	1.6607	2.3551	−0.1250	1.7090	2.3454	−0.1272	1.6845
Dec 13	3.1722	−0.0735	2.6904	2.4921	−0.1332	1.8875	2.4831	−0.1320	1.8579	2.5166	−0.1300	1.9242
Average	3.1971	−0.0652	2.6686	2.5372	−0.1126	1.8890	2.5708	−0.1101	1.9590	2.5510	−0.1114	1.9095

Table 72: Scoring rules of density forecasts from P-SLFPM with summer dummy for the Pittsburgh crime data. $\hat{y}_{iT+1} \sim Po(\hat{\mu}_{iT+1} = E[\widehat{y_{iT+1}}|.])$.

	(1)		(2)		(3)		(4)		(5)	
$WPartI_t$	x		x		x		x		x	
$WPartI_{t-1}$					x				x	
$PartII_{t-1}$	x		x		x		x			
fixed effects					x		x		x	
time dummies			x		x		x		x	
	RMSFE	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**
Jan 13	4.7457	3.2029	4.7028	3.2464	2.9952	2.1884	3.0288	2.2319	3.0657	2.2246
Feb 13	4.4673	3.4203	4.2758	2.9638	2.8233	2.1159	2.7452	2.0725	2.8194	2.1232
Mar 13	4.6974	3.2536	4.7105	3.3188	3.4829	2.5362	3.4631	2.5000	3.4599	2.5362
Apr 13	6.1432	3.9130	6.1426	3.8043	3.8307	2.7899	3.8505	2.8261	3.8382	2.7899
May 13	5.2316	3.9348	5.2847	4.0145	3.8420	2.9058	4.1980	3.0580	3.7936	2.8841
Jun 13	5.6202	3.6159	5.5769	3.6087	3.9791	2.9203	4.1380	2.9493	4.0154	2.9783
Jul 13	5.4100	3.1522	5.4785	3.2754	3.3297	2.5942	3.5763	2.7464	3.4093	2.6377
Aug 13	6.8826	3.7319	6.8079	3.9275	4.1064	2.8043	4.1868	2.9203	4.2169	2.8841
Sep 13	6.5097	3.8116	6.5779	3.8043	3.8373	2.6087	3.9955	2.7029	3.8183	2.6087
Oct 13	5.8117	3.7029	5.7879	3.6884	3.5509	2.6667	3.6435	2.7101	3.6286	2.6884
Nov 13	5.3717	3.6522	5.4407	3.6159	3.2348	2.3623	3.3922	2.5072	3.3166	2.4493
Dec 13	5.3805	3.6014	5.3024	3.5797	3.7965	2.5870	3.6968	2.5652	3.8523	2.6232
Average	5.5226	3.5827	5.5074	3.5707	3.5674	2.5900	3.6596	2.6492	3.6029	2.6190

Table 73: Point forecast evaluation of P-SLFPM for the Pittsburgh crime data.

	(1)			(2)			(3)			(4)			(5)		
$WPartI_t$		x			x			x			x			x	
$WPartI_{t-1}$								x						x	
$PartII_{t-1}$		x			x			x			x				
fixed effects								x			x			x	
time dummies					x			x			x			x	
	LogS	**QS**	**RPS**	**LogS**	**QS**	**RPS**	**LogS**	**QS**	**RPS**	**LogS**	**QS**	**RPS**	**LogS**	**QS**	**RPS**
Jan 13	3.1036	−0.0796	2.4149	3.0839	−0.0774	2.4567	2.4032	−0.1157	1.5580	2.4227	−0.1123	1.5772	2.4115	−0.1153	1.5695
Feb 13	2.9819	−0.0681	2.4413	2.8760	−0.0906	2.1988	2.2715	−0.1300	1.4613	2.2756	−0.1294	1.4618	2.2914	−0.1274	1.4854
Mar 13	3.3814	−0.0742	2.4438	3.3567	−0.0713	2.4778	2.5130	−0.1133	1.8036	2.5088	−0.1132	1.8127	2.5063	−0.1138	1.7957
Apr 13	3.5075	−0.0508	2.9602	3.5294	−0.0526	2.9349	2.6060	−0.1027	1.9702	2.6236	−0.1011	1.9957	2.6103	−0.1030	1.9823
May 13	3.4177	−0.0424	2.9316	3.4282	−0.0414	2.9757	2.6204	−0.1088	2.0315	2.6825	−0.1065	2.1643	2.6156	−0.1096	2.0128
Jun 13	3.2115	−0.0728	2.7031	3.2003	−0.0735	2.6924	2.7056	−0.0945	2.0702	2.7398	−0.0942	2.1283	2.7177	−0.0929	2.0926
Jul 13	2.9663	−0.0707	2.3216	2.9759	−0.0694	2.3776	2.4766	−0.1100	1.7729	2.5241	−0.1067	1.8760	2.4944	−0.1093	1.8025
Aug 13	3.3592	−0.0566	2.8400	3.3461	−0.0537	2.9084	2.6356	−0.1153	2.0511	2.6819	−0.1117	2.1238	2.6705	−0.1136	2.1071
Sep 13	3.2383	−0.0616	2.8727	3.2495	−0.0625	2.8709	2.5052	−0.1234	1.8643	2.5475	−0.1202	1.9289	2.5145	−0.1219	1.8798
Oct 13	3.1921	−0.0542	2.7397	3.1896	−0.0538	2.7554	2.5205	−0.1125	1.8429	2.5366	−0.1110	1.8879	2.5419	−0.1105	1.8735
Nov 13	3.2143	−0.0603	2.7043	3.2142	−0.0610	2.6895	2.3322	−0.1291	1.6588	2.3597	−0.1243	1.7181	2.3436	−0.1273	1.6810
Dec 13	3.2322	−0.0695	2.7037	3.2213	−0.0697	2.7049	2.4860	−0.1364	1.8585	2.4764	−0.1348	1.8309	2.5042	−0.1340	1.8838
Average	3.2338	−0.0634	2.6731	3.2226	−0.0647	2.6703	2.5063	−0.1160	1.8286	2.5316	−0.1138	1.8755	2.5185	−0.1149	1.8472

Table 74: Scoring rules of density forecasts from P-SLFPM for the Pittsburgh crime data. $\hat{y}_{iT+1} \sim Po(\hat{\mu}_{iT+1} = E[\widehat{y_{iT+1}}|.])$.

C.2 Empirical Results for the Dynamic Panel Model with Multiplicative Fixed Effects

This section contains the results of the dynamic panel model with multiplicative fixed effects (see Section 5.4) using a summer dummy, as well as the detailed results for the specifications discussed in the main text which employ monthly dummies.

	(1)	**(2)**	**(3)**
$PartI_{t-1}$	0.0297	0.0626	0.4102
	(0.3498)	(0.2052)	(0.9670)
$WPartI_t$	0.6264	1.0539	
	(0.4063)	(2.0784)	
$WPartI_{t-1}$	−0.0603		0.2823
	(0.3184)		(0.2077)
$log(PartII_{t-1})$	0.4854	−0.0774	1.5706
	(1.7490)	(0.6553)	(6.1786)
Summer	0.2412	−0.0942	−0.5537
	(2.6811)	(3.5259)	(4.5134)
Sargan	0.0001	0.0002	0.0017
	[1.0000]	[1.0000]	[1.0000]

Table 75: Estimation results from the multiplicative fixed effects model with summer dummy for the Pittsburgh crime data. The estimation sample is January 2008 to December 2012. Standard errors in parentheses; W is a queen contiguity spatial weighting matrix; ** and *** denote 5% and 1% statistical significance.

	(1)		**(2)**		**(3)**	
$PartI_{t-1}$	x		x		x	
$WPartI_t$	x		x			
$WPartI_{t-1}$	x				x	
$log(PartII_{t-1})$	x		x		x	
summer dummy	x		x		x	
	RMSFE	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**
Jan 13	6.3975	3.8551	10.8964	8.3696	35.0525	9.6667
Feb 13	4.7610	3.5217	4.6811	3.4928	9.4282	5.6449
Mar 13	4.6943	3.4565	4.7898	3.5217	5.6773	4.1739
Apr 13	7.3834	4.7609	6.9413	4.5000	22.4359	9.8913
May 13	6.1479	4.4058	5.9026	4.1014	17.3029	7.7971
Jun 13	6.3731	4.5145	6.2479	4.4420	24.3211	8.1667
Jul 13	6.3045	3.9928	7.6447	5.6449	19.5926	7.1594
Aug 13	4.9152	3.6522	4.3171	3.3478	10.2073	6.0435
Sep 13	5.8198	3.8261	5.5443	3.6812	27.2823	9.8623
Oct 13	4.5802	3.3986	4.6757	3.4420	7.8482	4.9710
Nov 13	5.2212	3.7391	5.0440	3.6449	13.9860	6.3478
Dec 13	6.2537	4.2536	6.2241	4.2029	21.1031	6.9348
Average	5.7376	3.9481	6.0758	4.3659	17.8531	7.2216

Table 76: Point forecast evaluation of the multiplicative fixed effects model with summer dummy for the Pittsburgh crime data.

	(1)		(2)		(3)		(4)		(5)	
$PartI_{t-1}$	x		x		x		x		x	
$WPartI_t$	x		x				x		x	
$WPartI_{t-1}$	x				x		x			
$log(PartII_{t-1})$	x		x		x		x		x	
time dummies	x		x		x					
	RMSFE	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**
Jan 13	6.8271	4.0000	5.8508	3.6667	36.0714	9.7826	7.2372	4.1449	6.2118	3.7464
Feb 13	4.6415	3.4130	4.3862	3.2681	9.3843	5.6159	4.9454	3.6739	4.6819	3.5000
Mar 13	4.6811	3.4348	4.5636	3.3478	5.8588	4.1957	4.7868	3.4928	4.7860	3.5145
Apr 13	7.5939	4.8696	7.0072	4.5507	23.4126	9.7609	7.8850	4.9855	7.0243	4.5290
May 13	6.1972	4.3043	5.9381	4.1014	15.7347	7.3188	6.2843	4.3623	5.9338	4.1087
Jun 13	6.4847	4.5580	6.2130	4.4420	22.7995	7.7899	6.6042	4.5870	6.2797	4.4638
Jul 13	6.3177	4.0290	5.9338	3.8913	19.6506	7.1159	6.4807	4.0870	5.9509	3.8913
Aug 13	4.9731	3.7029	4.8492	3.6594	8.6510	5.5362	5.0188	3.7246	4.8335	3.5942
Sep 13	6.1444	4.1304	5.5730	3.6667	24.4388	9.1812	6.0762	3.9928	5.5443	3.6812
Oct 13	4.8447	3.5435	4.7746	3.5217	7.4547	4.7174	4.7212	3.5072	4.6780	3.4493
Nov 13	5.2963	3.8478	5.1259	3.7826	12.0545	5.9493	5.4534	3.8986	5.0727	3.6594
Dec 13	6.3850	4.3188	6.2693	4.2319	27.0225	8.1594	6.5203	4.3406	6.2837	4.2101
Average	5.8656	4.0127	5.5404	3.8442	17.7111	7.0936	6.0011	4.0664	5.6067	3.8623

Table 77: Point forecast evaluation of the multiplicative fixed effects model for the Pittsburgh crime data.

C.3 Empirical Results for the Dynamic Panel Model with Additive Fixed Effects

This last section reports the results of the dynamic panel model with additive fixed effects (see Section 5.5) using a summer dummy, as well as the detailed results for the specifications discussed in the main text which employ monthly dummies.

	(1)	(2)	(3)	(5)
$PartI_{t-1}$	0.0653***	0.0396	0.0236	0.0776***
	(0.0207)	(0.0247)	(0.0236)	(0.0178)
$WPartI_t$	0.9207***	0.3684		0.8246***
	(0.1651)	(0.2522)		(0.0823)
$WPartI_{t-1}$	0.0178		0.0348	−0.0048
	(0.0486)		(0.0261)	(0.0477)
$log(PartII_{t-1})$	−1.0567	0.6299***	0.7990***	
	(5.7351)	(0.1621)	(0.0204)	
Summer	−0.7181	0.2769	0.8857***	0.2696
	(0.4923)	(0.4915)	(0.1999)	(0.3999)
Sargan	0.0913	0.1140	0.1163	0.0247
	[1.0000]	[1.0000]	[1.0000]	[0.9999]

Table 78: Estimation results from the non-linear additive fixed effects model with summer dummy for the Pittsburgh crime data, $g(x_it) = \exp(x_{it}\beta)$. The estimation sample is January 2008 to December 2012. Standard errors in parentheses; W is a queen contiguity spatial weighting matrix; ** and *** denote 5% and 1% statistical significance.

	(1)		(2)		(3)		(5)	
$PartI_{t-1}$	x		x		x		x	
$WPartI_t$	x		x				x	
$WPartI_{t-1}$	x				x		x	
$log(PartII_{t-1})$	x		x		x			
summer dummy	x		x		x		x	
	RMSFE	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**
Jan 13	4.2264	3.1957	4.3788	3.3913	4.8611	3.7899	4.2767	3.1739
Feb 13	4.0638	3.0942	4.1046	3.1232	4.4306	3.3986	4.0984	3.1304
Mar 13	4.6039	3.5580	4.5739	3.5290	4.6070	3.5580	4.5556	3.4638
Apr 13	5.7294	4.1449	5.5358	4.0072	5.5795	4.0290	4.8871	3.5652
May 13	5.1259	3.6812	5.9618	4.4565	5.5808	4.1739	5.3771	4.0435
Jun 13	5.4686	4.0507	5.5397	4.1087	5.8092	4.3841	5.3263	3.9058
Jul 13	4.4632	3.3551	4.4274	3.3116	5.1288	3.8986	4.3879	3.2971
Aug 13	4.3197	3.2681	4.3489	3.2899	5.1330	3.9420	4.3138	3.2899
Sep 13	5.1633	3.5725	7.6053	5.1449	5.6862	3.8261	5.1506	3.5000
Oct 13	5.5599	4.0870	5.5069	4.0507	4.7404	3.6739	4.6920	3.3478
Nov 13	4.2944	3.1812	4.6827	3.6522	4.9651	3.8406	4.2238	3.1594
Dec 13	5.0641	3.8768	4.9241	3.6812	4.9454	3.6884	4.8663	3.6522
Average	4.8402	3.5888	5.1325	3.8122	5.1222	3.8502	4.6796	3.4607

Table 79: Point forecast evaluation of the non-linear additive fixed effects models with summer dummy for the Pittsburgh crime data, $g(x_it) = \exp(x_{it}\beta)$.

	(1)	(2)	(3)
$PartI_{t-1}$	0.0735***	0.0865***	0.0462**
	(0.0179)	(0.0197)	(0.0184)
$WPartI_t$	1.1051***	1.5961***	
	(0.2629)	(0.3679)	
$WPartI_{t-1}$	0.0106		0.0379
	(0.0399)		(0.0374)
$log(PartII_{t-1})$	−0.7010	−2.3338	2.7099***
	(0.8068)	(1.2130)	(0.1987)
summer	−0.7953	−1.5875**	1.1295***
	(0.5506)	(0.7233)	(0.3309)
Sargan	0.0760	0.0571	0.1393
	[1.0000]	[1.0000]	[1.0000]

Table 80: Estimation results from the linear additive fixed effects model with summer dummy for the Pittsburgh crime data, $g(x_i t) = x_{it}\beta$. The estimation sample is January 2008 to December 2012. Standard errors in parentheses; W is a queen contiguity spatial weighting matrix; ** and *** denote 5% and 1% statistical significance.

	(1)		(2)		(3)	
$PartI_{t-1}$	x		x		x	
$WPartI_t$	x		x			
$WPartI_{t-1}$	x				x	
$log(PartII_{t-1})$	x		x		x	
summer dummy	x		x		x	
	RMSFE	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**
Jan 13	4.7143	3.6884	13.4579	9.1304	4.6219	3.5797
Feb 13	12.3951	9.3333	19.5163	13.0000	4.3121	3.3768
Mar 13	4.6780	3.5507	5.5384	4.3261	4.5309	3.4565
Apr 13	5.4214	4.1014	16.7267	11.7101	5.2689	4.0072
May 13	10.9167	8.3768	10.1317	7.5942	5.3406	3.9710
Jun 13	5.3188	3.8841	5.3202	3.9130	5.3270	3.9420
Jul 13	4.4175	3.2971	4.4151	3.2754	4.5253	3.4348
Aug 13	4.3163	3.3116	10.2999	9.4493	4.7304	3.6957
Sep 13	9.5246	7.6449	11.0552	8.0145	5.3710	3.7609
Oct 13	24.4961	15.8841	6.0834	4.6449	4.6835	3.5290
Nov 13	4.4510	3.3768	5.1591	4.1087	4.6819	3.6304
Dec 13	15.7914	10.7029	21.9627	13.8551	4.8252	3.6159
Average	8.8701	6.4293	10.8055	7.7518	4.8516	3.6667

Table 81: Point forecast evaluation of the linear additive fixed effects model with summer dummy for the Pittsburgh crime data, $g(x_i t) = x_{it}\beta$.

	(1)		(2)		(3)		(4)		(5)	
$PartI_{t-1}$	x		x		x		x		x	
$WPartI_t$	x		x				x		x	
$WPartI_{t-1}$	x				x				x	
$log(PartII_{t-1})$	x		x		x		x			
monthly dummies	x		x		x				x	
	RMSFE	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**
Jan 13	4.8797	3.8261	4.6664	3.6159	4.9935	3.8768	4.3871	3.2609	5.2496	4.0507
Feb 13	4.1450	3.1667	4.0136	3.0362	4.2032	3.2029	4.1798	3.1957	4.1117	3.1377
Mar 13	4.5165	3.4275	4.5373	3.4275	4.5707	3.4565	7.3391	5.7609	4.3738	3.3333
Apr 13	5.1906	3.7681	4.8871	3.6232	5.4970	3.9710	36.9568	34.9928	4.9782	3.6667
May 13	5.3872	3.8913	5.0684	3.6159	5.5168	4.0145	5.7269	4.2029	5.0051	3.5435
Jun 13	5.5325	4.1014	5.4706	4.0435	5.7647	4.3333	6.0893	4.4565	5.3297	3.9565
Jul 13	4.8968	3.7029	4.3381	3.2101	5.3031	4.0217	4.6540	3.5435	4.7549	3.6087
Aug 13	4.8469	3.7391	4.6974	3.6304	5.0898	3.9348	4.4216	3.4348	4.3920	3.3188
Sep 13	5.4167	3.5870	5.2378	3.5652	5.7167	3.7246	5.1780	3.5507	5.1556	3.4928
Oct 13	4.8327	3.5725	4.7792	3.5072	4.8938	3.6594	4.8477	3.4275	4.8222	3.4420
Nov 13	4.6376	3.6232	4.4494	3.4058	4.9424	3.8623	4.2469	3.1957	4.3622	3.2754
Dec 13	5.0123	3.7319	4.7966	3.5580	5.0260	3.7536	4.9738	3.7391	4.9964	3.7319
Average	4.9412	3.6781	4.7451	3.5199	5.1265	3.8176	7.7501	6.3967	4.7943	3.5465

Table 82: Point forecast evaluation of the non-linear additive fixed effects model for the Pittsburgh crime data, $g(x_i t) = \exp(x_{it}\beta)$.

	(1)		(2)		(3)		(4)	
$PartI_{t-1}$	x		x		x		x	
$WPartI_t$	x		x				x	
$WPartI_{t-1}$	x				x			
$log(PartII_{t-1})$	x		x		x		x	
monthly dummies	x		x		x			
	RMSFE	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**	**RMSFE**	**MAFE**
Jan 13	4.7784	3.7319	4.4510	3.4058	4.8139	3.7536	4.3862	3.2391
Feb 13	4.0271	3.1449	3.9572	2.9928	4.0522	3.1739	4.3489	3.3768
Mar 13	4.4802	3.4203	4.2936	3.2609	4.4859	3.4275	5.2184	3.9855
Apr 13	5.1584	3.8406	4.9593	3.5942	5.3263	4.0362	7.6267	6.1232
May 13	5.1485	3.7246	5.0007	3.5725	5.2716	3.8043	11.5980	10.5870
Jun 13	5.3757	3.9710	5.2805	3.8696	5.4267	4.0290	5.6376	4.1304
Jul 13	4.4923	3.3551	4.4036	3.3043	4.6672	3.5507	4.6227	3.4420
Aug 13	4.5794	3.5942	4.4729	3.4565	4.7777	3.7391	4.4093	3.4275
Sep 13	5.2254	3.5942	5.1633	3.4855	5.3697	3.7319	5.2094	3.6159
Oct 13	4.8124	3.5072	4.7777	3.4058	4.9123	3.5797	4.7296	3.3551
Nov 13	4.4932	3.4348	4.4003	3.3188	4.5755	3.5580	4.3614	3.3261
Dec 13	4.8574	3.6377	4.7686	3.5507	4.8931	3.6812	4.9174	3.7029
Average	4.7857	3.5797	4.6607	3.4348	4.8810	3.6721	5.5888	4.3593

Table 83: Point forecast evaluation of the linear additive fixed effects model for the Pittsburgh crime data, $g(x_it) = x_{it}\beta$.

References

Abdelmoula, M. and Bresson, G. (2005). Patents and R&D spillovers in some European regions: a dynamic count panel data model. Working Papers ERMES 33.

Abdelmoula, M. and Bresson, G. (2007). Spatial and Technological Spillovers in European Patenting Activities: A Dynamic Count Panel Data Model. *Annals of Economics and Statistics*, 87/88:167–194.

Acosta, M., Coronado, D., and Martínez, M. Á. (2012). Spatial differences in the quality of university patenting: Do regions matter? *Research Policy*, 41(4):692–703.

Aguero-Valverde, J. and Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident; analysis and prevention*, 38(3):618–625.

Alañón Pardo, Á., Arauzo-Carod, J. M., and Myro Sánchez, R. (2007). Accessibility, agglomeration and location. In Arauzo- Carod, J. M. and Manjón-Antolín, M., editors, *Entrepreneurship, industrial location and economic growth*, pages 247–267. Chentelham: Edward Elgar.

Andersson, R., Quigley, J. M., and Wilhelmsson, M. (2009). Urbanization, productivity, and innovation: Evidence from investment in higher education. *Journal of Urban Economics*, 66(1):2–15.

Anselin, L. (2002). Under the hood. Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3):247–267.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2):310–330.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1):3–25.

Arauzo-Carod, J.-M. and Manjón-Antolín, M. (2012). (Optimal) spatial aggregation in the determinants of industrial location. *Small Business Economics*, 39(3):645–658.

Arbia, G. (2014). *A Primer for Spatial Econometrics*. Palgrave Macmillam.

Arbia, G., Bee, M., Espa, G., and Santi, F. (2014). Fitting Spatial Econometric Models through the Unilateral Approximation. DEM Discussion Papers.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58(2):277.

Arribas-Bel, D., Koschinsky, J., and Amaral, P. V. (2012). Improving the multi-dimensional comparison of simulation results: a spatial visualization approach. *Letters in Spatial and Resource Sciences*, 5(2):55–63.

Arzaghi, M. and Henderson, J. V. (2008). Networking off Madison Avenue. *Review of Economic Studies*, 75(4):1011–1038.

Augustin, N. H., McNicol, J., and Marriott, C. a. (2006). Using the truncated auto-Poisson model for spatially correlated counts of vegetation. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1):1–23.

Baltagi, B., Bresson, G., and Etienne, J. (2014). Hedonic Housing Prices in Paris: An Unbalanced Spatial Lag Pseudo-Panel Model with Nested Random Effects. *Journal of Applied Econometrics*.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data.* Chapman & Hall/CRC, New York.

Basile, R., Benfratello, L., and Castellani, D. (2013). Geoadditive Models for Regional Count Data: An Application to Industrial Location. *Geographical Analysis*, 45(1):28–48.

Beger, A. (2012). *Predicting The Intensity And Location Of Violence In War.* PhD thesis, Florida State University, College of Social Science and Public Policy.

Bernardinelli, L. and Clayton, D. (1995). Bayesian analysis of spacetime variation in disease risk. *Statistics in Medicine*, 14:2433–2443.

Bernardinelli, L. and Montomoli, C. (1992). Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in medicine*, 11(8):983–1007.

Bernardinelli, L., Pascutto, C., Best, N. G., and Gilks, W. R. (1997). Disease mapping with errors in covariates. *Statistics in medicine*, 16(7):741–52.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B*, 36(2):192–236.

Besag, J. (1975). Statistical Analysis of Non-lattice Data. *The Statistician: Journal of the Insitute of Statisticians*, 24(3):179–195.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82:733–746.

Best, N., Richardson, S., and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14(1):35–59.

Bhat, C. R., Paleti, R., and Singh, P. (2014). A spatial multivariate count model for firm location decisions. *Journal of Regional Science*, 54(3):462–502.

Bhati, A. S. (2005). Robust spatial analysis of rare crimes: An information-theoretic approach. *Statistical Methodology*, 35:227–289.

Bhati, A. S. (2008). A Generalized Cross-Entropy Approach for Modeling Spatially Correlated Counts. *Econometric Reviews*, 27(4-6):574–595.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Blundell, R., Griffith, R., and Reenen, J. V. (1995). Dynamic Count Data Models of Technological Innovation. *The Economic Journal*, 105(429):333.

Blundell, R., Griffith, R., and Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics*, 108:113–131.

Bonaccorsi, A., Colombo, M. G., Guerini, M., and Rossi-Lamastra, C. (2013). University specialization and new firm creation across industries. *Small Business Economics*, 41:837–863.

Brännäs, K. (1995). Explanatory variables in the AR(1) count data model. Umea Economic Studies (DP).

Brännäs, K. (2013). Simultaneity in the Multivariate Count Data Autoregressive Model. Umea Economic Studies (DP).

Brännäs, K. (2014). Adaptations of Conventional Spatial Econometric Models to Count Data. Umea Economic Studies (DP).

Briet, O. J. T. (2009). *Towards malaria prediction in Sri Lanka: Modelling spatial and temporal variability of malaria case counts*. PhD thesis, Universität Basel.

Brown, J. P. and Lambert, D. M. (2014). Location decisions of natural gas extraction establishments: a smooth transition count model approach. The Federal Reserve Bank of Kansas City Research Working Paper 14-05.

Brown, J. P. and Lambert, D. M. (2016). Extending a Smooth Parameter Model To Firm Location Analyses: the Case of Natural Gas Establishments in the United States. *Journal of Regional Science*, (forthcoming).

Buczkowska, S., Coulombel, N., and Lapparent, M. D. (2014). Euclidean distance versus travel time in business location: A probabilistic mixture of hurdle- Poisson models. In *ERSA conference papers*.

Buczkowska, S. and de Lapparent, M. (2014). Location choices of newly created establishments: Spatial patterns at the aggregate level. *Regional Science and Urban Economics*, 48:68–81.

Buddhavarapu, P., Scott, J. G., and Prozzi, J. A. (2016). Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data. *Transportation Research Part B: Methodological*, 91:492–510.

Cai, Q., Lee, J., Eluru, N., and Abdel-Aty, M. (2016). Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. *Accident Analysis & Prevention*, 93:14–22.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data*. Cambridge University Press, 2 edition.

Carlin, B. P. and Banerjee, S. (2003). Hierarchical Multivariate CAR Models for Spatio-Temporally Correlated Survival Data. In Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Hecherman, D., Smith, A., and West, M., editors, *Bayesian Statistics 7*, pages 45–63. Oxford University Press.

Castellani, D., Meliciani, V., and Mirra, L. (2016). The Determinants of Inward Foreign Direct Investment in Business Services across European Regions. *Regional Studies*, 50(4):671–691.

Castro, M., Paleti, R., and Bhat, C. R. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B: Methodological*, 46(1):253–272.

Chakraborty, A., Beamonte, M., Gelfand, A. E., Alonso, M. P., Gargallo, P., and Salvador, M. (2013). Spatial interaction models with individual-level data for explaining labor flows and developing local labor markets. *Computational Statistics & Data Analysis*, 58:292–307.

Chamberlain, G. (1992). Comment: Sequential Moment Restrictions in Panel Data. *Journal of Business & Economic Statistics*, 10(1):pp. 20–26.

Chiou, Y.-C., Fu, C., and Chih-Wei, H. (2014). Incorporating spatial dependence in simultaneously modeling crash frequency and severity. *Analytic Methods in Accident Research*, 2:1–11.

Chun, Y. (2014). Analyzing Space-Time Crime Incidents Using Eigenvector Spatial Filtering: An Application to Vehicle Burglary. *Geographical Analysis*, 46(2):165–184.

Cliff, A. and Ord, J. (1981). *Spatial processes: models and applications*. Pion, London.

Corsatea, T. D. and Jayet, H. (2014). Spatial patterns of innovation activities in France: market's role versus public research efforts. *Annals of Regional Science*, 52:739–762.

Cox, D. (1981). Statistical analysis of time series: some recent developements. *Scandinavian Journal of Statistics*, 8:93–115.

Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons Ltd, revised ed edition.

Cressie, N. and Chan, N. N. H. N. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, 84(406):393–401.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–61.

Czado, C., Schabenberger, H., and Erhardt, V. (2014). Non nested model selection for spatial count regression models with application to health insurance. *Statistical Papers*, 55(2):455–476.

Dawid, A. (1984). Statistical Theory: The Prequential Approach. *Journal of the Royal Statistical Society. Series A*, 147(2):278–292.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1):40–57.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–883.

Dimakos, X. and Rattalma, A. D. (2002). Bayesian premium rating with latent structure. *Scandinavian Actuarial Journal*, 3:162–184.

Elhorst, J. P. (2008). Serial and spatial error correlation. *Economics Letters*, 100(3):422–424.

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

Elhorst, J. P. (2014). *Spatial Econometrics. From Cross-Sectional Data to Spatial Panels.* Springer, Heidelberg.

Ferrandiz, J., Lopez, A., Llopis, A., Morales, M., and Tejerizo, M. L. (1995). Spatial interaction between neighbouring counties: cancer mortality data in Valencia Spain. *Biometrics*, 51(2):665–678.

Fischer, M. M., Bartkowska, M., Riedl, A., Sardadvar, S., and Kunnert, A. (2009). The impact of human capital on regional labor productivity in Europe. *Letters in Spatial and Resource Sciences*, 2(2-3):97–108.

Fischer, M. M. and Wang, J. (2011). *Spatial Data Analysis. Models, Methods and Techniques.* Springer Berlin Heidelberg.

Frühwirth-Schnatter, S. (1996). Recursive residuals and model diagnostics for normal and non-normal state space models. *Environmental and Ecological Statistics*, 3(4):291–309.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–25.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Ghodsi, A. (2015). Conditional Maximum Likelihood Estimation of the First-Order Spatial Integer-Valued Autoregressive (SINAR (1,1)) Model. *Journal of Iranian Statistical Society*, 14(2):15–36.

Ghodsi, A., Shitan, M., and Bakouch, H. S. (2012). A First-Order Spatial Integer-Valued Autoregressive SINAR(1, 1) Model. *Communications in Statistics - Theory and Methods*, 41(15):2773–2787.

Gorr, W., Olligschlaeger, A., and Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting*, 19:579–594.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700.

Graham, D. J. J., McCoy, E. J., and Stephens, D. A. (2013). Quanitifying the effect of area deprivation on child pedestrian casualties by using longitudinal mixed models to adjust for confounding, interference and spatial dependence. *Journal of the Royal Statistical Society. Series A*, 176(Part 3).

Griffith, D. A. (2002). A spatial filtering specification for the auto-Poisson model. *Statistics & Probability Letters*, 58:245–251.

Griffith, D. A. (2003). *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Springer-Verlag Berlin Heidelberg.

Griffith, D. A. (2006). Assessing Spatial Dependence in Count Data: Winsorized and Spatial Filter Specification Alternatives to the Auto-Poisson Model. *Geographical Analysis*, 38(2):160–179.

Griffith, D. A. and Paelinck, J. H. P. (2007). An equation by any other name is still the same: on spatial econometrics and spatial statistics. *The Annals of Regional Science*, 41:209–227.

Gschlößl, S. and Czado, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 2007(3):202–225.

Gschlößl, S. and Czado, C. (2008). Does a Gibbs sampler approach to spatial Poisson regression models outperform a single site MH sampler? *Computational Statistics & Data Analysis*, 52(9):4184–4202.

Haining, R., Law, J., and Griffith, D. (2009). Modelling small area counts in the presence of overdispersion and spatial autocorrelation. *Computational Statistics & Data Analysis*, 53(8):2923–2937.

Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica*, 52(4):909–938.

Hays, J. and Franzese, R. (2009). A comparison of the small-sample properties of several estimators for spatial-lag count models. Working Paper.

Held, L., Höhle, M., and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, 5(3):187–199.

Hilbe, J. M. (2011). *Negative Binomial Regression*. Cambridge University Press, 2 edition.

Jiang, X., Abdel-Aty, M., and Alamili, S. (2013). Application of Poisson random effect models for highway network screening. *Accident; analysis and prevention*, 63C:74–82.

Jones-Webb, R., McKee, P., Hannan, P., Wall, M., Pham, L., Erickson, D., and Wagenaar, A. (2008). Alcohol and malt liquor availability and promotion and homicide in inner cities. *Substance use & misuse*, 43(2):159–77.

Jung, R. C., McCabe, B. P. M., and Tremayne, A. R. (2016). Model Validation and Diagnostics. In Davis, R. A., Holan, S. H., Lund, R., and Ravishanker, N., editors, *Handbook of Discrete-Valued Time Series*, pages 189–218. Chapman and Hall/CRC.

Jung, R. C. and Tremayne, A. R. (2011). Useful models for time series of counts or simply wrong ones? *AStA Advances in Statistical Analysis*, 95(1):59–91.

Kaiser, M. S. and Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistical Papers*, 35:423–432.

Kauermann, G., Haupt, H., and Kaufmann, N. (2012). A hitchhiker's view on spatial statistics and spatial econometrics for lattice data. *Statistical Modelling*, 12(5):419–440.

Kelejian, H. H. and Prucha, I. R. (1998). A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics*, 17(1):99–121.

Kelling, G. L. and Coles, C. M. (1996). *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities.* Free Press, New York.

Kukenova, M. and Monteiro, J.-A. (2009). Spatial Dynamic Panel Model and System GMM: A Monte Carlo Investigation. IRENE Working Papers.

Lambert, D., McNamara, K. T., and Garrett, M. I. (2006). An application of spatial Poisson models to manufacturing investment location analysis. *Journal of Agricultural and Applied Economics*, 38(1):105–121.

Lambert, D. M., Brown, J. P., and Florax, R. J. (2010). A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application. *Regional Science and Urban Economics*, 40(4):241–252.

Lee, L.-f. and Yu, J. (2009). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, 40(5):255–271.

LeSage, J. P. (1999). The Theory and Practice of Spatial Econometrics. Working Paper.

LeSage, J. P., Fischer, M. M., and Scherngell, T. (2007). Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science*, 86(3):393–421.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics.* Taylor & Francis.

Li, L., Zhu, L., and Sui, D. Z. (2007). A GIS-based Bayesian approach for analyzing spatialtemporal patterns of intra-city motor vehicle crashes. *Journal of Transport Geography*, 15(4):274–285.

Liesenfeld, R., Nolte, I., and Pohlmeier, W. (2006). Modelling financial transaction price movements: a dynamic integer count data model. *Empirical Economics*, 30(4):795–825.

Liesenfeld, R., Richard, J.-F., and Vogler, J. (2016a). Likelihood-Based Inference and Prediction in Spatio-Temporal Panel Count Models for Urban Crimes. *Journal of Aplied Econometrics.*

Liesenfeld, R., Richard, J.-f., and Vogler, J. (2016b). Likelihood evaluation of high-dimensional spatial latent Gaussian models with non-Gaussian response variables. *Spatial and Spatiotemporal Econometrics, Advances in Econometrics*, (forthcoming).

Liviano, D. and Arauzo-Carod, J.-M. (2013). Industrial location and interpretation of zero counts. *The Annals of Regional Science*, 50(May 2013):515–534.

Liviano, D. and Arauzo-Carod, J.-M. (2014). Industrial Location and Spatial Dependence: An Empirical Application. *Regional Studies*, 48(4):727–743.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Sage Publications, Inc., Thousand Oaks.

Martinez Ibañez, O., Manjón Antolín, M., and Arauzo-Carod, J.-M. (2013). The Geographical Scope of Industrial Location Determinants: An Alternative Approach. *Tijdschrift voor economische en sociale geografie*, 104(2):194–214.

McKenzie, E. (1985). Some simple models for discrete variate time series1. *Journal of the American Water Resources Association*, 21(4):645–650.

Mears, D. P. and Bhati, A. S. (2006). No Community is an Island: The Effects of Resource Deprivation on Urban Violence in Spatially and Socially Proximate Communities. *Criminology*, 44(3):509–548.

Melo, O. O., Mateu, J., and Melo, C. E. (2015). A generalised linear spacetime autoregressive model with spacetime autoregressive disturbances. *Journal of Applied Statistics*, 4763(October):1–28.

Miaou, S., Song, J., and Mallick, B. (2003). Roadway traffic crash mapping: A space-time modeling approach. *Journal of Transportation and Statistics*, 6(1).

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1):17–23.

Neelon, B., Ghosh, P., and Loebs, P. F. (2013). A Spatial Poisson Hurdle Model for Exploring Geographic Variation in Emergency Department Visits. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, 176(2):389–413.

Nowrouzian, R. and Srinivasan, S. (2014). A Spatial Quasi-Poisson Model for Car Ownership. *Transportation Research B*, 32611:1–18.

Pace, R. K. (2003). Spatial Statistics Toolbox 2.0. Manual.

Paelinck, J. and Klaassen, L. (1979). *Spatial Econometrics.* Farnborough, Hants, Saxon House.

Palmgren, J. (1981). The Fisher information matrix for log-linear models arguing conditionally in the observed explanatory variables. *Biometrika*, 68:563–566.

Parent, O. and LeSage, J. P. (2011). A spacetime filter for panel data models containing random effects. *Computational Statistics & Data Analysis*, 55(1):475–490.

Payton, S. B., Stucky, T. D., and Ottensmann, J. R. (2015). The spatial extent of the effect of foreclosures on crime. *Social Science Research*, 49:288–298.

Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. *Accident; analysis and prevention*, 40(4):1486–1497.

Roodman, D. (2009). Practitioners' corner: A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23:470–472.

Savun, B. and Hays, J. C. (2011). Foreign Aid as a Counterterrorism Tool: Can Democracy Aid Reduce Terrorism? APSA 2011 Annual Meeting Paper.

Stuart, T. E. and Sorenson, O. (2003). Liquidity Events and the Geographic Distribution of Entrepreneurial Activity. *Administrative Science Quarterly*, 48(2):175–201.

Tevie, J., Bohara, A., and Valdez, R. B. (2014). Examination of the geographical variation in human West Nile virus: a spatial filtering approach. *Epidemiology and Infection*, 142(12):2522–2529.

Tobler, A. W. R. (1970). A Computer Movie Simulation Urban Growth in Detroit Region. *Economic Geography*, 46(332):234–240.

Torabi, M. (2016). Hierarchical multivariate mixture generalized linear models for the analysis of spatial data: An application to disease mapping. *Biometrical Journal*, (forthcoming).

Truong, L. T., Kieu, L.-M., and Vu, T. A. (2016). Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam. *Accident Analysis & Prevention*, 94:153–161.

U.S. Department of Justice, Federal Bureau of Investigation (2004). Uniform Crime Reporting Handbook.

Villalta, D., Guenni, L., Rubio-Palis, Y., and Ramírez Arbeláez, R. (2012). Bayesian spacetime modeling of malaria incidence in Sucre state, Venezuela. *AStA Advances in Statistical Analysis*.

Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8(2):158–83.

Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). Hierarchical Spatio-Temporal Mapping of Disease Rates. *Journal of the American Statistical Association*, 92(438):607–617.

Wang, Y. and Kockelman, K. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis and Prevention*, 60:71–84.

Ward, M. P. and Gleditsch, K. S. (2008). *Spatial Regression Models*. Sage Publications, Inc.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48(4):817–838.

Whittle, P. (1954). On Stationary Processes in the Plane. *Biometrika*, 41(3/4):434–449.

Wilson, J. Q. and Kelling, G. L. (1982). The Police and Neighborhood Safety: Broken Windows. *The Atlantic*, 249:29–38.

Windmeijer, F. (2008). GMM for Panel Count Data Models. In Mátyás, L. and Sevestre, P., editors, *The Econometrics of Panel Data*, pages 603–624. Springer Berlin Heidelberg.

Winkelmann, R. (2010). *Econometric Analysis of Count Data*. Springer Berlin Heidelberg, 5 edition.

Wooldridge, J. M. (1997). Multiplicative Panel Data Models Without the Strict Exogeneity Assumption. *Econometric Theory*, 13(5):667–678.

Xia, H., Carlin, B., and Waller, L. (1997). Hierarchical models for mapping Ohio lung cancer rates. *Environmetrics*, 8(June 1996):107–120.

Xia, H. and Carlin, B. P. (1998). Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Statistics in medicine*, 17(18):2025–43.

Zeger, S. L. (1988). A Regression Model for Time Series of Counts. *Biometrika*, 75(4):621.