

# Article Modelling and Diagnostics of Spatially Autocorrelated Counts

Robert C. Jung <sup>1,\*</sup> and Stephanie Glaser <sup>2</sup>

- <sup>1</sup> Institut f
  ür Volkswirtschaftslehre (520K), Computational Science Lab (CSL) Hohenheim, Universit
  ät Hohenheim, 70593 Stuttgart, Germany
- <sup>2</sup> Institut f
  ür Volkswirtschaftslehre (520K), Universit
  ät Hohenheim, 70593 Stuttgart, Germany; ssg.glaser@gmail.com
- \* Correspondence: robert.jung@uni-hohenheim.de

**Abstract:** This paper proposes a new spatial lag regression model which addresses global spatial autocorrelation arising from cross-sectional dependence between counts. Our approach offers an intuitive interpretation of the spatial correlation parameter as a measurement of the impact of neighbouring observations on the conditional expectation of the counts. It allows for flexible likelihood-based inference based on different distributional assumptions using standard numerical procedures. In addition, we advocate the use of data-coherent diagnostic tools in spatial count regression models. The application revisits a data set on the location choice of single unit start-up firms in the manufacturing industry in the US.

Keywords: count data models; spatial econometrics; spatial autocorrelation; firm location choice

# 1. Introduction

Our paper seeks to address global spatial autocorrelation arising form cross-sectional dependence between counts, a research area which constitutes a fundamental challenge in the spatial econometrics literature according to Billé and Arbia (2019). We propose a spatial autoregressive specification for the expected mean of counts at location *i* as a function of its *j* neighbouring region counts, an approach which is in the spirit of the class of observation-driven models of Cox (1981). It allows us to obtain a reduced form of the conditional mean function of the counts as a function of the so-called Leontief inverse or spatial multiplier as in the spatial autoregressive (SAR) model for continuous (dependent) variables.

We contrast our model with the well-received SAR-Poisson count model of Lambert et al. (2016), which also employs a spatial autoregressive structure for the expected mean of counts at location *i*, but relates this to its *j* neighbouring regions' Poisson log-intensities. We argue that this modelling of the spatial dependence between neighbouring regions is less intuitive in terms of interpretation and is not directly in the spirit of a continuous SAR model, because the intensity is an inherently unobserved measure while the counts are observable. Moreover, in the SAR-Poisson count model, the exogenous regressors have to explain the spatial dependence in the data fully, as there is no unexplained part in the intensity modelled explicitly.

Lambert et al. (2016) also derived a full-information maximum likelihood approach for parameter estimation, which, however, suffers from numerical difficulties mainly caused by the need to invert a transformation of the  $n \times n$  spatial weight matrix, where n denotes the sample size. As an alternative, the authors suggested a two-step estimation framework for their model, with an ad hoc specification to deal with the problem of computing the logarithm from zero counts<sup>1</sup>.

Our proposed spatial count regression model addresses all the issues raised and aims at supplying an alternative, which provides a useful interpretation of the model's parameters for the empirical economist and is straightforwardly to estimate using a likelihood-based approach with standard numerical procedures.



**Citation:** Jung, Robert C., and Stephanie Glaser. 2022. Modelling and Diagnostics of Spatially Autocorrelated Counts. *Econometrics* 10: 31. https://doi.org/10.3390/ econometrics10030031

Academic Editor: Claudio Morana

Received: 30 September 2021 Accepted: 19 August 2022 Published: 13 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Our contribution extends the literature on econometric models for spatially correlated count in two further directions: a more flexible functional form assumption to cope with the empirically highly relevant phenomenon of overdispersion, and the introduction of diagnostic tools and model validation methods. In non-spatial count data regression, the Poisson model is regularly not able to deal with unobserved heterogeneity in the data. One way to overcome this is to allow for a more flexible variance function as, e.g., in the negative binomial (NB) regression model. We propose to allow for such flexibility in spatial count data models as well and employ a negative binomial variant of our proposed model specification.

Our paper also advocates the routine use of diagnostic tools and model validation methods to assess the adequacy of a fitted model and to compare two or more competing model specifications. For this purpose we propose, i.a., a suitably adjusted variant of the probability integral transform (PIT) and scoring rules.

The reminder of paper is structured as follows: In Section 2, we briefly outline the current state of the literature on econometric models for spatially correlated counts and introduce our proposed alternative model specification. Section 3 contains goodness of fit analysis tools for spatial count data models. In Section 4, we report the results of a small Monte Carlo study to show the appropriateness of our estimation approach. In Section 5, we revisit the start-up births data set of Lambert et al. (2016) and apply our spatial count regression model to it. Section 6 concludes.

#### 2. Spatial Lag Models for Count Data

As already pointed out, the econometrics literature on models accounting for spatial autocorrelation in count data is quite sparse. The starting point of this literature is deemed to be the widely studied auto-Poisson model of Besag (1974)<sup>2</sup>. However, the model suffers from related problems to the exponential feedback models for time series models for counts<sup>3</sup>. The recent surveys of Glaser (2017) and Billé and Arbia (2019) list the few approaches available so far. Among them, the SAR-Poisson model of Lambert et al. (2016) features prominently, as it allows one to compute marginal effects which, due to the autoregressive nature of the model, can be decomposed in direct and indirect effects and allow for useful econometrics interpretations in many fields of applications.

To provide a starting point for the following discussion, the SAR-Poisson model specification for observed counts *y* is given here:

$$y_i | \mu_i \sim Po(\mu_i)$$
 with  $\mu_i = E[y_i | y_{-i}, x_i] = \exp\left(\rho \sum_{\substack{j=1 \ j \neq i}}^n w_{ij} \ln \mu_j + x_i \beta\right)$ , (1)

where  $w_{ij}$  is an element of the time invariant  $(n \times n)$  spatial weight matrix W, which is assumed to be row-standardised,  $\rho$  is the spatial autocorrelation parameter,  $x_i$  is the *i*-th row of the matrix of exogenous variables X, and  $\beta$  denotes the corresponding parameter vector. The symbol  $y_{-i}$  denotes the vector of counts for all neighbours of  $y_i$ . Note that Equation (11) given in Lambert et al. (2016) differs slightly from (1); after careful checking, we believe our result to be correct.

An attractive feature of the SAR-Poisson model is the availability of a reduced form. This can be seen by collecting the  $\mu_i$ 's in the vector  $\mu$  and employing a log transformation of (1) to obtain  $\ln \mu = (I - \rho W)^{-1} X \beta$ . Written in this reduced form makes it obvious that the SAR-Poisson way of introducing spatial dependence only allows for spatial dependence in the regressors, not in the unexplained part of the observations.

An explicit way to alter this is presented in Liesenfeld et al. (2016), where the SAR-Poisson model is extended with an additional error term, allowing also for spatial dependence in the unexplained part of the variation of the observed counts. However, while the SAR-Poisson model of Lambert et al. (2016) can be estimated via (limited information) maximum likelihood, the estimation of the model proposed by Liesenfeld et al. (2016) is not

straightforward, as the likelihood function is not available in closed form. Their proposal is then to use efficient importance sampling estimation which is not routinely available.

To overcome the shortcomings of the approaches discussed above, we propose a model specification that adjusts Besag's auto-Poisson model by introducing spatially lagged counts linearly in the conditional expectation function of a Poisson regression. This is in the spirit of the linear feedback model of Blundell et al. (2002), which has its origins in the integer valued autoregressive models for count time series (see, e.g., Jung and Tremayne 2011). Using the generic notation *D* for a (discrete) probability model, our spatial count regression model is given by

$$y_i | \mu_i \sim D(\mu_i)$$
 with  $\mu_i = E[y_i | y_{-i}, x_i] = \rho \sum_{\substack{j=1 \ i \neq i}}^n w_{ij} y_j + \exp(x_i \beta)$ . (2)

To insure the non-negativity of the conditional mean function in (2),  $\rho$  must fulfil a possibly more restrictive condition as compared to SAR models for continuous data. It can be shown that

$$\rho > -\frac{\exp(x_i\beta)}{w_i y}$$

must hold for all *i* and  $w_i$  denotes the *i*th row of the *W* matrix. The lower bound for  $\rho$  will therefore be negative, although it might be greater than -1. In our empirical example below, however, the bound is found to be far away from -1. Accordingly, a change in the observation of a single neighbour *j* of *i* causes a change in  $y_i$  by  $\rho w_{ij}$ . As argued in the Introduction, this interpretation of the spatial dependence between neighbouring regions is more intuitive as compared to that in the SAR-Poisson specification.

The proposal to use neighbouring counts instead of unobservable neighbouring intensities was also pursued in the recent paper by Apardian and Smirnov (2020), who, however, relied on the exponential feedback framework, the drawbacks of which have already been outlined above.

Several count data distributions can be assumed for *D* in (2), the most prominent being the Poisson. The resulting model will be denoted as the Poisson spatial count regression model below. One way to deal with unobserved heterogeneity in the observed counts is to use the negative binomial (NB) distribution. Employing an NB distribution in (2) leads to the NB spatial count regression model below. Many other extensions of the Poisson distribution which have been introduced in the literature—e.g., the zero-inflated Poisson and hurdle Poisson models—could be used in the modelling framework of (2). As these more flexible specifications have turned out not to substantially improve the model fit in our application below, they are not discussed further here.

In non-linear regression models such as the spatial count regression model (2), marginal effects are regularly employed to support the interpretation of the model parameters. In order to derive these, we restate the conditional mean of our model specification in (2) using the identity  $y = E[y|y_{-i}, X] + \epsilon$ , where  $\epsilon$  represents the (vector of) regression errors, as follows.

$$\mathsf{E}[y|y_{-i},X] = (I-\rho W)^{-1} \exp(X\beta) + (I-\rho W)^{-1} \rho W\epsilon .$$
(3)

Based on the form (3), it is straightforward to derive the total marginal effects for continuous regressors  $X_k$  for as follows.

$$A_i^{-1} \exp(X\beta)\beta_k$$

where  $A^{-1} = (I - \rho W)^{-1}$  and  $A_i^{-1}$  denotes the *i*th row of  $A^{-1}$ .

Following LeSage and Pace (2009), this can be decomposed into a direct and an indirect marginal effect. The former is given as

$$rac{\partial \mu_i}{\partial x_{ik}} = a_{ii}^{-1} \exp(x_i eta) eta_k$$
 ,

where  $a_{ii}^{-1}$  is the according element of the  $A^{-1}$ , and the latter as

$$\sum_{\substack{j=1\\j\neq i}}^n a_{ij}^{-1} \exp(x_j \beta) \beta_k \ .$$

Average marginal effects can be computed in various ways, e.g., by averaging over all observations *i*.

The specific form of the proposed conditional mean function in (2) does not allow for an easy to implement factorisation of the joint probability density function of the observed counts. Additionally, an operational high-dimensional multivariate count distribution is not available. Therefore, we propose a pseudo-likelihood approach for the estimation of the model parameters and marginal effects.

The idea to compose a pseudo-likelihood function of conditional probability functions such as the ones described in (2) stems from Besag (1975), who proposed this technique for the auto-normal schemes of Besag (1974) and sketched a proof of the estimator's consistency, relating it to the coding technique. Besag pointed 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).

Pseudo and composite likelihood functions have been either employed in the literature to obtain an approximation for an intractable likelihood function (see for example Jensen and Moller 1991) or to reduce the complexity of the maximisation of the likelihood function either by using the joint distribution of subsets of the data (i.e., 'subsetting methods') or by omitting redundant information (i.e., 'omission methods') in the likelihood (Varin and Vidoni 2005). Such a procedure was first proposed for spatial models by Vecchia (1998), building upon Besag (1974). In the case of a spatial model, one omits irrelevant information and conditions only on the relevant neighbours (in our case the eight nearest neighbours) instead of giving a full decomposition of the likelihood into conditional distributions. See also the description of the data generating process employed in the Monte Carlo experiments in Section 4 below. Finally note that a pseudo-likelihood procedure has also been proposed in the paper by Apardian and Smirnov (2020) in the context of their spatial count regression model.

In case of a Poisson distributional assumption in specification (2), the corresponding pseudo-log likelihood function takes the specific form

$$\log L_{\mathcal{C}} = \sum_{i=1}^{n} y_i \log \left( \rho \sum_{\substack{j=1\\j\neq i}}^{n} w_{ij} y_j + \exp(x_i \beta) \right) - \left( \rho \sum_{\substack{j=1\\j\neq i}}^{n} w_{ij} y_j + \exp(x_i \beta) \right) - \log(y_i!) .$$

For the NB model variant, we propose to use the Negbin 2 specification of the NB distribution which is popular in econometric applications due to is quadratic variance function.

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

where the dispersion parameter is denoted as  $\alpha$ . The resulting pseudo-log likelihood function for the NB spatial count regression model is then given as

$$\log L_{\mathcal{C}} = \sum_{i=1}^{n} \left( \sum_{j=1}^{y_i} \log(\alpha^{-1} + j - 1) \right) - \log(y_i!)$$
  
-(y\_i + \alpha^{-1}) log  $\left( 1 + \alpha \left( \rho \sum_{\substack{j=1\\j \neq i}}^{n} w_{ij} y_j + \exp(x_i \beta) \right) \right)$   
+y\_i log(\alpha) + y\_i log  $\left( \rho \sum_{\substack{j=1\\j \neq i}}^{n} w_{ij} y_j + \exp(x_i \beta) \right).$ 

The parameter estimates from the pseudo-likelihood for both models can straightforwardly be obtained using standard numerical optimisation methods. The properties of the estimators were investigated in the Monte Carlo study in Section 4. Robust standard errors were computed from the Godambe sandwich information matrix.

## 3. Diagnostics

In this section, we propose the application of several diagnostic tools adopted from the literature on models for non-spatially correlated counts. Although checking the adequacy of a specified model is an important part of any empirical (regression) modelling, diagnostic tools specially designed for count data are rarely considered in the spatial econometrics literature. In an attempt to bridge this gap, we employ the non-randomised probability integral transform histogram, scoring rules and a relative deviations plot.

# 3.1. Non-Randomised Probability Integral Transform

The probability integral transform (PIT) of Rosenblatt (1952) can be used to obtain an informal check for the model calibration. In the case of continuous random variables, the PIT is calculated as the values of the predictive cumulative distribution function (CDF) for the observations. If the 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 (see, e.g., Diebold et al. 1998).

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.

A popular solution for this problem is provided by the nonrandomised PIT, introduced by Czado et al. (2009). The nonrandomised PIT approach uses the predictive CDF for each observed count  $y_i$  to obtain the distribution of the PIT values directly as follows:

$$F(u|y_i) = \begin{cases} 0, & u \le P_{y_i-1}, \\ (u - P_{y_i-1})/(P_{y_i} - P_{y_i-1}), & P_{y_i-1} \le u \le P_{y_i}, \\ 1, & u \ge P_{y_i}, \end{cases}$$

where  $P_{y_i}$  is the predictive CDF of observation  $y_i$ , evaluated at  $y_i$ . The mean PIT is obtained by aggregating over all  $F(u|y_i)$  and dividing by n and compared to the cumulative distribution function of standard uniform random variable. A straightforward way to interpret PIT histogram can then be obtained by transforming the average  $F(u|y_i)$  into J equally spaced bins j = 1, ..., J and checking for uniformity. A u-shaped pattern of the histogram indicates overdispersion in the data (compared to the predictive distribution) and an inverse u-shaped underdispersion in the data.

# 3.2. 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, they negatively penalise a worse result, and are therefore to be minimised. We adopted this method of model selection from the work of Czado et al. (2009) and Jung et al. (2016), and average over the score of all observations for model selection. Several scores are suggested in the count data literature. We will present the logarithmic, quadratic and ranked probability scores, which place their emphases on different aspects of the estimated distributions.

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

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

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:

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

As a third scoring rule we use the ranked probability score, which especially penalises a flat estimated distribution:

$$rps = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{\infty} (F(j) - \mathbb{I}(y_i \le j))^2.$$

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 its best. This property holds for all three scoring rules presented here (Czado et al. 2009; Jung et al. 2016).

#### 3.3. Relative Deviations Plot

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

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

with

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

and  $\hat{\mu}_i$  being the estimate for the conditional expectation of observation  $y_i$  in the case of a Poisson distributional assumption.

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 of how well the predictive distributions display the different features of the data, including modus, tails, etc.

# 4. Monte Carlo Study

In this section, we assess the small sample properties of the parameter estimates obtained from the pseudo-log likelihood functions of our proposed spatial count regression model by means of a simulation study.

#### 4.1. Data Generating Process

Due to the fact that spatially lagged  $y_is$  are directly included into the conditional mean specification of our model and no natural ordering of the data as compared to a time series model is available, we propose to employ a Gibbs sampling approach to generate counts according to our conditional model specification (2) using Poisson and NB distributions, 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, ..., n, given all other  $y_js$ , we asymptotically obtain draws from the joint distribution of  $y_1, ..., y_n$ . E.g., in the k iteration of this procedure we draw from the following distributions:

$$y_1^{(k)} \sim D(y_1|y_2^{(k-1)}, \dots, y_n^{(k-1)})$$
  

$$y_2^{(k)} \sim D(y_2|y_1^{(k)}, y_3^{(k-1)}, \dots, y_n^{(k-1)})$$
  

$$\vdots$$
  

$$y_n^{(k)} \sim D(y_2|y_1^{(k)}, \dots, y_{n-1}^{(k)})$$

where *D* is defined as above. The starting values  $y_i^{(0)}$  were drawn from a non-spatial Poisson and negative binomial distribution, respectively, with  $\mu_i = exp(X_i\beta)$ . For the negative binomial case, the dispersion parameter  $\alpha$  was set to 0.2 in our experiments to obtain a modest level of overdispersion.

We employed an 8-nearest neighbours inverse distance matrix for the spatial weighting matrix. To calculate this matrix, we first generated *n* random coordinates using random numbers from a U(0, 1) distribution. Then, we computed the Euclidian inverse distance matrix for these points which represent the *n* spatial units of our simulated data set. After selecting the 8 nearest neighbours of each unit, we set all other entries of the matrix equal to zero, and finally row-standardised the resulting matrix. *X* consists of a constant,  $X_1 \sim U(0, 2)$ , and  $X_2 \sim N(1, 2)$ . The parameter vector  $\beta$  was set [0.5, 0.5, 0.5]', and data were generated for  $\rho^* = \{0, 0.2, 0.4, 0.6, 0.8\}$ , to cover a wide range of positive spatial dependence. Sample sizes employed were  $n = \{100, 250, 500, 1000, 5000, 10, 000\}$ ; in particular, a range of low sample sizes was used to check for potential finite sample estimation bias. In total, 10,000 replications for each simulation were used to obtain estimates of bias and root mean squared error (RMSE) with a high precision.

## 4.2. Monte Carlo Results

The results of our simulation experiments are presented in Tables 1 and 2. With a few exceptions, the root mean square errors and biases of the parameter estimates are in general small. For both the Poisson and the NB case, the finite sample bias of the parameter estimates decreased quickly, as the sample size was larger than 100. From simulation results not reported in the tables, it emerged that fitting the Poisson spatial count regression models to an even lower sample size than 100 is not recommended. To obtain precise and unbiased estimates of the regression intercept  $\beta_0$  in both the Poisson and the NB case is more difficult as compared to the slope coefficients  $\beta_1$  and  $\beta_2$ . A high level of spatial dependence (lager than 0.6) leads to larger biases and root mean squared errors for all parameter estimates employed. Unsurprisingly, the RMSE and bias results for the NB spatial count regression model are larger than those obtained for the Poisson variant. For overdispersed counts, the minimal sample size recommendations are even higher than in the equidispersed case.

$\rho \setminus n$	100	250	500	1000	5000	10,000
			ĥ	)		
RMSF			•			
0	0.0444	0.0159	0.0141	0.0090	0.0040	0.0023
0.2	0.0233	0.0221	0.0210	0.0146	0.0059	0.0041
0.4	0.0417	0.0350	0.0205	0.0136	0.0059	0.0044
0.6	0.0497	0.0291	0.0175	0.0139	0.0058	0.0038
0.8	0.0455	0.0224	0.0152	0.0082	0.0050	0.0043
BIAS						
0	-0.0034	-0.0007	-0.0006	-0.0002	0.0001	-0.0001
0.2	-0.0009	-0.0018	-0.0021	-0.0016	-0.0010	-0.0010
0.4	-0.0040	-0.0041	-0.0018	-0.0015	-0.0009	-0.0012
0.6	-0.0048	-0.0030	-0.0014	-0.0010	-0.0009	-0.0007
0.8	-0.0113	-0.0045	-0.0036	-0.0030	-0.0030	-0.0031
			β	0		
RMSE						
0	0.1493	0.0819	0.0624	0.0425	0.0188	0.0118
0.2	0.1174	0.0926	0.0778	0.0560	0.0239	0.0159
0.4	0.2378	0.1554	0.0842	0.0628	0.0281	0.0195
0.6	0.2951	0.1871	0.1042	0.0900	0.0351	0.0237
0.8	0.6341	0.2598	0.1649	0.0732	0.0447	0.0331
BIAS						
0	0.0019	0.0012	0.0001	0.0007	-0.0011	0.0001
0.2	-0.0007	0.0045	0.0051	0.0049	0.0040	0.0042
0.4	0.0058	0.0103	0.0056	0.0058	0.0047	0.0055
0.6	-0.0002	0.0083	0.0051	0.0057	0.0062	0.0054
0.8	-0.0104	0.0020	0.0011	-0.0011	0.0027	0.0032
			β	1		
RMSE						
0	0.0645	0.0422	0.0296	0.0206	0.0093	0.0062
0.2	0.0763	0.0426	0.0346	0.0238	0.0107	0.0071
0.4	0.0955	0.0569	0.0356	0.0269	0.0114	0.0084
0.6	0.1021	0.0704	0.0401	0.0326	0.0130	0.0096
0.0	0.1005	0.0044	0.0550	0.0500	0.0175	0.0125
BIAS	-0.0002	-0.0004	0.0002	-0.0001	0.0006	0.0001
02	0.0002	-0.0004	-0.0002	-0.0001	-0.0000	-0.0007
0.4	-0.0008	-0.0012	-0.0012	-0.0008	-0.0008	-0.0009
0.6	-0.0029	-0.0008	0.0001	-0.0013	-0.0016	-0.0006
0.8	0.0057	-0.0012	-0.0003	0.0005	-0.0008	-0.0008
			β	2		
RMSE						
0	0.0274	0.0135	0.0119	0.0075	0.0033	0.0021
0.2	0.0131	0.0145	0.0140	0.0096	0.0040	0.0027
0.4	0.0323	0.0263	0.0141	0.0099	0.0045	0.0029
0.6	0.0594	0.0257	0.0149	0.0136	0.0054	0.0037
0.8	0.1056	0.0401	0.0254	0.0090	0.0064	0.0046
BIAS						
0	-0.0006	-0.0003	-0.0001	-0.0001	0.0001	0.0001
0.2	-0.0003	-0.0007	-0.0011	-0.0008	-0.0006	-0.0007
0.4	-0.0013	-0.0020	-0.0009	-0.0011	-0.0008	-0.0008
0.6	0.0008	-0.0016	-0.0012	-0.0009	-0.0008	-0.0010
0.8	0.0023	-0.0008	-0.0005	-0.0001	-0.0005	-0.0004

**Table 1.** Simulation results for the Poisson spatial count regression model. BIAS is calculated as the average difference between estimates and true value. RMSE is calculated as the square root of the average squared difference between estimates and the true value.

$\rho \setminus n$	100	250	500	1000	5000	10,000
			ĥ	)		
RMSE						
0	0.0581	0.0352	0.0162	0.0102	0.0041	0.0031
0.2	0.0705	0.0438	0.0333	0.0229	0.0125	0.0102
0.4	0.0653	0.0580	0.0413	0.0276	0.0148	0.0108
0.6	0.1113	0.0652	0.0391	0.0291	0.0129	0.0103
0.8	0.0910	0.0539	0.0429	0.0305	0.0129	0.0100
BIAS						
0	-0.0101	-0.0064	-0.0009	-0.0008	0.0001	-0.0001
0.2	-0.0189	-0.0116	-0.0103	-0.0095	-0.0080	-0.0079
0.4	-0.0099	-0.0132	-0.0129	-0.0096	-0.0087	-0.0074
0.6	-0.0299	-0.0157	-0.0087	-0.0067	-0.0049	-0.0054
0.8	-0.0291	-0.0118	-0.0108	-0.0102	-0.0057	-0.0056
			β	60		
RMSE						
0	0.2771	0.1549	0.0990	0.0653	0.0300	0.0196
0.2	0.3330	0.1987	0.1498	0.1091	0.0594	0.0493
0.4	0.3564	0.2672	0.1868	0.1402	0.0864	0.0706
0.6	0.7309	0.4176	0.2675	0.1958	0.1069	0.0908
0.8	1.2300	0.6524	0.5012	0.3573	0.1637	0.1334
BIAS						
0	0.0142	0.0124	0.0003	0.0017	-0.0028	0.0016
0.2	0.0581	0.0419	0.0436	0.0460	0.0397	0.0395
0.4	0.0481	0.0562	0.0663	0.0617	0.0647	0.0590
0.6	0.0836	0.0804	0.0897	0.0761	0.0732	0.0726
0.8	0.0421	0.0439	0.0761	0.1181	0.0928	0.0951
			β	1		
RMSE						
0	0.1297	0.0744	0.0527	0.0373	0.0172	0.0113
0.2	0.1598	0.0935	0.0679	0.0470	0.0222	0.0168
0.4	0.1734	0.1092	0.0767	0.0560	0.0290	0.0207
0.6	0.2461	0.1619	0.1071	0.0761	0.0368	0.0282
0.8	0.6660	0.2615	0.1854	0.1267	0.0574	0.0428
BIAS						
0	-0.0037	-0.0024	-0.0002	-0.0003	0.0018	-0.0007
0.2	-0.0105	-0.0078	-0.0090	-0.0096	-0.0078	-0.0075
0.4	-0.0132	-0.0126	-0.0118	-0.0108	-0.0140	-0.0112
0.6	-0.0076	-0.0132	-0.0180	-0.0165	-0.0164	-0.0157
0.8	-0.0230	-0.0058	-0.0153	-0.0221	-0.0200	-0.0198
			β	32		
RMSE						
0	0.0572	0.0363	0.0202	0.0142	0.0061	0.0044
0.2	0.0601	0.0397	0.0311	0.0209	0.0118	0.0098
0.4	0.0628	0.0520	0.0367	0.0271	0.0163	0.0139
0.6	0.1147	0.0784	0.0436	0.0355	0.0195	0.0175
0.8	0.1742	0.1007	0.0779	0.0650	0.0278	0.0248
BIAS						
0	-0.0040	-0.0033	-0.0004	-0.0004	0.0002	-0.0003
0.2	-0.0129	-0.0095	-0.0094	-0.0091	-0.0079	-0.0079
0.4	-0.0107	-0.0131	-0.0142	-0.0133	-0.0127	-0.0121
0.6	-0.0224	-0.0198	-0.0171	-0.0156	-0.0147	-0.0146
0.8	-0.0221	-0.0167	-0.0166	-0.0250	-0.0183	-0.0198

**Table 2.** Simulation results for NB spatial count regression model. BIAS is calculated as the average difference between estimates and true value. RMSE is calculated as the square root of the average squared difference between estimates and true value.

$\rho \setminus n$	100	250	500	1000	5000	10,000
			C	K		
RMSE						
0	0.0554	0.0358	0.0243	0.0174	0.0075	0.0056
0.2	0.0503	0.0314	0.0225	0.0155	0.0077	0.0065
0.4	0.0424	0.0295	0.0213	0.0161	0.0098	0.0092
0.6	0.0406	0.0256	0.0195	0.0145	0.0095	0.0087
0.8	0.0353	0.0227	0.0165	0.0125	0.0079	0.0067
BIAS						
0	-0.0129	-0.0050	-0.0028	-0.0012	-0.0005	-0.0002
0.2	-0.0088	-0.0020	0.0004	0.0022	0.0036	0.0042
0.4	-0.0034	-0.0002	0.0056	0.0071	0.0075	0.0081
0.6	0.0035	0.0025	0.0082	0.0074	0.0078	0.0077
0.8	-0.0004	0.0040	0.0046	0.0058	0.0061	0.0056

Table 2. Cont.

In particular, the good performance of the parameter estimators for large sample sizes supports our conjecture that our pseudo-log likelihood function for the spatial count regression model can be seen as compatible with the form provided in Besag (1975). The estimations for small sample sizes showed satisfactory results as well. With regard to our empirical application, which will be presented in Section 5 below, the simulation showed that for our sample size of about 3000 observations, the estimation based on the conditional likelihoods works quite well.

# 5. Empirical Application

# 5.1. Data

For the application of the proposed spatial count regression models, we revisit the cross-sectional data of Lambert et al. (2016) on firm births in the manufacturing sector of the United States between 2000 and 2004. The data contain the number of start-up firms during this time period for 3078 U.S. counties and variables measuring the location factors, such as market structure, labour market and infrastructure for each county (see Table 3 and Lambert et al. (2016), p. 249, for more details). Figure 1 displays the spatial structure of the number of firm births. The frequency distribution of the number of start-up firms for values up to 100 is shown in Figure 2. It is evident that the counts do not exhibit excess zeros but an extraordinarily large range (from 0 to 6938), which is very difficult to capture for any single index model adequately.

We experimented with several forms of the weighting matrix *W*. All of these specifications produced very similar results. Therefore, only the results for a row-standardised inverse distance matrix considering only the eight nearest neighbours (which equals the weight matrix used by Lambert et al. 2016) are reported.



**Figure 1.** Observations for the subirth variable in dezils with dark blue representing the lowest values and dark red representing the highest ones.



**Figure 2.** Histogram of start-up firm birth counts in the manufacturing sector, U.S. counties, 2000–2004; x-axis is cut at 100.

Table 3.	Short	descri	ption	of the	variables.
----------	-------	--------	-------	--------	------------

Dependent variable	subirths	Single unit start-ups in the lower 48 United States during 2000–2004 in the manufactur- ing sector (NAICS 31-33)
Agglomeration economies	msemp tfdense pel10emp pem100emp	Manufactoring share of employment Total establishment density (in 100 s) Percent of manufacturing establishments with less than 10 employees Percent of manufacturing establishments with more than 100 employees
Market structure	mhhi pop cclass	Median household income (in 1000 s) Population (in 10,000 s) Share of workers in creative occupations
Labor availability and cost	uer pedas avg_wage netflow_emp	Unemployment rate Pecent of adults with an associate's degree Average wage per job (in 1000 s) Net flow of wages per commuter (in 1000 s)

Infrastucture	proad	Public road density
	interst	Interstate highway miles
	hwy_pc	Government expenditures on highways per capita (in 100 s)
	avland	Percent of farmland to total county
Fiscal policy	educ_pc	Government expenditures on education per capita (in 100 s)
	bci	State tax business climate index (higher values indicate more favorable business climates)
Area	metro	Dummy variable indentifying counties as be- longing to metropolitan areas
	micro	Dummy variable indentifying counties as be- longing to micropolitan areas

Table 3. Cont.

# 5.2. Results

Table 4 displays the estimation results for the Poisson and NB models discussed in Section 2, and for comparison, the non-spatial Poisson and non-spatial NB regression, whose conditional expectations are given by

$$E[y|X] = exp(X\beta) .$$

In the case of the non-spatial NB model, we chose again the Negbin2 (NB2) specification. In both spatial models, the spatial autocorrelation parameter  $\rho$  is highly significant, indicating that the proposed spatial structure exists in the data but is considerably different in size. Additionally, several of the other parameter estimates (and marginal effects, respectively) show clear differences between the Poisson and NB model specifications. Both the log score and the quadratic score clearly favour 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 penalises the relatively flat estimated distributions from the negative binomial models. All three scores give evidence in favour of the spatial model specification over their non-spatial counterpart.

To get a first visual impression of the fit of the models, Figure 3 shows the estimated conditional expectations of each county for the four models. The classes equal the deciles of the observations displayed in Figure 1 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 coasts and at the Great Lakes of the observations are recognisable, along with the cluster of lower values in the Midwest. The spatial models, however, predicted higher values for many counties, reproducing the map of the observations more properly. However, all models tended to estimate the outcomes of counties with observation 0 (dark blue in Figure 1) too high.

Next, the nonrandomised probability integral transform (PIT) is presented to evaluate the calibration of the models. Figure 4 displays the PIT histograms for the four estimated model specifications. The u-shaped PIT histograms of both Poisson models show clear signs of too little dispersion allowed by the model. The PIT histogram of the non-spatial negative binomial model has an even shape, but shows a decrease for PIT values larger than 0.8. This might be caused by the occurrence of large outliers in the data (>6000), and at the same time relatively high frequencies of small counts (<3). As already mentioned, single index models, as used here, generally have difficulties coping with such a data structure. The model having the PIT histogram closest to a uniform distribution, even though the difference to the non-spatial negative binomial is small, is the negative binomial spatial count regression model.

	Spat	ial	Non-Sp	oatial		Spatial N		Non-Sp	Non-Spatial	
Variable	Poisson	NB	Poisson	NB2		Poisson	NB	Poisson	NB2	
ρ	0.288 ***	0.166 ***			awage	0.033 ***	-0.058 ***	0.019 ***	-0.038 ***	
	(0.043)	(0.022)			Ū.	(0.008)	(0.012)	(0.007)	(0.007)	
const	-1.707 ***	-1.120 ***	-0.934 ***	-1.066 ***	netflow	0.003	-0.027	0.002	-0.016 ***	
	(0.397)	(0.249)	(0.281)	(0.195)		(0.003)	(0.006)	(0.820)	(0.003)	
msemp	0.035 ***	0.053 ***	0.031 ***	0.050 ***	proad	0.093 ***	$0.084^{***}$	0.103 ***	0.083 ***	
	(0.006)	(0.003)	(0.004)	(0.002)		(0.023)	(0.024)	(0.018)	(0.022)	
pelt10	-0.007 **	0.005 ***	-0.002	0.005 ***	interst	0.009 ***	0.005 ***	0.007 ***	0.005 ***	
	(0.003)	(0.001)	(0.002)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	
pemt100	-0.034 ***	-0.023 ***	-0.029 ***	-0.018 ***	avland	-0.007 ***	-0.006 ***	-0.009 ***	-0.007 ***	
	(0.006)	(0.003)	(0.004)	(0.002)		(0.002)	(0.001)	(0.001)	(0.001)	
tfdens	-0.013	-0.046 ***	0.006	-0.053 ***	bci	0.128 ***	0.032	0.080	0.034	
	(0.011)	(0.016)	(0.010)	(0.013)		(0.042)	(0.021)	(0.037)	(0.015)	
mhhi	-0.034 ***	0.024 **	0.000	0.027 ***	educpc	0.006 ***	0.006 **	0.004	0.004	
	(0.008)	(0.010)	(0.009)	(0.005)		(0.002)	(0.003)	(0.002)	(0.003)	
рор	0.002 ***	0.017 ***	0.002 ***	0.018 ***	hwypc	-0.039	-0.132	-0.030	-0.028	
	(0.000)	(0.003)	(0.000)	(0.003)		(0.023)	(0.031)	(0.019)	(0.021)	
cclass	0.088 ***	0.101 ***	0.048 ***	0.082 ***	metro	1.630 ***	$1.017^{***}$	1.265 ***	0.845 ***	
	(0.011)	(0.007)	(0.013)	(0.005)		(0.157)	(0.081)	(0.092)	(0.054)	
uer	0.037	0.076 ***	0.073 ***	0.080 ***	micro	0.839 ***	0.645 ***	0.573 ***	0.546 ***	
	(0.037)	(0.021)	(0.022)	(0.013)		(0.119)	(0.055)	(0.063)	(0.038)	
pedas	0.150 ***	0.062 ***	0.130 ***	0.044 ***	α		0.403 ***		0.437 ***	
	(0.022)	(0.011)	(0.021)	(0.009)			(0.026)		(0.024)	
Log L	-28,149	-10,300	-32,248	-10,401						
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	33.858						

**Table 4.** Estimation results from Poisson and the NB spatial count data regression and the non-spatial Poisson and NB2 regressions. N = 3078; robust standard errors in brackets. \*\* and \*\*\* denote a 5% and 1% significance, respectively.

A further way to look at the fit of the model is to predict the probabilities of each possible outcome and 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 5. Generally, the two negative binomial models outperform the Poisson models. Small counts are underestimated by the Poisson model; large counts are 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 Poisson spatial count regression model can be observed, whereas the non-spatial Poisson model fits the data less well.

For evaluating the impact of the regressors on the number of firm births in these nonlinear models, the marginal effects have to be considered. Table 5 displays the estimated median marginal effects for all four model specifications considered. We decided to look at the median rather than the mean, since our data set contains some unusually large observations of the dependent variable. As for that, the predicted effects for these regions distort the mean marginal effects. On the other hand, 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 neighbours could be considered.

As mentioned in Section 2, the total marginal effects in our spatial count regression model can be separated into the direct effect and indirect effect. The direct effect gives the impact of a change in the regressor  $x_{ik}$  on observation i 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 do not exist in a non-spatial model. Following (LeSage and Pace 2009, p. 39), the standard errors for the marginal effects were obtained using 2000 draws from the simulated joint distribution of the parameter estimators. As can be seen in Table 5, the majority of effects are 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.



(c) Non-spatial Poisson regression

(d) Non-spatial NB regression

**Figure 3.** Estimated conditional expectations. The colours indicate the deciles of the observations, with dark blue representing the lowest values and dark red representing the highest ones (see also Figure 1).





(a) Poisson spatial count regression

(b) NB spatial count regression



(c) Non-spatial Poisson regression (d) Non-spatial NB regression

Figure 4. Nonrandomised PIT histograms for the four regression models.



Figure 5. Relative deviations plot based on the four regression models employed.

When comparing the impacts between spatial and non-spatial models, we can see that the total marginal effects are in general of the same magnitude as the marginal effects of the corresponding non-spatial model. Therefore, the spatial models separate in some sense the effects in the non-spatial ones into direct and indirect and highlight the influence of neighbouring observations. Again, there are similar patterns for non-spatial and spatial versions of the Poisson and the negative binomial models.

**Table 5.** Median of marginal effects of the Poisson and NB spatial count regressions, and the non-spatial Poisson and non-spatial NB2 regressions. For the dummy variables *metro* and *micro*, the effect of a change from 0 to 1 is given. \*\* and \*\*\* denote 5% and 1% significance, respectively. Standard errors (in brackets) were estimated using their sample counterparts of 2000 draws of the asymptotic joint distribution of the coefficients.

	Poisson Spatial Reg.				NB Spatial Reg.	Poisson	NB2	
Variable	Total M.E.	Direct M.E.	Indirect M.E.	Total M.E.	Direct M.E.	Indirect M.E.	Direct M.E.	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)
рор	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)

Poisson Spatial Reg.					NB Spatial Reg.	Poisson	NB2	
Variable	Total M.E.	Direct M.E.	Indirect M.E.	Total M.E.	Direct M.E.	Indirect M.E.	Direct M.E.	Direct M.E.
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 5. Cont.

# 6. Conclusions

This article proposes a new regression model for spatially autocorrelated count data, denoted as the spatial count regression model. The model is observation-driven in the sense that neighbouring counts linearly affect the conditional expectation function of a Poisson or negative binomial regression model. This avoids the problems caused by exponential feedback specifications and provides a more intuitive interpretation of the spatial autocorrelation than most existing specifications. It also allows one to compute marginal effects, which can be decomposed into direct and indirect effects and allow for useful econometrics interpretations in many fields of application. The model can be estimated via pseudo-maximum likelihood using standard optimisation methods, making it easily adoptable in applied work. Finally, we advocate the usage of several model validation and diagnostics already standard in the regression analysis of non-spatial count data, i.e., PIT histograms, scoring rules and relative deviations plots.

In our empirical example, we revisited the start-up firm births data set of Lambert et al. (2016). With the help of the model validation devices and diagnostic tools, we illustrated the pros and cons of several model specifications. Among these, the best fit for the modelling of firm births in U.S. counties can be obtained with the negative binomial spatial count regression model. Nevertheless, this model considerably overestimated the smallest counts. A future improvement of the fit for this particular application may be obtained by moving away from single-index models, e.g., by employing a mixture of discrete distributions.

**Author Contributions:** All authors have contributed equally to all aspects of the preparation and writing of this paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Deutsche Forschungsgemeinschaft (DFG) under grant number JU 2276/3-1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The data used in this study are available upon request from the corresponding author.

Acknowledgments: We are grateful to anonymous referees, Roman Liesenfeld, Jan Vogler and the late Raymond Florax, for valuable and very constructive suggestions. For helpful comments, we also thank seminar and conference participants at the Universities of Hohenheim and Cologne and the World Conference of the Spatial Econometrics Association 2014 in Zurich, Switzerland.

Conflicts of Interest: The authors declare no conflict of interest.

# Notes

- Proença and Glórias (2021) also employed the SAR-Poisson specification (1), but proposed an alternative two-step Poisson pseudo-maximum likelihood approach to solve the problem of taking logarithms from zero counts. See also Simões et al. (2017), who employed the SAR-Poisson model of Lambert et al. (2016) to analyse the spatial correlation of calls to the Portugese National Healthline. The authors proposed a spatial lag Poisson Bayesian model to be estimated by a suitably augmented Nested Laplace Approximation (INLA) developed by Gómez-Rubio et al. (2015).
- <sup>2</sup> Besag's model has been used frequently to model spatial heterogeneity in context of a count data model (see, e.g., Gschlößl and Czado 2007; Gschlößl and Czado 2008; Apardian and Smirnov 2020).
- <sup>3</sup> See the very instructive discussion on this in Lambert et al. (2016).

# References

- Apardian, Rebecca E., and Oleg Smirnov. 2020. An analysis of pedestrian crashes using a spatial count data model. *Papers in Regional Science* 99: 1317–38. [CrossRef]
- Besag, Julian. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B* 36: 192–236. [CrossRef]
- Besag, Julian. 1975. Statistical Analysis of Non-lattice Data. The Statistician 24: 179–95. [CrossRef]
- Billé, Anna Gloria, and Giuseppe Arbia. 2019. Spatial limited dependent variable models: A review focused on specification, estimation, and health economics applications. *Journal of Economic Surveys* 33: 1531–54. [CrossRef]
- Blundell, Richard, Rachel Griffith, and Frank Windmeijer. 2002. Individual effects and dynamics in count data models. *Journal of Econometrics* 108: 113–31. [CrossRef]
- Cox, David R. 1981. Statistical analysis of time series: Some recent developements. Scandinavian Journal of Statistics 8: 93–115.
- Czado, Claudia, Tilmann Gneiting, and Leonhard Held. 2009. Predictive model assessment for count data. *Biometrics* 65: 1254–61. [CrossRef] [PubMed]
- Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay. 1998. Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review* 39: 863. [CrossRef]
- Geman, Stuart, and Donald Geman. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721–41. [CrossRef] [PubMed]
- Glaser, Stephanie. 2017. A review of spatial econometric models for count data. In *Hohenheim Discussion Papers in Business, Economics* and Social Sciences, No. 19-2017. Stuttgart: Universität Hohenheim.
- Gómez-Rubio, Virgilio, Roger Bivand, and Håvard Rue. 2015. A New Latent Class to Fit Spatial Econometrics Models with Integrated Nested Laplace Approximations. *Procedia Environmental Sciences* 27: 116–18. [CrossRef]
- Gschlößl, Susanne, and Claudia Czado. 2007. Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal* 2007: 202–25. [CrossRef]
- Gschlößl, Susanne, and Claudia Czado. 2008. Modelling count data with overdispersion and spatial effects. *Statistical Papers* 49: 531–52. [CrossRef]
- Jensen, Jens Ledet, and Jasper Moller. 1991. Pseudolikelihood for exponential family models of spatial point processes. Annals of Applied Probability 1: 445–61. [CrossRef]
- Jung, Robert C., Brendan P. M. McCabe, and Andrew R. Tremayne. 2016. Model validation and diagnostics. In *Handbook of Discrete Valued Time Series*. Edited by Richard A. Davis, Scott H. Holan, Robert Lund and Nalini Ravishanker. Boca Raton: Chapman and Hall, pp. 189–218.
- Jung, Robert C., and Andrew R. Tremayne. 2011. Useful models for time series of counts or simply wrong ones? AStA Advances in Statistical Analysis 95: 59–91. [CrossRef]
- Lambert, Dayton M., Jason P. Brown, and Raymond J. G. M. Florax. 2016. A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application. *Regional Science and Urban Economics* 40: 241–52. [CrossRef]
- LeSage, James P., and R. Kelley Pace. 2009. Introduction to Spatial Econometrics. New York: Chapman and Hall/CRC-Press.
- Liesenfeld, Roman, Jean-François Richard, and Jan Vogler. 2016. Likelihood Evaluation of High-Dimensional Spatial Latent Gaussian Models with Non–Gaussian Response Variables. In *Spatial Econometrics: Qualitative and Limited Dependent Variables*. Advances in Econometrics. Bingley: Emerald Group Publishing Limited, vol. 37, pp. 35–77.
- Long, J. Scott. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks: Sage Publications, Inc.
- Proença, Isabel, and Ludgero Glórias. 2021. Revisiting the spatial autoregressive exponential model for counts and other nonnegative variables, with application to the knowledge production function. *Sustainability* 13: 2843. [CrossRef]
- Rosenblatt, Murray. 1952. Remarks on a multivariate transformation. Annals of Mathematical Statistics 23: 470–72. [CrossRef]
- Simões, Paula, M. Lucília Carvalho, Sandro Aleixo, Sérgio Gomes, and Isabel Natário. 2017. A spatial econometric analysis of the calls to the Portugese national health line. *Econometrics* 5: 24. [CrossRef]
- Varin, B. Cristiano, and Paolo Vidoni. 2005. A note on composite likelihood inference and model selection. *Biometrika* 92: 519–28. [CrossRef]
- Vecchia, A. V. 1998. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society Series B* 50: 297–312. [CrossRef]