Richard Sellner and Manfred M. Fischer and Matthias Koch

A Spatial Autoregressive Poisson Gravity Model

Article (Draft)

Original Citation:
Sellner, Richard and Fischer, Manfred M. and Koch, Matthias (2013) A Spatial Autoregressive Poisson Gravity Model. *Geographical Analysis*, 45 (2). pp. 150-171. ISSN 1538-4632

A SAR Poisson gravity model

Richard Sellner[†, ‡], Manfred M. Fischer[†], and Matthias Koch[†]

[†]Institute for Economic Geography and GIScience, Vienna University of Economics and Business, Nordbergstrasse 15, A-1090 Vienna
[‡]Austrian Institute of Technology, Foresight & Policy Development Department, Donau-City-Strasse 1, A-1220 Vienna

February 20, 2012

Abstract

In this paper a Poisson gravity model is introduced that incorporates spatial dependence of the explained variable without relying on restrictive distributional assumptions of the underlying data generating process. The model comprises a spatially filtered component - including the origin, destination and origin-destination specific variables - and a spatial residual variable that captures origin- and destination-based spatial autocorrelation. We derive a 2-stage nonlinear least squares estimator that is heteroscedasticity-robust and, thus, controls for the problem of over- or underdispersion that is often present in the empirical analysis of discrete data. It can be shown that this estimator has desirable properties for different distributional assumptions, like the observed flows or (spatially) filtered component being either Poisson or Negative Binomial. In our spatial autoregressive model specification, the resulting parameter estimates can be interpreted as the implied total impact effects and, thus, include the indirect spatial feedback effects. Monte Carlo results indicate marginal biases in mean and standard deviation of the parameter estimates and convergence to the true parameter values in finite samples. Finally, patent citation flow data are used to illustrate the application of the model.

1 Introduction

Gravity models[1] represent a class of models that utilise origin-destination flow data to explain mean frequencies of interactions across space. Origin-destination flow data reflect (aggregate) interactions from a set of origin locations to a set of destination locations in some relevant geographic space. Such interactions may represent movements of various kinds. Examples include migration flows, journey-to-work flows, traffic and commodity flows, but also flows of information such as telephone calls or electronic messages and even the transmission of knowledge. Locations may be either area or point locations.

Gravity models typically rely on three types of factors to explain origin-destination flows: *first*, origin-specific variables that characterise the ability of origin locations to generate flows; *second*, destination-specific variables that represent the attraction of destination locations; and *third*, origin-destination functions that characterise the way interactions are impeded by separation (distance). In essence, gravity models assert a multiplicative relationship between mean interaction frequencies and the effects of origin, destination, and origin-destination variables[2], respectively (see Fischer and Wang 2011).

The gravity model has the advantage of simplicity but assumes independence of origin-destination flows. LeSage and Pace (2008) and Fischer and Griffith (2008) provide theoretical and empirical motivations that this may not be adequate, since flows might exhibit spatial dependence. In previous work, different approaches have been taken to account for the violation of the independence of flows assumption.

A simple way to overcome this weakness is to start with a log-additive version of the gravity model and allow for spatial dependence in the flow variable, as suggested in LeSage and Pace (2008). The major problem with this approach, especially when flows are a rare event, is the potential presence of zero flow magnitudes between origin-destination pairs; the so-called zero

[1]The term gravity model comes from the Newtonian origins of the models. In view of the enormous interest generated by gravity models, it is not surprising that a host of different theoretical approaches to these models have been proposed. For a discussion see Sen and Smith (1995) or Fischer and Reggiani (2004) among others.

[2]Origin-destination variables take the form of deterrence functions in some separation measure. At relatively large scales of geographical inquiry this might be simply the great circle distance separating an origin from a destination area (region), measured in terms of the distance between their respective centroids. In other cases, it might be transportation or travel time, cost of transportation, perceived travel time or any other sensible measure such as political distance, language distance or cultural distance measured in terms of nominal or categorical attributes.

flows problem. Mathematically, this would correspond to taking the logarithm of zero, which is not defined. In empirical applications, this problem can be avoided by adding an arbitrary constant to the observed zero flows. However, this might result in a downward bias in the parameter estimates for the model (see LeSage and Fischer 2010). Nevertheless, the log-additive version of the gravity model is widely used in practice.

Another way to overcome these deficiencies, is by assuming that the origin-destination flows are independently distributed Poisson variates and introducing two n-by-1 vectors of spatially structured regional effect parameters. These parameters contain one effect for each region treated as an origin and another for each region treated as a destination. These assumptions lead to the Bayesian hierarchical Poisson model suggested by LeSage et al (2007).

In contrast, Fischer and Griffith (2008) suggest to incorporate spatial dependence in the disturbance process, as in the case of serial correlations in time series regression models. They applied Griffith's spatial filtering methodology to deal with the issue of spatial dependence in Poisson gravity models[3]. Within this semi-parametric approach, synthetic variables are introduced to control for spatial dependence in the error term, arising from missing origin and destination variables that are spatially autocorrelated. These surrogate variables are constructed from the eigenvectors of a modified version of the spatial weight matrix (Griffith 2003).

In this paper, we propose a generalisation of the Poisson gravity model[4] as an alternative approach to account for spatial dependence in flows. Thus, we avoid the logarithmic transformation of the dependent variable and, hence, the zero flows problem. Spatial autocorrelation in the dependent flow variable is introduced by the origin- and destination-based dependencies as suggested by LeSage and Pace (2008). Consequently, unlike models that incorporate the spatial dependence in the error term, our model implies direct, indirect and total effects which are necessary for a proper model interpretation (as motivated by LeSage and Pace 2009, 33pp).

The remainder of the paper is organised as follows. In Section 2 we describe the standard Poisson gravity model along with its estimation approach. Section 3 then adds spatial lags of the

[3] The term *Poisson gravity model* (see Flowerdew and Aitkin 1982; Bailey and Gatrell 1995) might be misleading, since it does not assume the dependent variable to be Poisson distributed. Note that in the model of Fischer and Griffith (2008) the equidisperion property (i.e. mean equals variance) does not hold if the spatial autocorrelation parameter is different from zero.

[4] Note that in this paper we consider only the unconstrained gravity model version. See Davies and Guy (1987) for singly and doubly constrained versions corresponding to the family of gravity models identified in Wilson (1971).

dependent origin-destination flow variable to the Poisson gravity model. This is done by modelling the flow variable as the sum of a spatially filtered variable with a (specific) distribution and a residual spatial variable. The general structure for a gravity model including origin-based and destination-based spatial dependence is shown and the consequences regarding the dispersion properties of the data are discussed for a spatially autocorrelated Poisson variable. In Section 4, we first discuss the interpretation of the parameter estimates of the model. As a result of the model specification, the parameter estimates can be interpreted as total effects. The section continues with the derivation of a 2-stage nonlinear least squares estimator for the model. It is shown that the estimator (and the interpretation of the model) does not depend on the strict distributional assumptions of the model. In Section 5 a Monte Carlo experiment is conducted to show the performance of the estimator. The bias and standard deviation of the parameter estimates are computed for different sample sizes, spatial weight matrices and parameter values. Results of the Monte Carlo experiment indicate that the estimator has desirable properties regarding the bias, independent of the distributional assumptions of the model. Section 6 applies estimator to an empirical data set on patent citation flows in Europe. The model results are compared to those from conventional (non-spatial) Poisson and Negative Binomial model specifications, estimated by heteroscedasticity-robust Maximum Likelihood. A final section concludes.

2 A Poisson gravity model

Let us assume a spatial system consisting of n regions, and let $y = (y_1, \ldots, y_N)$ denote a sample of $N = n^2$ origin-destination pairs of regions. The n-by-n flow matrix that contains intraregional flows on the main diagonal and interregional flows as off-diagonal elements is vectorised by stacking the columns to form the N-by-1 vector of flows contained in y. We use $i = 1, \ldots, N$ to index observations, each of which represents an origin-destination pair.

In a Poisson gravity model, it is assumed that y_i is independently distributed and that the distribution of y_i is a Poisson distribution

$$y_i \sim \mathcal{P}(\mu_i) \tag{1}$$

whose mean parameter μ_i is given by

$$\mu_i = \exp(z_i\beta) = \exp\left(\sum_{k=1}^{K} z_{i,k}\beta_k\right) \tag{2}$$

where $z_i = (z_{i,1}, \ldots, z_{i,K})$ represents a vectors of (logged) origin-specific, destination-specific and origin-destination variables[5] associated with the i-th observation, and $\beta = (\beta_1, \ldots, \beta_K)'$ is a K-by-1 ($k = 1, \ldots, K$) parameter vector. The exponential function appearing in μ_i is justified by the positivity of μ_i, since a linear function $\mu_i = z_i\beta$ would imply possibly incompatible constraints on the parameters. The conditional mean and the conditional variance of y_i given z_i are equal to μ_i, and the density function of y_i is

$$\frac{\exp(-\mu_i)(\mu_i)^{y_i}}{y_i!}. \tag{3}$$

Given independent observations y_i, the standard estimator for this Poisson gravity model is the maximum likelihood estimator (MLE). The log-likelihood function then takes the following form

$$\mathcal{L}(\beta) = \sum_{i=1}^{N} \left\{ y_i \log(\exp(z_i\beta)) - \exp(z_i\beta)) - \ln y_i! \right\}, \tag{4}$$

which has to be optimised numerically.

In the Poisson gravity model y_i has mean $\mu_i = exp(z_i\beta)$ and variance[6] μ_i. Because flow data almost always reject the restriction that the variance equals the mean, Fischer et al (2006) suggest a heterogeneous Poisson gravity model. This model specification arises from introducing multiplicative heterogeneity in the mean of the Poisson model as a proxy for fixed effects parameters. The heterogeneity term is strategically assumed to follow a conjugate gamma distribution. This choice of a Poisson-gamma mixture is strategic in the sense that the conjugate gamma distribution yields a tractable Negative Binomial distribution maximum likelihood procedure (see

[5]Note that we consider a log-additive gravity model with a power deterrence function. Thus, the mean parameter μ_i is logarithmically linked to a linear combination of the logged origin-specific and destination-specific characteristics and the logged distances between origins and destinations. Then the coefficient estimates reflect elasticity responses of origin-destination flows to the various origin-, destination, and origin-destination characteristics.

[6]The restriction that the variance equals the mean in a Poisson regression framework is usually called equidispersion (see Cameron and Trivedi 1998, p4).

Cameron and Trivedi 1998).

In matrix notation the Poisson gravity model can be expressed as

$$y \sim \mathcal{P}(\mu) \tag{5}$$

$$\mu = exp(Z\beta) \tag{6}$$

where Z is a N-by-K matrix of origin, destination and origin-destination characteristics and a constant term, with the associated K-by-1 vector of regression parameters[7] β. Hence, we may write

$$Z\beta = \alpha\iota_N + X_d\gamma_d + X_o\gamma_o + \delta D \tag{7}$$

with

$$X_d = \iota_n \otimes X_1 \tag{8}$$

$$X_o = X_2 \otimes \iota_n \tag{9}$$

where X_1 is an n-by-K_1 matrix of (logged) destination-specific characteristics and X_2 an n-by-K_2 matrix of (logged) origin-specific characteristics in the n regions, ι_n is an n-by-1 unity-vector, $\otimes$ represents the Kronecker product, and γ_d and γ_o are K_1-by-1 and K_2-by-1 parameter vectors associated with the destination-specific and origin-specific characteristic of the regions, respectively. The K_3-by-1 parameter vector δ reflects the effect of the N-by-K_3 matrix of (logged) spatial separation variables (D) between each origin-destination pair. The parameter α denotes the constant term parameter[8].

[7] Note that $Z = (\iota_N\ X_d\ X_o\ D)$ and $\beta = (\alpha\ \gamma_d\ \gamma_o\ \delta)'$.

[8] Thus, the total number of parameters of the model is $K = K_1 + K_2 + K_3 + 1$.

3 The SAR Poisson gravity model

3.1 Introducing spatial autocorrelation in the Poisson gravity model

In gravity models, observations are usually assumed to be independent from each other. This is a heroic assumption – as LeSage and Pace (2009, p. 212) point out – since origin-destination flows are fundamentally spatial in nature. Models of a type given by Eqs. (5) and (6) neglect spatial dependencies of the origin-destination flows contained in the dependent variable y, hereafter termed *spatial autocorrelation*. For example, neighbouring origins and destinations may exhibit estimation errors of similar magnitude when underlying a spatially dependent variable, or missing variables reflecting origin and destination characteristics may exert a similar impact on neighbouring observations.

Following LeSage and Pace (2008), we define spatial dependence to mean that observed flows from an origin region r to a destination region j are either negatively or positively correlated with: (i) flows from regions nearby the origin r to the destination j, say regions r' and r'' that are neighbours to region r, which they label origin-dependence; and (ii) flows from origin region r to regions neighbouring the destination region j, say regions j' and j'', which they label destination-dependence[9].

In our model the original variable, say y, representing origin-destination flows, is given by a spatially filtered variable y^* and a residual spatial variable $\tilde{y}$. The latter is assumed to be given by

$$\tilde{y} = \rho_o W_o y + \rho_d W_d y, \tag{10}$$

$$W_o = W \otimes I_n, \tag{11}$$

$$W_d = I_n \otimes W, \tag{12}$$

where W is an n-by-n spatial weight matrix with diagonal elements set to zero and I_n is an n-by-

[9]We do not consider the third type of spatial dependence in flows of LeSage and Pace (2008) which would be present if observed flows from an origin region r to a destination region j are negatively or positively correlated with flows from regions neighbouring the origin region r to regions neighbouring the destination region j, say flows from regions r' and r'' to regions j' and j'', which they label origin-to-destination dependence.

n identity matrix. The spatial weight matrix is typically normalised to have row sums of unity[10], and thus produces linear combinations of flows from neighbouring regions (see LeSage and Pace 2009, p. 10). Non-zero entries in matrix W indicate that a neighbourhood relation exists between the corresponding regions. Neighbours may be defined using contiguity or measures of spatial proximity such as cardinal distance (measured, for example, in terms of travel time) and ordinal distance (for example, the five closest neighbours). Given an origin-centric organisation of the sample data, the spatial weight matrix W_o is used to form an N-by-1 spatial lag vector $W_o y$ that captures origin-based dependence arising from flows that neighbour the origin, similarly, a spatial lag of y is formed using the spatial weight matrix W_d that captures destination-based dependence using a linear combination of flows associated with observations that neighbour the destination region. ρ_o and ρ_d are the corresponding scalar spatial lag parameters (LeSage and Fischer 2010).

As a result, the spatial autoregressive version of the Poisson gravity model (in the following referred to as SPGM) takes the form:

$$\begin{aligned} y &= \tilde{y} + y^* = \rho_o W_o y + \rho_d W_d y + y^*, && (13) \\ E[y^*] &= \mu = \exp(Z\beta), && (14) \\ y^* &\sim \mathcal{P}(\mu). && (15) \end{aligned}$$

Note that if $\rho_o = \rho_d = 0$ the SPGM collapses to the conventional Poisson gravity model:

$$y = y^*, \quad E[y^*] = \mu = \exp(Z\beta) \text{ and } y^* \sim \mathcal{P}(\mu). \qquad (16)$$

As will be shown later, the interpretation and estimation of the model given by Eqs. (13) - (15) does not depend on the strict assumption given in Eq. (15), concerning the distribution of y^* being Poisson and, hence, the model would be valid for a more general model class. However, for the following discussion, we hold on to the assumption of $y^* \sim \mathcal{P}(\mu)$ and examine the resulting dispersion properties of a spatially autocorrelated Poisson distributed variable.

[10] From an econometric perspective, other normalisation procedures, such as the maximum absolute eigenvalue normalisation, are also possible.

3.2 Dispersion properties of spatially autocorrelated Poisson distributed variables

Adding spatial lags to the Poisson gravity model results in interesting dispersion properties for y, given the underlying data generating process is $y^* \sim \mathcal{P}(\mu)$. These properties can be derived from the first two central moments of y that are given by[11]

$$E[y] = (I_N - \rho_d W_d - \rho_o W_o)^{-1}\mu, \tag{17}$$

$$Var[y] = (I_N - \rho_d W_d - \rho_o W_o)^{-1}\mathrm{diag}[\mu](I_N - \rho_d W_d' - \rho_o W_o')^{-1}, \tag{18}$$

where the operator diag$[\cdot]$, transforms the vector μ into a diagonal matrix and I_N is an N-by-N identity matrix. From Eqs. (17) and (18) it can be seen that if $\rho_d \neq 0$ and $\rho_o \neq 0$ then $E[y] \neq Var[y]$ which then violates the equidispersion property of the Poisson distribution (see Cameron and Trivedi 1998, p. 4). Since each entry in the vectors $E[y]$ and diag$[Var[y]]$ may differ, an analytic comparison between Eq. (17) and Eq. (18) with respect to over- or underdispersion[12] is difficult. However, to give some insights regarding the dispersion dynamics caused by spatial autocorrelation in Poisson distributed variables, consider the following simple example. Assume that $\mu = \exp(Z\beta) = \iota_N$ and a spatial weight matrix W that reflects a simple first-order contiguity neighbourhood structure. As W is row-normalised, $E[y^*] = \iota_N$ and Eq. (17) simplifies to $E[y] = \frac{1}{1-\rho_d-\rho_o}\iota_N$. For a sample size of $n = 25$ (and thus $N = 625$) we simulate the first two moments given by Eqs. (17) and (18) over a grid of ρ_d and ρ_o ranging from +0.4 to -0.4 in steps of 0.05. As a measure of the deviation between mean and variance we construct τ, which denotes the average percentage distance between the mean and variance of y:

$$\tau = \varnothing(\mathrm{diag}[\mu]^{-1}(\mathrm{diag}[Var[y]] - E[y])), \tag{19}$$

whereas the operator $\varnothing[\cdot]$ takes the average of a vector. If $\tau > 0$ ($\tau < 0$), y will show tendencies towards overdispersion (underdispersion).

[11] It is assumed that the inverse of $(I_N - \rho_d W_d' - \rho_o W_o')$ exists.
[12] The terms overdispersion and underdispersion refer to $E[y] < Var[y]$ and $E[y] > Var[y]$, respectively.

Figure 1 depicts τ on the vertical axis and ρ_o and ρ_d on the horizontal axes. To test the statistical significance of over- or underdispersion, we use the test statistic suggested by Cameron and Trivedi (1998, 77p). This statistic corresponds to a simple OLS regression of the form

$$\begin{aligned} \widetilde{y} &= \eta\widehat{y} + \varepsilon, \qquad (20) \\ \widehat{y} &= (I_N - \rho_d W_d - \rho_o W_o)^{-1}\mu \\ \widetilde{y} &= \mathrm{diag}[\widehat{y}]^{-1}\left((y-\widehat{y})\odot(y-\widehat{y})\right), \end{aligned}$$

with $\widehat{y}$ being the fitted values of y, $\widetilde{y}$ is a measure of deviation between mean and variance and ε being a standard normal error term. The sign $\odot$ denotes the Hadamard product that refers to an elementwise multiplication of two vectors or matrices. The t-statistic of the resulting coefficient η is asymptotically normal distributed under the null hypothesis of equidispersion against the alternative hypothesis of over- or underdispersion.

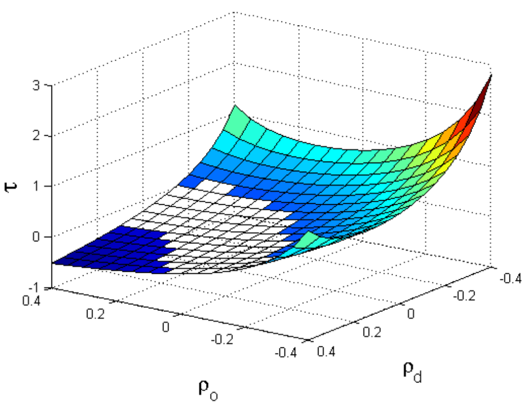

Figure 1: Simulated mean percentage distance between mean and variance of y
Coloured (transparent) cells indicate significant (insignificant) $sign(\tau)$ with 95 percent accuracy.

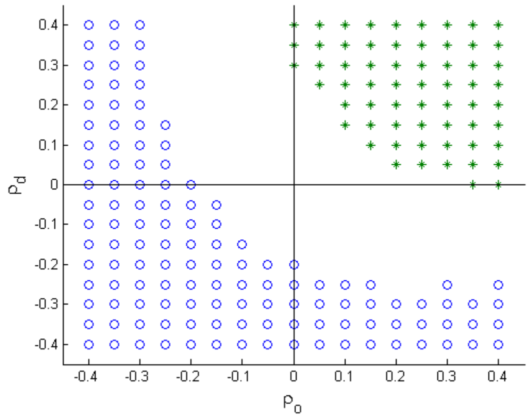

Figure 2: Significant over- and underdispersion for different spatial lag parameter values
Symbols: o indicates significant overdispersion and * significant underdispersion.

We simulated 100 replications of $y^* \sim \mathcal{P}(\mu_0)$ and derived the respective $y = (I_N - \rho_d W_d - \rho_o W_o)^{-1} y^*$ for each combination of ρ_o and ρ_d. If the corresponding p-value is less than 0.05 in more than or equal to 95 of the repetitions, we treat over- or underdispersion as being statistically significant. Such a case is indicated by shaded entries in Figure 1. For $\rho_o, \rho_d < 0$ we find statistically significant and substantial overdispersion. The opposite case $\rho_o, \rho_d > 0$ leads to significant underdispersion, however, to a somewhat smaller extent. Values around zero for both ρ_o, ρ_d do not lead to significant dispersion patterns. Mixtures of positive and negative spatial autocorrelation parameters translate to significant overdispersion for values of approximately $>$ $|0.2|$. For a better depiction of this pattern consider Figure 2, highlighting spatial lag parameters that result in significant over- or underdispersion. As can be seen more clearly in this figure, for mixed signs in spatial autocorrelation parameters $\rho_o < 0, \rho_d > 0$ or $\rho_o > 0, \rho_d < 0$, significant overdispersion is present for negative spatial autocorrelation parameters smaller than -0.20.

4 Model interpretation and estimation

4.1 Model interpretation

As the SPGM belongs to the class of spatial autoregressive regression models, the total effect of an explanatory variable on the dependent variable, has to include the indirect effects arising from the spatial feedback effects (see LeSage and Pace 2009, 33pp)[13]. Like LeSage and Pace (2009), we suggest *total impact effects* for the interpretation of the coefficients. We define the total average effects in a sense that they reflect the total average elasticity of the explanatory variables on the expected value of y. By construction[14], the elasticity of $E[y_i]$ to z is given by $\chi_{k,q,i}$:

$$\chi_{k,q,i} \quad := \frac{\partial \log\left(E\left[y_i\right]\right)}{\partial z_{q,k}}, \tag{21}$$

whereas $z_{k,q}$ denotes the k^{th} variable (or the k^{th} column of Z) and $q = 1, ..., N$ denotes the origin-destination pair (or row of Z). Hence, we can define the total impact effects of the k^{th} variable on y by

[13] Similar to linear models, the interpretation of the model parameters does not depend on the underlying distributional assumptions of the model. Therefore, we outline the interpretation of the model for a more general model class, covering all models with strictly positive mean realisations.

[14] Note that the explanatory variable matrix Z is already logged and therefore Eq. (21) represents the elasticities and not semi-elasticities. Additionally, μ is always strictly positive, by construction, and therefore $\log(E[y])$ is well defined.

$$Total_k = \frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{q=1}^{N}\chi_{k,q,i}. \tag{22}$$

As the explanatory variables in Z are in logarithms, the resulting total effects can be interpreted as elasticities. Rewriting $E\left[y_i\right]$ in an elementwise fashion yields:

$$E\left[y_i\right] = \sum_{q=1}^{N} s_{i,q}\exp\left(\sum_{k=1}^{K} z_{q,k}\beta_{0,k}\right), \tag{23}$$

where $s_{i,q} \in S$ and $S := \left(I_N - \rho_{d,0}W_d - \rho_{o,0}W_o\right)^{-1}$. True parameters of the data generating process in Eq. (23) are indexed with a zero: $\rho_{o,0}$, $\rho_{d,0}$ and β_0, respectively. As Eqs. (24) and (25) show, the total impact effects do not depend on $\rho_{d,0}, \rho_{o,0}$ and are equal to $\beta_{0,k}$. Thus, in contrast to spatial autoregressive models that are linear in the spatial autocorrelation coefficient, the total impact effects in the SPGM need no additional treatment since they only depend on the parameter vector β.

$$\frac{\partial \log E\left[y_i\right]}{\partial z_{q,k}} = \frac{s_{i,q}\exp\left(\sum_{\kappa=1}^{K} z_{q,\kappa}\beta_{0,\kappa}\right)\beta_{0,k}}{\sum_{u=1}^{N} s_{i,u}\exp\left(\sum_{\kappa=1}^{K} z_{u,\kappa}\beta_{0,\kappa}\right)} \tag{24}$$

$$\begin{aligned}\Rightarrow Total_k = \frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{q=1}^{N}\frac{s_{i,q}\exp\left(\sum_{\kappa=1}^{K} z_{q,\kappa}\beta_{0,\kappa}\right)\beta_{0,k}}{\sum_{u=1}^{N} s_{i,u}\exp\left(\sum_{\kappa=1}^{K} z_{u,\kappa}\beta_{0,\kappa}\right)} = \\ \frac{1}{N}\sum\nolimits_{i=1}^{N}\frac{\sum_{q=1}^{N} s_{i,q}\exp\left(\sum_{\kappa=1}^{K} z_{q,\kappa}\beta_{0,\kappa}\right)\beta_{0,k}}{\sum_{u=1}^{N} s_{i,u}\exp\left(\sum_{\kappa=1}^{K} z_{u,\kappa}\beta_{0,\kappa}\right)} = \beta_{0,k}.\end{aligned} \tag{25}$$

For the interested reader, the direct and indirect elasticities are outlined in Eqs. (26) and (27), respectively.

$$Direct_k := \frac{1}{N}\sum\nolimits_{i=1}^{N}\frac{\partial \log\left(E\left[y_i\right]\right)}{\partial x_{i,k}} = \frac{1}{N}\sum\nolimits_{i=1}^{N}\frac{s_{i,i}\exp\left(\sum_{\kappa=1}^{K} x_{i,\kappa}\beta_{0,\kappa}\right)\beta_{0,k}}{\sum_{j=1}^{N} s_{i,j}\exp\left(\sum_{\kappa=1}^{K} x_{j,\kappa}\beta_{0,\kappa}\right)} \tag{26}$$

$$Indirect_k := \frac{1}{N}\sum\nolimits_{i=1}^{N}\sum\nolimits_{q=1,i\neq q}^{N}\frac{\partial \log\left(E\left[y_i\right]\right)}{\partial x_{q,k}} = Total_k - Direct_k \tag{27}$$

It can be seen that the direct and indirect effects in this model still depend on $\rho_{d,0}$ and $\rho_{o,0}$ and therefore their standard deviation would have to be simulated. To derive an efficient simulation algorithm Eq. (26) should be rewritten in matrix form, given in Eq. (28)

$$Direct_k = \frac{\gamma_{0,k}}{N}\iota_N'(diag(S) \oslash \exp(X\gamma_0)) \odot S\exp(X\gamma_0), \tag{28}$$

where $\oslash$ denotes elementwise division of matrices or vectors. Note that in order to simulate Eq. (28) one has to calculate the elements $s_{i,i}$ which poses a similar computational problem as in spatial autoregressive models that are linear in the spatial autocorrelation coefficient. Since our model has two spatial weight matrices, the algorithms for the computation of the log-determinants given in LeSage and Pace (2009, Chapter 4) would have to be adopted accordingly.

4.2 A two-stage nonlinear least squares estimator for the model

As already mentioned before, the standard approach for estimating a Poisson regression model is to derive the likelihood function and apply Maximum Likelihood (ML) estimation methods (see Cameron and Trivedi 1998, 22pp). However, unlike the multivariate normal distribution for linear models, no analytically closed form for a multivariate Poisson distribution[15] has been derived so far (to our knowledge). Therefore estimation techniques like ML or Bayesian methods are infeasible. Hence, the estimator introduced in this paper builds upon the non-linear least squares (NLS) estimator for Poisson distributions as in Cameron and Trivedi (2005, 150pp).

To generalise the estimation results with respect to different distributional assumptions of the underlying data generating process (DGP), we assume that the spatially filtered variable y^* can have any distribution $C(\mu, \Omega)$ with a specified (finite) mean μ and an unspecified (finite) diagonal variance-matrix Ω with the diagonal entries: $(\sigma^2_{0,1,1}, \sigma^2_{0,1,2}, ..., \sigma^2_{0,n,1}, \sigma^2_{0,2,1}, ..., \sigma^2_{0,n,n})$ and off diagonals of zeros. Hence, our SPGM estimator can be considered heteroscedasticity-robust (i.e. robust against over- or underdispersion). Note that if $\rho_{d,0} = \rho_{o,0} = 0$ the DGP

[15]The likelihood for a multivariate Poisson distributed variable in our case is given by $\sum_{y^* \in M} \prod_{i=1}^{N} \frac{\exp(-\mu_i)\mu_i{}^{y_i^*}}{y_i^*!}$, with $M = (I_N - \rho_{d,0}W_d - \rho_{o,0}W_o)\, y$. In order to calculate the likelihood, recursive algorithms are needed as for example given in Karlis and Meligkotsidou (2005). However, these algorithms are much more computationally time consuming than the approach we suggest in this paper.

collapses to the one described in Cameron and Trivedi (2005, 150pp), which can be estimated by NLS used for Poisson distributed random variables. As suggested by Cameron and Trivedi (2005), we employ an NLS estimation framework to find estimates for $\rho_{d,0}$, $\rho_{o,0}$ and β_0.

Assume that $\rho_{d,0}$ and $\rho_{o,0}$ have values such that the maximum absolute eigenvalue of F is smaller than one[16], where $F := \rho_{d,0}W_d + \rho_{o,0}W_o$. It follows that the inverse of $I_N - F$ exists (for details see Kelejian and Prucha 1998). Hence, Eq. (13) can be solved for y to derive Eq. (29):

$$y = (I_N - \rho_{d,0}W_d - \rho_{o,0}W_o)^{-1}\, y^* \text{ where } y^* \sim C(\mu, \Omega)\,. \tag{29}$$

Since W_d and W_o represent spatial weight matrices, their entries can be treated as fixed weights. Hence, the mean of y is given by $E[y] = (I_N - \rho_{d,0}W_d - \rho_{o,0}W_o)^{-1} \exp(Z\beta_0)$. Note that the DGP in Eq. (29) is more general as the one given in Eq. (13), since y^* can be drawn from a variety of distributions. Given that y represents count data flows, three particular distributions are of importance: First y^* could be drawn from a Poisson distribution, hence $E[y^*] = \mu$ and $Var[y^*] = \mu$. Therefore, if $\rho_{d,0} = \rho_{o,0} = 0$ then y is Poisson distributed as well. Second y^* could be drawn from a Negative Binomial distribution, hence the expected value is unchanged with $E[y^*] = \mu$ and the variance is quadratic in mean $Var[y^*] = \mu + \lambda(\mu \odot \mu)$ with dispersion parameter λ. Therefore, if $\rho_{d,0} = \rho_{o,0} = 0$ then y is Negative Binomial distributed as well. Note that in both cases if $\rho_{d,0} \neq 0$, $\rho_{o,0} \neq 0$ the random variable y is no longer Poisson or Negative Binomial. Finally, y^* could be drawn from a distribution such that the random variable y is Poisson or Negative Binomial distributed. Given the case that y is Poisson, the DGP given in Eq. (29) would not account for possible overdispersion in y, since $E[y] = Var[y] = (I_N - \rho_{d,0}W_d - \rho_{o,0}W_o)^{-1}\mu$. However, independent of the distributional assumption, each DGP results in the same first moment of y. Since a heteroscedasticity-robust NLS estimator requires only a correctly specified first moment (mean), our estimation procedure will result in the same estimates independent of the underlying distribution of y^* or y and, hence, is robust against all possible misspecification in the distribution.

Applying NLS estimation methods to the DGP given in Eq. (29) yields an estimator for $\delta_0 := (\rho_{d,0}\ \rho_{o,0}\ \beta_0')'$:

[16] A sufficient condition would be: $|\rho_{d,0}| + |\rho_{o,0}| \leq 1$.

$$\widehat{\delta} = \min_{\rho_d,\rho_o,\beta} e(\delta)'e(\delta), \text{ where } e(\delta) = y - (I_N - \rho_d W_d - \rho_o W_o)^{-1} \exp(Z\beta). \tag{30}$$

Eq. (31) shows the gradient of the NLS criteria function presented in Eq. (30) for $\delta = \delta_0$:

$$E\left[\triangledown|_{\delta=\delta_0} e(\delta)'e(\delta)\right] = E\left[-2\begin{pmatrix} SW_dS\exp(Z\beta_0) \\ SW_oS\exp(Z\beta_0) \\ (\iota_K' \otimes (S\exp(Z\beta_0))) \odot Z \end{pmatrix}' e(\delta_0)\right] = 0. \tag{31}$$

Since[17] $E\left[\triangledown|_{\delta=\delta_0} e(\delta)'e(\delta)\right] = 0$ the minimisation problem defined in Eq. (30) yields a consistent estimate for δ_0 (for more details see Pötscher and Prucha 1997). This nonlinear least squares estimator has the following asymptotic distribution:

$$N \to \infty\colon \frac{1}{\sqrt{N}}\left(\widehat{\delta} - \delta_0\right) \sim \mathcal{N}\left(0, G^{-1}H^{-1}\right),$$
$$\text{where } G = \Upsilon'\Upsilon,\ H = \Upsilon'\overline{\Omega}\Upsilon,\ \Upsilon = \begin{pmatrix} SW_dS\exp(Z\beta_0) \\ SW_oS\exp(Z\beta_0) \\ (\iota_K' \otimes (S\exp(Z\beta_0))) \odot Z \end{pmatrix}, \tag{32}$$

and $\overline{\Omega} = S\Omega S'$. In Eq. (32) $\mathcal{N}(\cdot)$ denotes a multivariate normal distribution. Since G and H are not known, we use their empirical equivalents; especially we use for the typical diagonal element of Ω,

$$\widehat{\sigma}_b^2 := \left(y_b - \sum_{q=1}^{N} \widehat{\rho}_d w_{d,i,q} y_q - \sum_{q=1}^{N} \widehat{\rho}_o w_{o,i,q} y_q - \exp\left(Z_i\widehat{\beta}\right)\right)^2,$$

where $w_{d,i,q}$ and $w_{o,i,q}$ are typical elements of the matrices W_d and W_o.

Since the two spatial lags introduce heteroscedasticity in the error term $e(\delta_0)$, the minimisation procedure described in Eq. (30) is inefficient. Therefore, an estimator that filters this kind of heteroscedasticity pattern, would improve the efficiency of the estimates. Such a pro-

[17] This only reflects one of the necessary assumptions in order for the nonlinear least squares estimator to be consistent. For a detailed list of all assumptions see Pötscher and Prucha (1997). Additionally, for assumptions regarding least distance estimators in general and nonlinear least squares for spatial autoregressive data generating processes see Jenish and Prucha (2010).

cedure would correspond to a heteroscedasticity-robust second stage estimation procedure, as given in Cameron and Trivedi (2005, 667pp). Using $\widehat{\rho}_d$ and $\widehat{\rho}_o$ from the first stage, we can construct $\widehat{y^*} = (I_N - \widehat{\rho}_d W_d - \widehat{\rho}_d W_o)\, y$ which is essentially the spatially filtered version of our origin-destination flows y. In the second stage, we fit $\widehat{y^*}$ on the explanatory variables $\exp(Z\beta)$. Note that in the second stage we receive no estimates for $\widehat{\rho}_d$ and $\widehat{\rho}_o$ since we filtered out the two spatial components. However, this is unproblematic as we are mainly interested in the models implied total effects that are given by the estimates of the second stage of the model. For the remainder of this paper, we therefore refer to our estimator as a 2-stage NLS estimator (2NLS).

5 Monte Carlo simulation study

5.1 Monte Carlo design

In the following Monte Carlo study, the data generating process of the flows y is given by Eqs. (13) - (15). Thus, we demonstrate the properties of our estimator for the case that the underlying distribution of the non-spatial process y^* is given by a Poisson distribution with mean and variance μ. However, as discussed before, our 2NLS estimator would also be efficient for other distributional assumptions, as long as the mean is correctly specified[18].

For simplicity, the matrix of explanatory variables Z includes one origin-specific variable ($K_1 = 1$), one destination-specific variable ($K_2 = 1$) and one origin-destination variable ($K_3 = 1$) and no constant term. We simulate the explanatory variables, each of size n-by-1, from a standard normal distribution[19] and then apply the Kronecker product transformations, as outlined above. The three explanatory variables are simulated just once for the entire Monte Carlo experiment.

The true parameter vector $\beta_0 = (\gamma_{d,0}, \gamma_{o,0}, \delta_0)'$ is varied over three different specifications: $\beta_{0,low} = (0.5, 0.3, -0.7)$, $\beta_{0,med} = (1.5, 0.9, -0.7)$ and $\beta_{0,high} = (2.5, 1.5, -0.7)$, respectively, to allow for varying means of the Poisson distribution. Higher values of β_0 lead to higher means in the Poisson process y^* and therefore result in a higher probability of large realisations. These distributions, showing large outliers, are likely to be found in empirical data sets, so they are explicitly included in the Monte Carlo design. Illustrative examples for the influence of the

[18]Monte Carlo results for the other data generating processes are given in the Appendix B.
[19]Thus $X_1, X_2, D \sim \mathcal{N}(0, 1)$, with $K_1 = K_2 = K_3 = 1$.

β-values and different degrees of spatial autocorrelation are given in Appendix A.

The spatial autocorrelation (or lag) parameters $\rho_{d,0}$ and $\rho_{o,0}$ are set to $\rho_{0,zero} = (\rho_{d,0}, \rho_{o,0}) = (0,0)$, $\rho_{0,lolo} = (0.1, 0.1)$, $\rho_{0,hilo} = (0.4, 0.1)$ and $\rho_{0,hihi} = (0.4, 0.4)$, to simulate different degrees of spatial autocorrelation[20]. We use two different $n \times n$ spatial weight matrices W for the construction of W_d and W_o. First, we use a simple first-order contiguity neighbourhood W_{cont}, where each region is considered a neighbour to the region on the main diagonal. By construction, each region has exactly two neighbours except the first and the last region, which have only one neighbour[21]. To allow inference about the properties of the 2NLS estimator for less sparse (or denser) spatial weight matrices (i.e. containing less zeros), we also consider an ordinal distance-based six closest neighbours spatial weight matrix W_{ord}, using Euclidean distances between geographic coordinates taken from the Pace and Barry (1997) dataset[22]. The two spatial weight matrices are normalised with respect to the maximum row sum.

Our dependent variable y is then derived by

$$y = \left(I_N - \rho_{d,0} W_d - \rho_{o,0} W_o\right)^{-1} y^*, \text{ where } E[y^*] = \mu = \exp\left(Z\beta_0\right) \text{ and } y^* \sim \mathcal{P}\left(\mu\right). \tag{33}$$

Finally, to analyse the behavior of our 2NLS estimator for different sample sizes, we implemented our Monte Carlo simulations for three sample sizes $n = (25, 50, 100)$. Given the estimator is asymptotically efficient, the bias in the estimates should decrease with increasing sample size. For each specification, we decided to conduct 1,000 Monte Carlo runs of $y^* \sim \mathcal{P}(\mu)$. To summarize, our Monte Carlo experiment includes three different β_0 parameters, four different ρ_0, two different spatial neighbourhood structures (W_{cont} and W_{ord}) and three different sample sizes n. This results in a total of $4 \times 3 \times 3 \times 2 = 72$ Monte Carlo experiments, with 1,000 repetitions each.

[20] The subscripts *hi* denotes high, *lo* low and *zero* no spatial autocorrelation.

[21] The typical element in W_{cont}, $w_{cont,r,j} = 1$ if region j and r are contiguous and $w_{cont,r,j} = 0$ otherwise. This neighbourhood structure corresponds to the case where regions are ordered along a straight line (as in LeSage and Pace 2009, p. 9).

[22] The typical element in W_{ord}, $w_{ord,r,j} = 1$ if region j is one of the six closest neighbours of r and $w_{ord,r,j} = 0$ otherwise.

5.2 Monte Carlo results

For the computation of the 2NLS estimator, we used sparse algorithms in order to minimise the problem given in Eq. (30). Applying these algorithms, optimising over inverse matrices reflecting a sample size of $n = 300$ (or $N = 90,000$), three explanatory variables and a six closest neighbour matrix[23], the computational time is approximately 10 minutes[24]. This illustrative example reflects a common empirical application for regional economists, since $n = 300$ is roughly the number of NUTS-2 regions in Europe and the six closest neighbour concept is empirically widely used.

From the estimation results of the simulations, we calculate the mean percentage bias of the point estimate and the root mean squared error of the standard deviation of the vectors β_0 and ρ_0 for each of the 72 Monte Carlo experiments, discussed above. The mean percentage bias of β_0, expressed in matrix notation, is given by

$$BIAS_{\widehat{\beta}} = \varnothing \left[diag[\beta_0]^{-1} (\overline{\widehat{\beta}} - \beta_0) \right] 100, \tag{34}$$

with $\overline{\widehat{\beta}}$ being the mean of the point estimates $\widehat{\beta}$ over the 1,000 Monte Carlo repetitions[25]. As β_0, $\overline{\widehat{\beta}}$, ρ_0 and $\overline{\widehat{\rho}}$ are vectors of dimension 3×1 and 2×1, we report the mean of the bias and root mean squared error of the standard deviation for the full vectors $\widehat{\beta}$ and $\widehat{\rho}$ and not each single parameter, for brevity. As a measure of the precision of our 2NLS estimator, we define the root mean squared error (RMSE) of the empirical standard deviation of the point estimate $\widehat{\beta}$ in percent of the true parameter β_0 by

$$RMSE_{\widehat{\beta}} = \varnothing \left[diag[\beta_0]^{-1} \sqrt{(\overline{\widehat{\beta}} - \beta_0) \odot (\overline{\widehat{\beta}} - \beta_0)} \right] 100. \tag{35}$$

The resulting measures of bias and standard deviation of our 72 Monte Carlo experiments are shown in Table 1. The upper part of Table 1 shows the results for the first-order contiguity weight matrix (W_{cont}). The columns indicate the respective means of the sample, by varying over $\beta_{0,low}$ referred to as *low mean*, $\beta_{0,med}$ referred to as *medium mean* and $\beta_{0,high}$ referred to

[23]Note that the optimisation time increases with the density of the spatial weight matrix W.
[24]Computational time is based on a 3.3 GHz x86 with 8 GB of RAM with Matlab 7.11.0.
[25]The mean percentage bias for ρ_0 is calculated accordingly.

as *high mean*.

Overall, the results indicate that the estimates show virtually no bias in mean and standard deviation of the model ($\widehat{\beta}$) and spatial lag ($\widehat{\rho}$) parameters. In almost all cases the bias is less than 1%, even in the sample of the smallest size ($n = 25$). Comparing the bias of the estimates and the RMSE across sample sizes, indicates that our estimates converge to the true parameter values. Both, the bias and the RMSE, decrease with increasing sample size, independent of ρ_0 and β_0. Another result worth noting is that the bias and RMSE decrease with increasing sample mean (i.e. higher probability of large realisations). This is especially interesting for empirical researchers, since they often face data with higher means and substantial outliers. In these cases, the Monte Carlo results indicate nearly unbiased estimates. It should, however, be noted that in empirical applications, the mean function and the spatial neighbourhood structure of the model has to be correctly specified. The same patterns can be observed in the lower part of Table 1, showing the results for a data generating process that corresponds to the (denser) ordinal distance spatial weight neighbourhood structure (W_{ord}, i.e. six closest neighbours).

As discussed in Section 4, the interpretation of the parameter estimates and the estimation method of the SPGM is irrespective of the distributional assumptions of y^* and y, given the first moment of the model is correctly specified. To demonstrate the performance of the 2NLS estimator, we conducted an additional Monte Carlo experiment of the same design for the following distributional assumption:

$$\begin{aligned} y &\sim \mathcal{P}(\breve{\mu}), \\ \breve{\mu} &= \left(I_N - \rho_{d,0}W_d - \rho_{o,0}W_o\right)^{-1} y^*, \end{aligned} \tag{36}$$

The Monte Carlo result for this data generating processes is shown in Appendix B. Bias and root mean squared error of the standard deviation are – similar to the DGP $y^* \sim \mathcal{P}(\mu)$ – negligible. Increasing the sample size further decreases the bias in the point estimates and the standard deviations.

Table 1: Monte Carlo experiment results for $y^* \sim \mathcal{P}(\mu)$: Mean percentage bias and root mean squared percentage error

				low mean				medium mean				high mean			
				$\hat{\beta}$		$\hat{\rho}$		$\hat{\beta}$		$\hat{\rho}$		$\hat{\beta}$		$\hat{\rho}$	
W	n	ρ_o	ρ_d	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
W_{cont}	25	0,0	0,0	0,015	0,034	-0,001	0,041	0,001	0,023	0,000	0,008	0,000	0,017	0,000	0,003
		0,1	0,1	-0,032	0,036	-0,001	0,036	0,000	0,023	-0,001	0,008	0,000	0,017	0,000	0,003
		0,4	0,1	0,008	0,034	-0,018	0,038	0,000	0,022	-0,003	0,008	0,000	0,017	0,000	0,003
		0,4	0,4	-0,006	0,036	0,000	0,040	0,000	0,022	0,000	0,008	0,000	0,017	0,000	0,004
	50	0,0	0,0	0,011	0,023	-0,001	0,022	0,002	0,032	0,000	0,014	0,000	0,003	0,000	0,001
		0,1	0,1	0,023	0,023	-0,001	0,020	-0,001	0,031	-0,001	0,011	0,000	0,003	0,000	0,001
		0,4	0,1	-0,021	0,023	0,002	0,021	0,001	0,033	-0,006	0,013	0,000	0,003	0,000	0,001
		0,4	0,4	0,008	0,024	0,000	0,027	0,000	0,031	0,000	0,012	0,000	0,003	0,000	0,001
	100	0,0	0,0	-0,002	0,010	0,000	0,010	0,000	0,005	0,000	0,003	0,000	0,003	0,000	0,001
		0,1	0,1	-0,005	0,010	0,000	0,009	0,000	0,005	0,000	0,003	0,000	0,003	0,000	0,001
		0,4	0,1	0,002	0,010	0,001	0,010	0,000	0,005	0,001	0,003	0,000	0,003	0,000	0,001
		0,4	0,4	0,005	0,011	0,000	0,012	0,000	0,005	0,000	0,003	0,000	0,003	0,000	0,001
W_{ord}	25	0,0	0,0	0,004	0,033	-0,001	0,048	0,002	0,066	-0,003	0,039	0,000	0,008	0,000	0,002
		0,1	0,1	-0,002	0,032	-0,001	0,040	-0,001	0,067	-0,002	0,027	0,000	0,008	0,000	0,001
		0,4	0,1	0,017	0,033	-0,007	0,045	0,001	0,067	-0,016	0,034	0,000	0,008	0,000	0,001
		0,4	0,4	0,032	0,035	0,000	0,042	-0,001	0,069	-0,001	0,029	0,000	0,008	0,000	0,001
	50	0,0	0,0	0,002	0,026	0,000	0,031	-0,001	0,017	0,000	0,010	0,000	0,009	0,000	0,004
		0,1	0,1	-0,009	0,025	0,000	0,025	0,000	0,017	0,000	0,007	0,000	0,009	0,000	0,003
		0,4	0,1	-0,002	0,027	-0,004	0,029	0,000	0,017	0,003	0,009	0,000	0,009	0,000	0,004
		0,4	0,4	0,007	0,026	0,000	0,027	0,000	0,017	0,000	0,006	0,000	0,009	0,000	0,002
	100	0,0	0,0	-0,002	0,010	0,000	0,014	0,000	0,004	0,000	0,003	0,000	0,004	0,000	0,001
		0,1	0,1	0,003	0,010	0,000	0,011	0,000	0,004	0,000	0,003	0,000	0,004	0,000	0,001
		0,4	0,1	-0,003	0,010	-0,001	0,013	0,000	0,004	0,000	0,003	0,000	0,004	0,000	0,001
		0,4	0,4	-0,002	0,010	0,000	0,013	0,000	0,004	0,000	0,002	0,000	0,004	0,000	0,001

6 An empirical illustration

We use the European patent citation flow data described in Fischer et al (2010) to illustrate how the suggested model works in a real data environment in comparison to the Poisson and the Negative Binomial gravity model specifications. Europe is represented in form of $n = 112$, generally the NUTS-2 regions of the countries Germany (38 regions), France (21 regions), Italy (20 regions), the Netherlands (12 regions), Belgium (11 regions), Austria (8 regions) and the NUTS-0 regions Luxembourg and Switzerland.

The explanatory variables matrix contains an origin-specific variable measured in terms of the log number of high-technology patents in the knowledge-producing region in the time period 1985-1997, a destination-specific variable measured in terms of the log number of high-technology patents in the knowledge-absorbing region in the time period 1990-2002, and a separation variable measured in terms of great circle distances [in km] between the economic centres of the regions[26]. We employ a binary first-order contiguity matrix implemented in row-standardized form to represent the neighborhood structure.

Table 2 shows the parameter estimates of the three models, with the estimates of the SPGM in the first column, the Poisson MLE estimates in the second and the Negative Binomial MLE estimates in the third column. Both MLE models show only highly significant coefficients. Larger stocks of high-tech patents in the origin and destination region are associated with larger patent citation flows between regions, with somewhat higher coefficents in the Negative Binomial model. Geographical distance impedes regional interaction measured by the patent citation flows. Again, the impeding impacts of the deterrence measures are much more distinct in the Negative Binomial model.

For the SPGM, given in the first column, we also find highly statistically significant parameters for the origin, destination and origin-destination variables. The size of the coefficient is compareable to that of the Poisson MLE model, given in the second column of Table 2. The coefficient on the origin-destination variable of the SPGM is negative and highly statistically significant. However, geographic distance has a much smaller impact on the patent citation flows compared to the other two models. The coefficient in the SPGM is -0.249 compared to -0.313

[26] Intra-zonal distances were set to zero.

Table 2: Estimation results for patent citation flows in Europe

	Spatial Poisson gravity model+		Poisson MLE++		Negative Binomial MLE++	
Constant	-8.965	***	-9.356	***	-8.367	***
	(0.852)		(0.236)		(0.167)	
Origin variable	0.774	***	0.825	***	0.863	***
	(0.042)		(0.015)		(0.011)	
Destination variable	0.771	***	0.794	***	0.827	***
	(0.042)		(0.015)		(0.009)	
Geographic distance	-0.249	***	-0.313	***	-0.588	***
	(0.023)		(0.019)		(0.017)	
Destination-based dependence	0.132	*	-		-	
	(0.075)		-		-	
Origin-based dependence	0.074		-		-	
	(0.061)		-		-	
R^2		0.83		-		-
RMSE+++		10.21		12.73		121.23
Pseudo Log-Likelihood		-		-24772.33		-16102.67

+ Estimated with heteroscedasticity-robust 2-stage nonlinear least squares. The spatial lag parameters ρ_o and ρ_d are taken from the first stage, whereas the remaining parameter estimates and statistics are derived from the second stage.
++ Maximum Likelihood estimation with robust standard errors.
+++ RMSE = Root mean squared error of $\widehat{y}$ (predicted outome).
Standard errors of coefficients are in brackets. $*$, $**$ and $***$ denote statistical significance at a 90, 95 and 99% confidence level .

(Poinsson MLE) and -0.588 (Negative Binomial).

The sizeable reductions in the parameter estimates of the spatial deterrence variable might be due to the positive and statistically significant destination-based spatial lag parameter. To some extent, the destination-based spatial lag may capture similar spatial patterns as the spatial impedence measure. This argument becomes more intuitive given the definition of the spatial lag parameter and the underlying neighbourhood structure. The destination-based lag parameter means that a patent citation flow from an origin region, say r, to a destination region j is positively correlated with a flow from origin r to a neighbouring region of j, say j'. The neigbourhood in our application is defined by means of contiguity, i.e. j and j' share a border and can be assumed to be close to each other, since our spatial units are NUTS-2 regions. Thus the distance from r to j and r to j' will be somewhat similar. Given that distance between regions deters patent citation flows between them, flows to more (less) distant regions and their neighbours will be smaller (larger) and, thus, positively correlated.

Concerning model selection criteria, the SPGM cannot be directly compared to the conventional Poisson and Negative Binomial models. As a matter of fact, the model with spatial autocorrelation incorporates two additional parameters: the spatial lags ρ_o and ρ_d. Thus, measures of model fit that can be applied to all three models must yield better results for the Poisson model with spatial autocorrelation, by definition. This can be seen, for example, by the root mean squared error (RMSE) statistics in Table 2. Other model information criteria like the adjusted R^2 or the Log-Likelihood are not defined for all three types of models. Still, given the difference in the point estimate of the origin-destination variable and the statistically significant destination-based spatial lag, a parametrically richer model like the SAR Poisson Gravitiy Model might be preferred compared to more restricted models like the Poisson or the Negative Binomial gravity model.

7 Conclusions

We introduce a Poisson gravity model with spatial dependence in the dependent (flow) variable. Previous methods for modeling discrete flow variables either: (i) did not adequately account for the zero flow problem; (ii) failed to account for the violation of the independence of flow assumption; or (iii) modelled the spatial dependence in the error term rather than in the dependent flow variable and, thus, misinterpreted the resulting parameter estimates. The model described in this paper cirumvents such deficiencies.

We start by augmenting a standard Poisson gravity model by introducing origin- and destination based spatial lags in a way suggested by LeSage and Pace (2008). It is shown that the model can be estimated within a 2-stage nonlinear least squares (2NLS) framework, yielding an estimator that does not rely on strict distributional assumptions of the data generating process such as the Poisson or Negative Binomial distribution, given that the first moment of the model is correctly specified. The estimator is heteroscedasticity-robust, i.e. it can account for over- or underdispersion in the data which is often experienced in empirical research.

As the SAR Poisson gravity model (SPGM) belongs to the family of spatial autoregressive models, the effect of the explanatory variables on the dependent variable has to include the indirect effects arising from spatial feedback effects (see LeSage and Pace 2009, 33pp). Due to the specification of the model, the parameter estimates can be interpreted as the implied total impact effects without further calculation. As a consequence of the flexibility of the estimator, the model interpretation is also valid for all distributional assumptions of the model.

We conducted Monte Carlo experiments for the distributional assumptions of (i) the observable (flow) variable being Poisson and (ii) the spatially filtered variable being Poisson distributed. The results indicate that our SPGM estimator shows virtually no bias in the parameter estimates, even for small sample sizes. Furthermore, bias in mean and standard deviation of the parameters decrease with increasing sample size, thus, indicating convergence towards the true parameter values.

Finally, the SPGM is illustrated using patent citation data. The results of our model indicate significant destination-based spatial dependence. Compared to conventional (non-spatial) Poisson and Negative Binomial models, the size of the coefficient on the spatial separation variable decreases substantially. This result might hint towards common spatial influences reflected by both, the spatial lag parameters and the variable used as origin-destination separation measure.

Acknowledgements

The first author, Richard Sellner, gratfully acknowledges the scholarship provided by the Innovation Economics Vienna - Knowledge and Talent Development Programme that is directed by the Institute of Economic Geography and GIScience at the Vienna University of Economics and Business in cooperation with the Foresight and Policy Development Departement at the Austrian Institute of Technology. The Institute of Economic Geography and GIScience has kindly given access to the institute's patent database. The authors further like to thank Prof. Jesus Crespo Cuaresma for his comments on the paper.

References

Bailey TC, Gatrell AC (1995) Interactive spatial data analysis. Addison-Wesley Longman, Harlow (UK)

Cameron AC, Trivedi PK (1998) Regression analysis of count data, 1st edn. Cambridge University Press, Cambridge

Cameron AC, Trivedi PK (2005) Microeconometrics: Methods and applications. Cambridge University Press, Cambridge

Davies RB, Guy CM (1987) The statistical modeling of flow data when the Poisson assumption is violated. Geographical Analysis 19(4):300–314

Fischer MM, Griffith DA (2008) Modeling spatial autocorrelation in spatial interaction data: An application to patent citation data in the European Union. Journal of Regional Science 48(5):969–989

Fischer MM, Reggiani A (2004) Spatial interaction models: From the gravity to the neural network approach. In: Capello R, Nijkamp P (eds) Urban dynamics and growth. Advances in urban economics. Elsevier, Amsterdam, pp 319-346

Fischer MM, Wang J (2011) Spaital data analysis: Models, methods and techniques. Springer, Berlin, Heidelberg and New York

Fischer MM, Scherngell T, Jansenberger E (2006) The geography of knowledge spillovers between high-technology firms in Europe: Evidence from a spatial interaction modell. Geographical Analysis 38(3):288–309

Fischer MM, Reismann M, Scherngell T (2010) Spatial interaction and spatial autocorrelation. In: Anselin L, Rey S (eds) Perspectives on spatial data analysis., Springer, Berlin, Heidelberg and New York, pp 61–79

Flowerdew R, Aitkin M (1982) A method of fitting the gravity model based on the Poisson distribution. Journal of Regional Science 22(2):191–202

Griffith DA (2003) Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization. Springer, Berlin, Heidelberg and New York

Jenish N, Prucha IR (2010) On spatial processes and asymptotic inference under near-epoch dependence. Working paper, Department of Economics, University of Maryland

Karlis D, Meligkotsidou L (2005) Multivariate Poisson regression with covariance structure. Statistics and Computing 15(4):255–265

Kelejian HH, Prucha IR (1998) A generalized spatial two-stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17(1):99–121

LeSage JP, Fischer MM (2010) Spatial econometric methods for modeling origin-destination flows. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis, Springer, Berlin, Heidelberg und New York, pp 409–433

LeSage JP, Pace K (2008) Spatial econometric modeling of origin-destination flows. Journal of Regional Science 48(5):941–967

LeSage JP, Pace K (2009) Introduction to spatial econometrics. CRC Press (Taylor and Francis Group), Boca Raton [FL], London and New York

LeSage JP, Fischer MM, 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

Pace RK, Barry R (1997) Quick computation of spatial autoregressive estimators. Geographical Analysis 29(3):232–247

Pötscher BM, Prucha IR (1997) Dynamic nonlinear econometric models: Asymptotic theory. Springer, New York

Sen A, Smith TE (1995) Gravity models of spatial interaction behavoir. Springer Berlin, Heidelberg and New York

Wilson A (1971) A family of spatial interaction models and associated developments. Environment and Planning A 3(1):1–32

A Appendix: Poisson distributions with different parameter values and spatial autocorrelation patterns

To illustrate the impact of the magnitude of the β parameter on the distributional form of y, consider the following experiment. We consider realisations of Poisson distributions with low (corresponding to $\beta_{0,low}$) and high means (corresponding to $\beta_{0,high}$) with either no spatial autocorrelation ($\rho_{d,0} = \rho_{o,0} = 0$) or substantial positive spatial autocorrelation ($\rho_{d,0} = \rho_{o,0} = 0.4$). For this simulation, we used a simple first-order contiguity neighbourhood matrix W_{cont}. Furthermore, we chose a sample size of $n = 100$, corresponding to $N = 10,000$ realisations from $y^* \sim \mathcal{P}(exp(Z\beta))$ and the according $y = (I_N - \rho_d W_d - \rho_d W_o)^{-1} y^*$. Figure A.1 shows the resulting four graphs.

For ease of visual comparison, we restrict the horizontal axes of the four graphs to a maximum value of 50. The upper parts of the figure, (a) and (b), show the distribution plots of y for the case of no spatial autocorrelation for a Poisson distribution with (a) a low mean $\beta_{0,low} = (0.5, 0.3, -0.7)$ and (b) a high mean $\beta_{0,high} = (2.5, 1.5, -0.7)$, respectively. Sample means of the simulated distributions are given in brackets at the bottom of each figure. The highest five realisations of y from a typical distribution of type (a) are between 25 and 30, whereas of type (b) they are between 30,000 and 85,000. Still, both distributions show a high probability of small realisations. Introducing spatial autocorrelation in the distributions, shown in (c) and (d) of

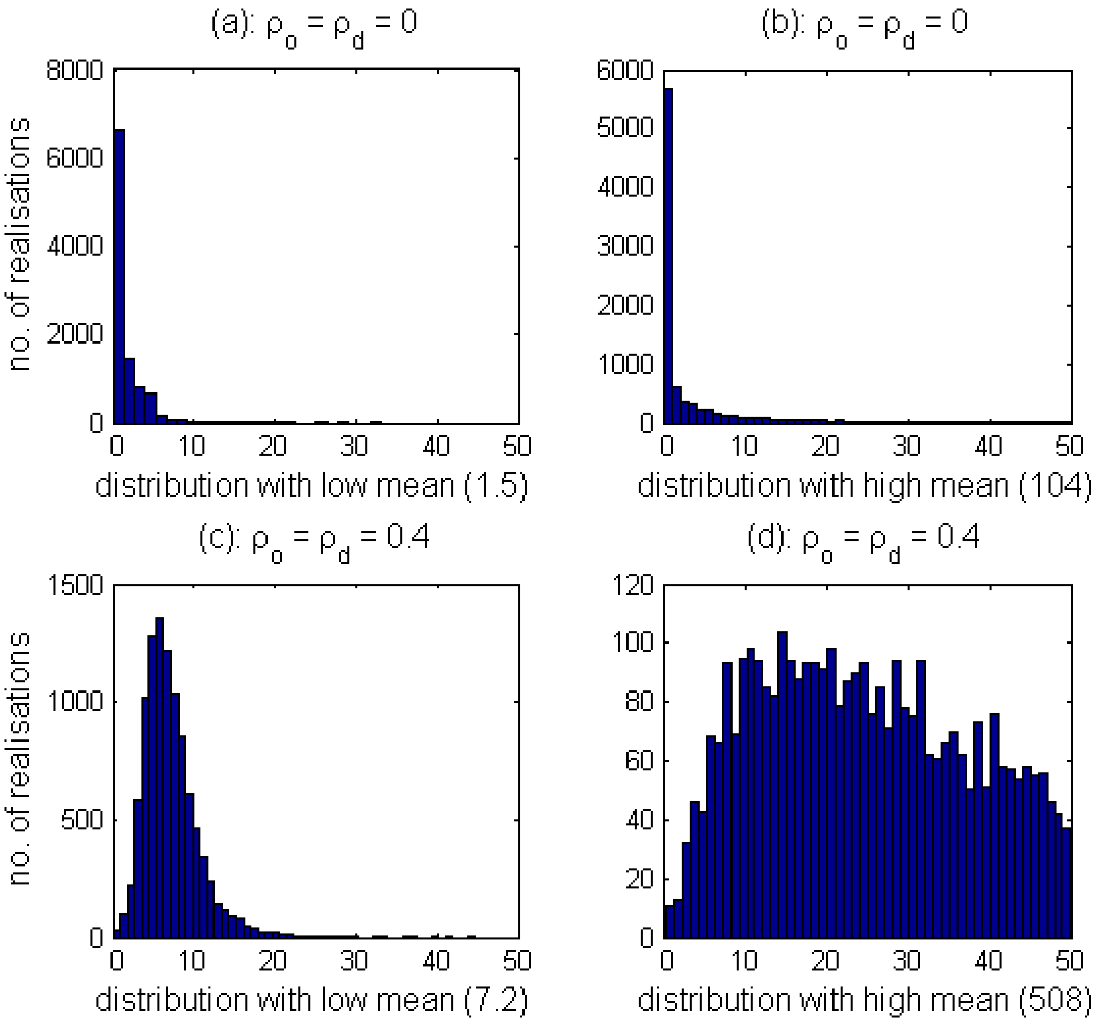

Figure A.1: Distribution of y for different means and spatial dependence patterns
Remarks: n=100, $W = W_{cont}$.

Figure A.1, a rather different picture emerges. Due to the spatial autocorrelation, the probability of very small realisations decreases significantly. The sample mean of y for an underlying Poisson distribution with low mean increases from (a) 1.5 to (c) 7 and for a distribution with high mean from (b) 104 to (d) 508, when spatial autocorrelation is present. The five largest realisations of these typical distributions are between (c) 35 and 45, and (d) 45,000 and 95,000.

B Appendix: Monte Carlo results for other distributional forms

The Monte Carlo results for the distributional assumption of y given in Eqs. (36) are summarized in Table B.1.

Table B.1: Monte Carlo experiment results for $y \sim \mathcal{P}(\breve{\mu})$: mean percentage bias and root mean squared percentage error

				low mean				medium mean				high mean			
				$\hat{\beta}$		$\hat{\rho}$		$\hat{\beta}$		$\hat{\rho}$		$\hat{\beta}$		$\hat{\rho}$	
W	n	ρ_o	ρ_d	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
W_{cont}	25	0.0	0.0	0.004	0.042	-0.001	0.044	-0.001	0.027	-0.001	0.015	0.000	0.007	0.000	0.000
		0.1	0.1	-0.028	0.047	-0.002	0.045	-0.002	0.031	-0.001	0.015	0.000	0.008	0.000	0.001
		0.4	0.1	-0.010	0.042	-0.009	0.047	0.000	0.027	-0.003	0.016	0.000	0.008	0.000	0.001
		0.4	0.4	-0.040	0.057	-0.001	0.045	0.001	0.037	0.000	0.015	0.000	0.010	0.000	0.001
	50	0.0	0.0	0.014	0.026	-0.001	0.022	0.001	0.014	0.000	0.004	0.000	0.002	0.000	0.000
		0.1	0.1	-0.016	0.027	-0.001	0.021	0.000	0.015	-0.001	0.007	0.000	0.002	0.000	0.001
		0.4	0.1	0.024	0.026	-0.010	0.023	0.000	0.014	0.001	0.007	0.000	0.002	0.000	0.001
		0.4	0.4	0.001	0.035	0.000	0.023	0.001	0.017	0.000	0.009	0.000	0.002	0.000	0.001
	100	0.0	0.0	-0.003	0.011	0.000	0.010	0.000	0.013	0.000	0.007	0.000	0.003	0.000	0.001
		0.1	0.1	-0.001	0.012	-0.001	0.010	0.001	0.014	0.000	0.007	0.000	0.003	0.000	0.001
		0.4	0.1	-0.003	0.011	-0.001	0.010	-0.001	0.013	0.001	0.008	0.000	0.003	0.000	0.001
		0.4	0.4	-0.002	0.015	0.000	0.011	0.000	0.016	0.000	0.006	0.000	0.003	0.000	0.001
W_{ord}	25	0.0	0.0	0.003	0.045	-0.003	0.057	0.000	0.040	0.000	0.025	-0.001	0.059	-0.002	0.022
		0.1	0.1	0.014	0.049	-0.002	0.053	-0.001	0.042	0.000	0.019	-0.001	0.059	0.001	0.016
		0.4	0.1	0.024	0.046	-0.019	0.060	0.001	0.040	-0.009	0.023	-0.001	0.058	-0.011	0.020
		0.4	0.4	-0.023	0.062	-0.001	0.066	0.000	0.048	0.000	0.018	0.001	0.067	0.000	0.017
	50	0.0	0.0	-0.017	0.025	-0.001	0.033	0.000	0.012	0.000	0.010	0.000	0.014	0.000	0.005
		0.1	0.1	-0.002	0.026	-0.001	0.028	0.000	0.012	0.000	0.011	0.000	0.014	0.000	0.004
		0.4	0.1	0.023	0.025	-0.006	0.032	0.000	0.012	-0.005	0.011	0.000	0.014	-0.002	0.005
		0.4	0.4	-0.017	0.032	0.000	0.033	-0.001	0.013	0.000	0.010	0.000	0.015	0.000	0.004
	100	0.0	0.0	0.000	0.014	0.000	0.017	0.000	0.009	0.000	0.006	0.000	0.003	0.000	0.001
		0.1	0.1	0.005	0.015	0.000	0.014	0.000	0.009	0.000	0.004	0.000	0.003	0.000	0.001
		0.4	0.1	0.006	0.014	-0.001	0.016	0.000	0.009	0.000	0.005	0.000	0.003	0.000	0.001
		0.4	0.4	0.001	0.017	0.000	0.015	0.000	0.010	0.000	0.004	0.000	0.003	0.000	0.001