A spatial autoregressive Poisson gravity model

Sellner, Richard and Fischer, Manfred M. and Koch, Matthias

Institute for Advanced Studies, Vienna, Vienna University of Economics and Business

2010

A spatial autoregressive Poisson gravity model

Richard Sellner[1], Manfred M. Fischer[2], and Matthias Koch[2]

[1]Institute for Advanced Studies, Economics and Finance, Applied Research, Stumpergasse 56, A-1060 Vienna
[2]Institute for Economic Geography and GIScience, Vienna University of Economics and Business, Nordbergstrasse 15, A-1090 Vienna

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 often is present in the empirical analysis of discrete data or, in the case of overdispersion, if spatial autocorrelation is present. This estimator can be shown to have 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 defined as the sum of direct and indirect spatial feedback effects. Monte Carlo results indicate marginal finite sample biases in the mean and standard deviation of the parameter estimates and convergence to the true parameter values as the sample size increases. In addition, the paper illustrates the model by analysing patent citation flows data across European regions.

Corresponding author: Richard Sellner, Institute for Advanced Studies, Economics and Finance, Applied Research, Stumpergasse 56, Vienna AT-1060, e-mail: sellner@ihs.ac.at.

Introduction

Gravity models[1] — sometimes also called spatial interaction models — represent a class of models that utilize 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, as well as flows of information such as telephone calls or electronic messages, and even the transmission of knowledge. Locations may be either area or point units.

Gravity models typically rely on three types of factors to explain origin-destination flows: *first*, origin-specific variables that characterize 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 characterize 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 specification 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, because 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[3] 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 flows problem. Mathematically, this approach involves 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, the selected constant 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 of accounting for spatial dependence 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[4] Poisson model suggested by LeSage et al. (2007).

In contrast, Fischer and Griffith (2008) suggest incorporating spatial dependence in the disturbance process, as in the case of serial correlations in time series regression models. They apply Griffith's spatial filtering methodology to deal with the issue of spatial dependence in Poisson gravity models[5]. Within this semiparametric approach[6], 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[7] 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 that are necessary for a proper model interpretation (as motivated by LeSage and Pace 2009).

A Poisson gravity model

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

In a Poisson gravity model, y_i is assumed to follow an independently distributed Poisson process, such that

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

whose mean parameter μ_i is given by[8]

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

where $z_i = (z_{i,1}, \dots, z_{i,K})$ represents vectors of (logged) origin-specific, destination-specific and origin-destination variables[9] associated with the i-th observation, and $\beta = (\beta_1, \dots, \beta_K)'$ is a K-by-1 ($k = 1, \dots, K$) parameter vector. The exponential function appearing in μ_i is justified by the positivity of μ_i, because a linear function

$\mu_i = z_i\beta$ possibly would imply 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

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

which has to be optimized numerically.

In the Poisson gravity model, y_i has mean $\mu_i = exp(z_i\beta)$ and variance[10] 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 with gamma-distributed unobserved individual heterogeneity reflecting the fact that the true mean is not perfectly observed.

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 (logged) origin, destination and origin-destination characteristics and a constant term, with the associated K-by-1 vector of regression parameters[11] β. Hence, we may write

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

with

$$X_d = \iota_n \otimes X_1 \quad \text{and} \tag{8}$$
$$X_o = X_2 \otimes \iota_n \tag{9}$$

where X_1 is an n-by-K_1 matrix of (logged) destination-specific characteristics, and X_2 is 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 parameter[12].

The spatial autoregressive Poisson gravity model

In gravity models, the assumption that observations are mutually independent is a heroic assumption — as LeSage and Pace (2009, p. 212) point out — because origin-destination flows are fundamentally spatial in nature. Models of the type given by Eqs. (5) and (6) neglect spatial dependencies in the origin-destination flows contained in the dependent variable y, hereafter termed *spatial autocorrelation*. Following LeSage and Pace (2008), we define spatial autocorrelation 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 neighbors to region r, which they label origin-dependence; and, (ii) flows from origin region r to regions neighboring the destination region j, say regions j' and j'', which they label destination-dependence[13].

Introducing spatial autocorrelation in the Poisson gravity model

In our model, the original variable, 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

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

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

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

where W is an n-by-n spatial weight matrix with diagonal elements set to zero, and I_n is an n-by-n identity matrix. Typically the spatial weight matrix is normalized to have row sums of unity[14], and thus produces linear combinations of flows from neighboring regions (see LeSage and Pace 2009, p. 10). Non-zero entries in matrix W indicate that a neighborhood relation exists between the corresponding regions. Neighbors 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 nearest neighbors). Given an origin-centric organization of 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 neighbor 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 neighbor 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 (SPGM) takes the form[15]:

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

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

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

As is shown subsequently, 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 is valid for a more general model class. However, for the following discussion, we retain the assumption of $y^* \sim \mathcal{P}(\mu)$ and examine the resulting dispersion properties of a spatially autocorrelated Poisson distributed variable.

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[16]

$$E[y] = (I_N - \rho_d W_d - \rho_o W_o)^{-1}\mu, \quad (17)$$
$$Var[y] = (I_N - \rho_d W_d - \rho_o W_o)^{-1}\text{diag}[\mu](I_N - \rho_d W_d' - \rho_o W_o')^{-1}, \quad (18)$$

where the operator diag$[\cdot]$ transforms the vector μ into a diagonal matrix, and I_N is an N-by-N identity matrix. Equations (17) and (18) imply 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). Because each entry in the vectors $E[y]$ and $\text{diag}[Var[y]]$ may differ, an analytic comparison between Eqs. (17) and (18) with respect to over- or underdispersion[17] 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 one-step-forward contiguity neighborhood structure. Because W is row-normalized, $E[y^*] = \iota_N$ 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 simulated 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 constructed τ, which denotes the average percentage distance between the mean and variance of y:

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

whereas the operator $\varnothing[\cdot]$ takes the average of a vector. If $\tau > 0$ ($\tau < 0$), y shows a tendency toward overdispersion (underdispersion).

Figure 1 portrays the relationship between τ and both ρ_o and ρ_d. 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

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

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

We simulated 10,000 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 at least 9,500 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 of $\rho_o, \rho_d > 0$ leads to significant underdispersion, but, to a somewhat lesser extent. Values around zero for both

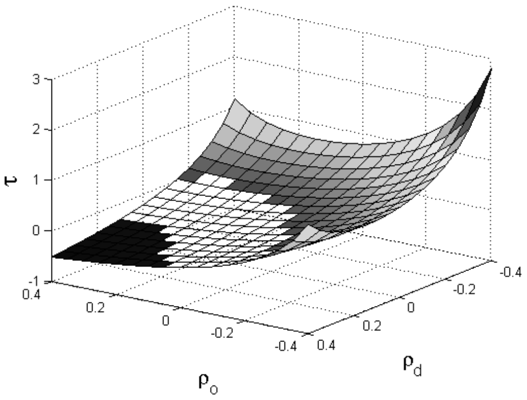

Figure 1: Simulated mean percentage distance between the mean and variance of y

Coloured (transparent) cells indicate significant (insignificant) $sign(\tau)$ with 95 percent accuracy.

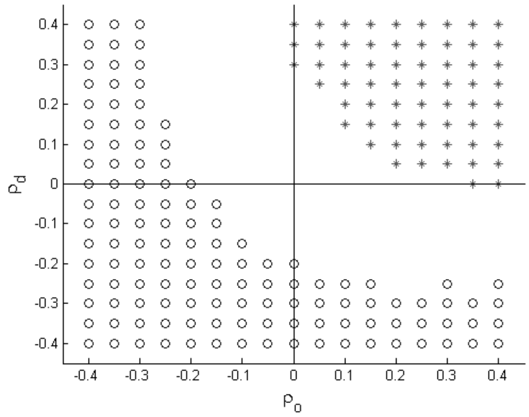

Figure 2: Significant over- and underdispersion for different spatial lag parameter values

Symbols: o indicates significant overdispersion and * significant underdispersion.

ρ_o and ρ_d do not lead to significant dispersion patterns. Mixtures of positive and negative spatial autocorrelation parameters translate into significant overdispersion for values approximately greater than $|0.2|$. For a better depiction of this pattern, consider Figure 2, which highlights spatial lag parameters that result in significant over- or underdispersion[18]. As can be seen more clearly in this figure, for mixed signs in spatial autocorrelation parameters — $\rho_o < 0$ and $\rho_d > 0$, or $\rho_o > 0$ and $\rho_d < 0$ — significant overdispersion is present for negative spatial autocorrelation parameters smaller than -0.20.

Model interpretation and estimation

As the SPGM belongs to the class of spatial autoregressive 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)[19]. Like LeSage and Pace (2009), we suggest *total impact effects* for the interpretation of the coefficients.

Model interpretation

We define the total average effects in a sense that they reflect the total average elasticity of the explanatory variables with respect to the expected value of y. By construction[20], the elasticity of $E\left[y_i\right]$ with respect to z is given by

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

where $z_{k,q}$ 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 average total impacts of the k^{th} variable on y by

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

as suggested in LeSage and Pace (2009, pp. 36-38). Because 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 element-wise 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 DGP 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 $\rho_{d,0}$ or $\rho_{o,0}$, and are equal to $\beta_{0,k}$. Thus, in contrast to spatial autoregressive models that are linear in the spatial autocorrelation coefficient, there is no need for a simulation approach to produce an empirical distrubtion of the parameters because the SPGM total effects only depend on the true parameter vector β_0. In other words,

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

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

The direct and indirect elasticities respectively, are given by

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

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

The direct and indirect effects in this model still depend on $\rho_{d,0}$ and $\rho_{o,0}$, and therefore their standard deviations need to be simulated. To derive an efficient simulation algorithm, Eq. (26) needs to be rewritten in the matrix form

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

where $\oslash$ denotes element-wise division of matrices or vectors. 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. Because our model has two spatial weight matrices (W_d and W_o), the algorithms for the computation of the log-determinants given in LeSage and Pace (2009, Chapter 4) need to be adopted.

A two-stage nonlinear least squares estimator for the model

As mentioned previously, 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[21] has been derived to date. 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 generalize the estimation results with respect to different distributional assumptions of the underlying DGP, we assume that the spatially filtered variable y^* can have any distribution $C\left(\mu, \Omega\right)$ with a specified (finite) mean μ, and an unspecified (finite) diagonal variance-matrix Ω with 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[22]). Note that if $\rho_{d,0} = \rho_{o,0} = 0$ the DGP collapses to the one described in Cameron and Trivedi (2005, 150pp), which can be estimated by NLS used for Poisson distributed random variables. Similar to 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 $\digamma$ is smaller than one[23], where $\digamma := \rho_{d,0}W_d + \rho_{o,0}W_o$. Thus the inverse of $I_N - \digamma$ exists (for details, see Kelejian and Prucha 1998). Hence, Eq. (13) can be solved for y to obtain

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

Because 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 $E\left[y\right] = (I_N - \rho_{d,0}W_d - \rho_{o,0}W_o)^{-1} \exp\left(Z\beta_0\right)$. The DGP in Eq. (29) is more general than the one given in Eq. (13), because 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, and the variance is quadratic in mean such that $Var[y^*] = \mu + \lambda(\mu \odot \mu)$ with dispersion parameter λ. Therefore, if $\rho_{d,0} = \rho_{o,0} = 0$, then y is Negative Binomial distributed as well. In both cases, if $\rho_{d,0} \neq 0$ and $\rho_{o,0} \neq 0$, 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) does not account for possible overdispersion in y, because $E[y] = Var[y] = \left(I_N - \rho_{d,0}W_d - \rho_{o,0}W_o\right)^{-1}\mu$. However, independent of the distributional assumption, each DGP results in the same first moment of y. Because a heteroscedasticity-robust NLS estimator requires only a correctly specified first moment (mean), our estimation procedure results in the same estimates independent of the underlying distribution of y^* or y, and, hence, is robust against a huge variety of possible misspecification of the distribution[24].

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

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

The gradient of the NLS criteria function presented in Eq. (30) for $\delta = \delta_0$ is

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

Because[25] $E\left[\nabla|_{\delta=\delta_0} e(\delta)'e(\delta)\right] = 0$, the minimization problem defined by Eq. (30) yields a consistent estimate for δ_0 (for more details, see Pötscher and Prucha 1997). This NLS estimator has the following asymptotic distribution:

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

and $\overline{\Omega} = S\Omega S'$. In expression (32), $\mathcal{N}(\cdot)$ denotes a multivariate normal distribution. Because 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.

Because the two spatial lags introduce heteroscedasticity into the error term $e(\delta_0)$, the minimisation procedure described in Eq. (30) yields inefficient estimates. Therefore, an estimator that filters this kind of heteroscedasticity pattern improves the efficiency of the estimates. Such a procedure corresponds 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 $\widehat{y^*} = \left(I_N - \widehat{\rho}_d W_d - \widehat{\rho}_d W_o\right) y$, which essentially is the spatially filtered version of our origin-destination flows y. In the second stage, we predict $\widehat{y^*}$ with the explanatory variables $\exp(Z\beta)$. In the second stage, we obtain no estimates for $\widehat{\rho}_d$ and $\widehat{\rho}_o$ because we filter out the two spatial components. However, this is unproblematic because we are mainly interested in the models implied total effects that are given by the estimates of the second stage of the model. We therefore refer to our estimator as a 2-stage NLS estimator (2NLS).

Monte Carlo simulation study

In the following Monte Carlo study, the DGP 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 previously, our 2NLS estimator also is efficient for other distributional assumptions, as long as the mean is correctly specified.

Monte Carlo design

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[26], and then apply the Kronecker product transformations, as previously outlined. The three explanatory variables are simulated just once for each sample size.

The true parameter vector $\beta_0 = (\gamma_{d,0}, \gamma_{o,0}, \delta_0)'$ is varied over the following three different specifications (respectively, to allow for varying means of the Poisson distribution): $\beta_{0,low} = (0.5, 0.3, -0.7)$, $\beta_{0,med} = (1.5, 0.9, -0.7)$ and $\beta_{0,high} = (2.5, 1.5, -0.7)$. 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. The Appendix presents illustrative examples for the influence of the β-values and different degrees of spatial autocorrelation.

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[27]. We use two different n-by-n spatial weight matrices W for the construction of W_d and W_o. First, we use a simple one-step-forward first-order contiguity neighborhood W_{cont}, where the regions are organized in a chain (see LeSage and Pace 2009, p.3). By construction, each region has exactly two neighbors, except for the first and the last regions, which have only one neighbor[28]. To allow inference about the properties of the 2NLS estimator for less sparse (or denser) spatial weight matrices (i.e., containing fewer zeros), we also consider an ordinal distance-based six nearest neighbors spatial weight matrix W_{ord}, using Euclidean distances between geographic coordinates taken from the Pace and Barry (1997) data set[29]. The two spatial weight matrices are normalized with respect to the maximum row sum (called row-standardisation).

Our dependent variable y is generated by

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

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 and 100. Given that 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 neighborhood structures (W_{cont} and W_{ord}), and three different sample sizes n. This design results in a total of $4 \times 3 \times 3 \times 2 = 72$ Monte Carlo experiments, with 1,000 repetitions each.

Monte Carlo results

For the computation of the 2NLS estimator, we used sparse algorithms in order to minimize 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 nearest neighbor matrix[30], the computational time is approximately 10 minutes[31]. This illustrative example reflects a common empirical application for regional economists, because $n = 300$ is roughly the number of NUTS-2 regions in Europe, and the six nearest neighbor concept is widely used in practice.

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 previously discussed. 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 $\widehat{\beta}$ for the 1,000 Monte Carlo repetitions[32]. Because β_0, $\overline{\widehat{\beta}}$, ρ_0 and $\overline{\widehat{\rho}}$ are vectors of dimension 3-by-1 and 2-by-1, for brevity we report the mean of the bias and root mean squared error (RMSE) of the standard deviation for the full vectors $\widehat{\beta}$ and $\widehat{\rho}$, and not each single parameter. 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

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

Table 1 summarizes the resulting measures of bias and standard deviation of our 72 Monte Carlo experiments. 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 one percent, 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 is that the bias and RMSE decrease with increasing sample mean (i.e., a higher probability of large realisations). This finding is especially interesting for empirical researchers, because they often face data with higher means and substantial outliers. In these cases, the Monte Carlo results indicate nearly unbiased estimates. For these properties to hold in empirical applications, the mean function and the spatial neighborhood structure of the model has to be correctly specified.

As discussed previously, the interpretation of the parameter estimates and the estimation method of the SPGM is irrespective of the distributional assumptions for y^* and y, given that the first moment of the model is correctly specified. In addition to the DGP given by Eq. (33), we conducted a Monte Carlo experiment using the same design for the following distributional assumption to demonstrate the performance of the 2NLS estimator:

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

Table 2 summarizes the Monte Carlo results for this DGP. 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 (BIAS), and root mean squared percentage error (RMSE)

				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

Table 2: Monte Carlo experiment results for $y \sim \mathcal{P}(\check{\mu})$: mean percentage bias (BIAS), and root mean squared percentage error (RMSE)

				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

An empirical illustration

We use the European patent citation flow data described in Fischer et al. (2010) to illustrate how the proposed model works in a real data environment in comparison to the Poisson and the Negative Binomial gravity model specifications. Europe comprises $n = 112$ areal units, 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 of 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[33]. We employ a binary first-order contiguity matrix implemented in row-standardized form to represent the neighborhood structure.

Table 3 summarizes the parameter estimates of the three models, with the estimates of the SPGM in the first, 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 regions 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 comparable to that of the Poisson MLE model, given in the second column of Table 3. 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 (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[34]. These results connect to the early findings of Cliff et al. (1974), Curry et al. (1975) or Griffith and Jones (1980). To some extent, the destination-based spatial lag captures similar spatial patterns as the spatial impedence measure. This argument becomes more intuitive given the definition of the spatial lag parameter and the underlying neighborhood structure. The destination-

Table 3: 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 (pseudo-R^2)		0.83		(0.81)		(0.29)
RMSE+++		10.21		12.73		121.23
Pseudo Log-Likelihood		-		-24772.33		-16102.67

+ Estimated with heteroscedasticity-robust 2NLS. 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 denotes root mean squared error of $\widehat{y}$ (predicted outome).
Standard errors of coefficients are in brackets. $*$, $**$ and $***$ denote statistical significance at 90%, 95% and 99% confidence levels.

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 neighboring 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 neighbors 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 gravity 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 3. 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 spatial autoregressive (SAR) Poisson Gravitiy Model (SPGM) might be preferred compared to more restricted models like the Poisson or the Negative Binomial gravity model.

Turning to the issue of over- and underdispersion, it can be seen from Table 3 that the Log-Likelihood of the Negative Binomial model is larger than that of the non-spatial Poisson model. This indicates that the spatially unfiltered data is overdispersed. However, as both spatial lags in the SPGM turn out to be positive (with one being significant), Figure 2 suggests that the data is located in the underdispersion region. Connecting these two results suggests that the spatially unfiltered data is highly overdispersed.

Conclusions

We introduce a Poisson gravity model with spatial dependence in the dependent (flow) variable. Previous methods for modeling discrete flow variables: (i) did not adequately account for the zero flows problem; (ii) fail to account for the violation of the independence of flow assumption; or (iii) model the spatial dependence in the error term rather than in the dependent flow variable and, thus, misinterpret 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). We show that the model can be estimated within a 2NLS framework, yielding an estimator that does not rely on strict distributional assumptions of the DGP 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 data which often is experienced in empirical research.

Because 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). 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 also is 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 decreases with increasing sample size, thus indicating convergence toward 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 for 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 an origin-destination separation measure.

Acknowledgements

The first author, Richard Sellner, gratfully acknowledges the scholarship provided by the Knowledge and Talent Development Programme that is directed by the Institute of Economic Geography and GIScience at the Vienna University of Economics and Business, and the Foresight and Policy Development Departement of the Austrian Institute of Technology. The Institute of Economic Geography and GIScience has kindly given access to the institute's patent database.

Notes

[1]The term gravity model comes from the Newtonian analogy for the models. In view of the enormous interest generated by gravity models, a host of different theoretical approaches

proposed for these models is not surprising. Wilson (1970) provides a motivation for spatial interaction models based on entropy-maximizing theory, Sen and Smith (1995) propose a choice-theoretical foundation, Niedercorn and Bechdolt (1969) derive the model from utility maximization, while Fischer and Reismann (2002) use the concept of neural computing. For a discussion, see Fischer and Reggiani (2004) or Sheppard (1978), 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.

[3]For a detailed discussion of models see, for example, Haworth and Vincent (1979). The log-additive approach assumes a lognormal distribution of the error term. If this assumption holds, then the resulting ordinary least squares (OLS) estimator is the best linear unbiased estimator (BLUE). If the log of the error term is identical, independent and not normal-distributed then the OLS estimator is still unbiased and in general considered a very 'reasonable' estimator by most textbooks. However, two assumptions are really vital for transforming the nonlinear data generating process to a log-linear data generating process: First, the error term has to enter the model in a multiplicative way. Second, none of the realizations of the dependent variable is zero. If the second assumption is violated, the logarithm is not defined, and hence the zero flows problem arises.

[4]The term 'hierarchical' in Bayesian statistics refers to models with two or more levels of random variables or models with latent variables.

[5]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.

[6]This approach combines parametric and nonparametric models. For instance, the eigenvector spatial filtering approach in Griffith (2003) eliminates the spatial autocorrelation by nonparametric methods. These filters can then be used in parametric regression models.

[7]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).

[8]For a more detailed discussion of the link between the exponential model and the multiplicative form, see Fischer and Wang (2011, pp. 53-59).

[9]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. Accordingly, the coefficient estimates reflect elasticity responses of origin-destination flows to the various origin, destination, and origin-destination characteristics.

[10]The restriction that the variance equals the mean in a Poisson specification usually is called equidispersion (see Cameron and Trivedi 1998, p.4).

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

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

[13]We do not consider the third type of spatial dependence of LeSage and Pace (2008) in flows 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 neighboring the origin region r to regions neighboring the destination region j, say flows from regions r' and r'' to regions j' and j'', which they label origin-to-destination dependence.

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

[15]We focus on an econometric solution to the problem of including spatial autocorrelation in the data generating process (DGP). However, a spatially autocorrelated (at least in an econometric sense) Poisson data generating process may not exist from a statistical point of view, because a spatially autocorrelated variable has non-integer outcomes. One possibility for a Poisson process that has some properties of a spatially autocorrelated process would be to generate a Poisson distributed vector with the mean $(I - \rho_d W_d - \rho_o W_o)^{-1} \exp(X\beta)$. Such a 'statistical' DGP would share the same interpretation as our proposed econometric DPG, and our nonlinear least squares estimation framework could be applied to it as well without any modifications. However, because $\exp(X\beta)$ is by construction positive, only negative spatial autocorrelation parameters can assure a non-negative mean which is required, per definition, for a Poisson distribution outcome.

[16]We assume that the inverse of $(I_N - \rho_d W_d' - \rho_o W_o')$ exists.

[17]The terms overdispersion and underdispersion refer to $E[y] < Var[y]$ and $E[y] > Var[y]$, respectively. Two types of overdispersion might exist in a data generating process. First, the data can be overdispersed, as often is found in empirical discrete data sets. The second type can be observed if the data are spatially autocorrelated. The estimator proposed in this paper deals with both types of overdispersion. Therefore, we do not distinguish between the two throughout the paper.

[18]Darker colours at the lower left of the graph indicate underdispersion, whereas brighter colours at the upper right indicate overdispersion. The specific colour does not matter (if, for instance, the paper is viewed in gray tones), because we know which combination of values of ρ_o and ρ_d lead to over- or underdispersion.

[19]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.

[20]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.

[21]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, such as in Karlis and Meligkotsidou (2005). However, these algorithms are much more computationally time consuming than the approach we suggest in this paper.

[22]Again, we refer to both types of overdispersion that might be present in discrete data (first type), and that is possible spatial autocorrelation (second type).

[23]A sufficient condition is $|\rho_{d,0}| + |\rho_{o,0}| \leq 1$.

[24]For instance, our estimator yields efficient results given a Poisson or Negative Binomial distribution. However, our method is inefficient given distrubtions like Cauchy, which has no finitely defined moments.

[25]This reflects only one of the necessary assumptions in order for the NLS 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 NLS for spatial autoregressive DGP in particular, see Jenish and Prucha (2012).

[26]Thus, $X_1, X_2, D \sim \mathcal{N}(0, 1)$, with $K_1 = K_2 = K_3 = 1$.

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

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

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

[30]The optimization time increases with the density of the spatial weight matrix W.

[31]Computational time is based on a 3.3 GHz x86 machine with 8 GB of RAM running Matlab 7.11.0.

[32]The mean percentage bias for ρ_0 is calculated accordingly.

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

[34]Note that the size of this coefficient is not negligible given our parameter restriction that the sum of both spatial lag coefficients is less than one.

References

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive spatial data analysis.* Addison-Wesley Longman, Harlow (UK).

Cameron, A. C. and Trivedi, P. K. (1998). *Regression analysis of count data.* Cambridge University Press, Cambridge, 1 edition.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and applications.* Cambridge University Press, Cambridge.

Cliff, A. D., Martin, R. L., and Ord, J. K. (1974). Evaluating the friction of distance parameter in gravity models. *Regional Studies*, 8:281–86.

Curry, L., Griffith, D. A., and Sheppard, E. S. (1975). Those gravity parameters again. *Regional Studies*, 9:289–96.

Davies, R. B. and Guy, C. M. (1987). The statistical modeling of flow data when the Poisson assumption is violated. *Geographical Analysis*, 19(4):300–314.

Fischer, M. M. and Griffith, D. A. (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, M. M. and 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, M. M. and Reismann, M. (2002). A methodology for neural spatial interaction models. *Geographical Analysis*, 34(3):207–228.

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

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

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

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

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

Griffith, D. A. and Jones, K. (1980). Explorations into the relationship between spatial structure and spatial interaction. *Environment and Planning A*, 12(2):187–201.

Haworth, J. M. and Vincent, P. J. (1979). The stochastic disturbance specification and its implications for log-linear regression. *Environment and Planning A*, 11(7):781–790.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics*, 170(1):178–190.

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

Kelejian, H. H. and Prucha, I. R. (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, J. P. and Fischer, M. M. (2010). Spatial econometric methods for modeling origin-destination flows. In Fischer, M. M. and Getis, A., editors, *Handbook of applied spatial analysis*, pages 409–433. Springer, Berlin, Heidelberg und New York.

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

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

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

Niedercorn, J. H. and Bechdolt, B. V. (1969). An economic derivation of the 'gravity law' of spatial interaction. *Journal of Regional Science*, 9(2):273–282.

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

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic nonlinear econometric models: Asymptotic theory.* Springer, New York.

Sen, A. and Smith, T. E. (1995). *Gravity models of spatial interaction behavior.* Springer Berlin, Heidelberg and New York.

Sheppard, E. S. (1978). Theoretical underpinnings of the gravity hypothesis. *Geographical Analysis*, 10(4):386–401.

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

Wilson, A. G. (1970). *Entropy in urban and regional modelling.* Pion Limited, London.

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 created 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 ($\rho_{d,0} = \rho_{o,0} = 0.4$)[35]. For this simulation, we used a simple first-order contiguity neighborhood 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 its corresponding $y = (I_N - \rho_d W_d - \rho_d W_o)^{-1} y^*$. Figure 3 portrays the resulting four graphs.

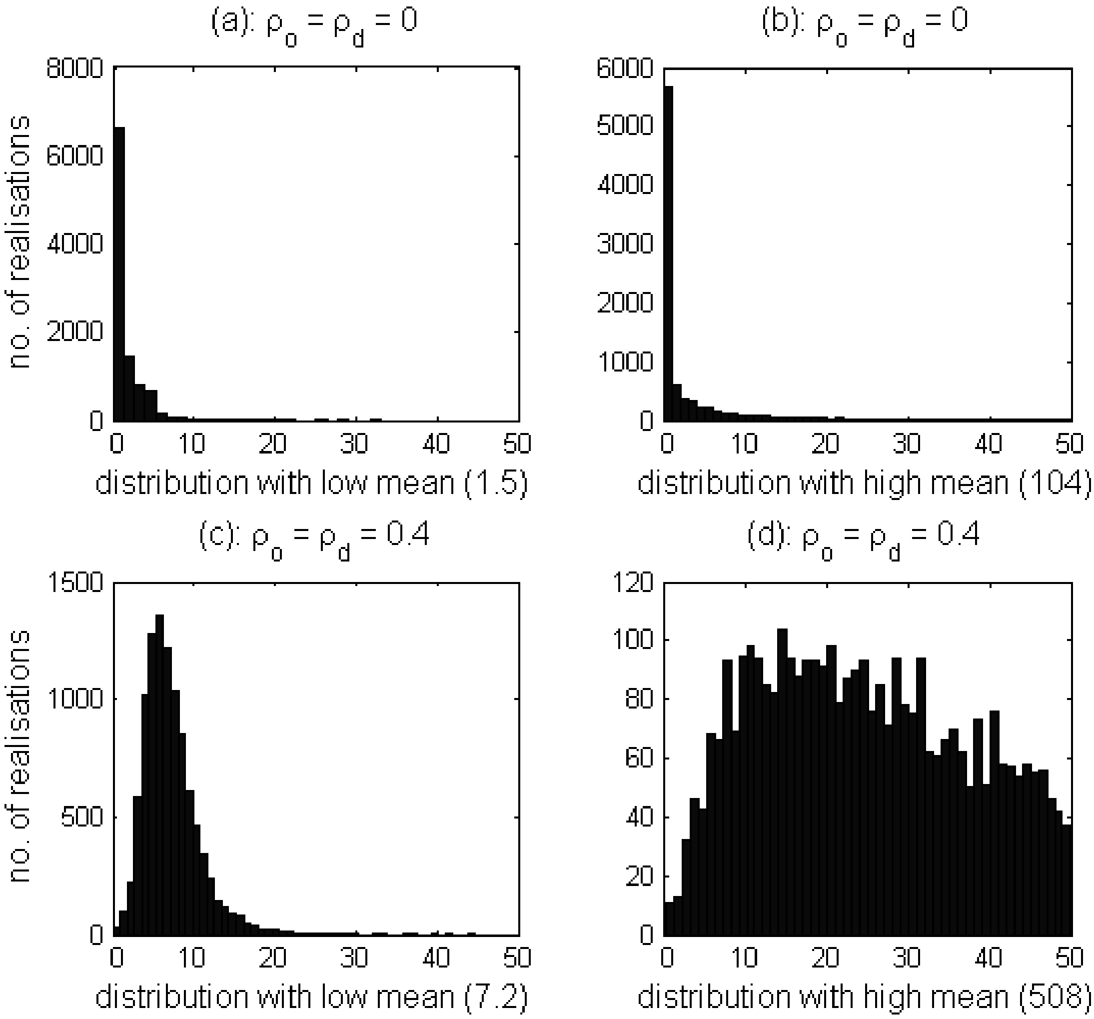

Figure 3: Distribution of y for different means and spatial dependence levels
Remarks: n=100, $W = W_{cont}$.

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 below 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 Fig. 3 (c) and (d), 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.