ePub^{WU} Institutional Repository

James P. LeSage and Manfred M. Fischer and Thomas Scherngell

Knowledge Spillovers across Europe. Evidence from a Poisson Spatial Interaction Model with Spatial Effects

Paper

Knowledge Spillovers across Europe.
Evidence from a Poisson Spatial Interaction Model with Spatial Effects

James P. LeSage*

McCoy Endowed Chair for Urban and Regional Economics
McCoy College of Business Administration
Department of Finance and Economics
Texas State University
San Marcos, Texas 78666
jlesage@spatial-econometrics.com

Manfred M. Fischer

Chair for Economic Geography
Institute for Economic Geography and GIScience
Vienna University of Economics
and Business Administration, A-1090 Vienna, Austria
manfred.fischer@wu-wien.ac.at

and

Thomas Scherngell

Institute for Economic Geography and GIScience
Vienna University of Economics
and Business Administration, A-1090 Vienna, Austria
thomas.scherngell@wu-wien.ac.at

16 June 2006

* Corresponding author

Abstract. By following the paper trail left by patent citations between high-technology patents in Europe we use a Bayesian hierarchical Poisson spatial interaction modelling approach to identify and measure spatial separation effects to interregional knowledge flows, as captured by patent citations. The model introduced here is novel in that it allows for spatially structured origin and destination effects for the regions. Estimation of the model is carried out within a Bayesian framework using data augmentation and Markov Chain Monte Carlo (MCMC) methods, related to recent work in Frühwirth-Schnatter and Wagner (2004). This allows MCMC sampling from well-known distribution families and, thus, provides a substantial improvement over MCMC estimation based on Metropolis-Hastings sampling from non-standard conditional distributions.

Estimation results from our model provides evidence that geography matters. First, geographical distance between origin and destination regions has a significant impact on knowledge spillovers, and this effect is substantial. Second, national border effects are important and dominate geographical distance effects. Third, the latent spatial effects exhibit weak spatial dependence. Not only geography, but also technological proximity matters. Interregional knowledge flows are industry specific and occur most often between regions located close to each other in technological space.

JEL Classification: C11, C13, C31, R15

Keywords: Origin-destination flows, spatially structured random effects, Bayesian Markov Chain Monte Carlo, knowledge spillovers, patent citations.

1 Introduction

The recent past has seen the development of a significant body of empirical research on knowledge spillovers. Generally speaking, this research has shown that new knowledge spills over (see Griliches 1992) and complements R&D in some industry, especially in high-technology industries (see Bernstein and Nadiri 1998). But we know very little about where spillovers go. The objective of this paper is to identify and measure those types of spatial separation that tend to impede the likelihood of knowledge spillovers between regions in Europe. In particular, we are interested in the question whether or not knowledge – as captured by patent citations – flows more easily within countries than between, and to what extent geographic distance between inventors has an influence on these knowledge flows. As we consider spatial separation effects to interregional spillovers in a multiregional setting it is important to control for technological proximity between regions as geographical distance could be just proxying for technological proximity.

We adopt the view that finds thinking in terms of a spatial interaction modelling perspective congenial and useful to investigate origin-destination knowledge flows as captured by high-technology patent citations in Europe. High-technology is defined to include the ISIC-sectors aerospace (ISIC 3845), electronics-telecommunication (ISIC 3825), computers and office equipment (ISIC 3842), and pharmaceuticals (ISIC 3522). The European coverage is given by patent applications at the European Patent Office (EPO) that are assigned to high-technology firms located in the EU-25 member states (except Cyprus and Malta), the two accession countries, Bulgaria and Romania, and Norway and Switzerland.

By following the paper trail left by patent citations between high-technology patents in Europe we use a Bayesian hierarchical Poisson spatial interaction model. The model is novel in that it allows for spatially structured origin and destination latent effects. A spatial autoregressive structure serves as a prior for these effects vectors, one for regions reflecting origins of the cited patents and another for the regions that cite patents. Posterior estimates of the origin and destination latent effects may be used to identify regions that exhibit positive and negative effects magnitudes since the effects parameters have a prior mean of zero. Positive and negative posterior effects estimates can be interpreted as measuring the magnitude and influence of latent unobservable factors on the knowledge flow process.

It is this model that distinguishes the current study from prior work by Fischer, Scherngell and Jansenberger (2006) which produces more conventional maximum likelihood estimates based on a heterogeneous Poisson spatial interaction model. Their 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 leads to a tractable negative binomial distribution maximum likelihood procedure (see Cameron and Trivedi 1998). The negative binomial distribution can be derived by assuming true contagion, allowing us to interpret the model in two quite different ways that are opposed to each other. This represents a serious drawback of the model that arises from relying on the conjugate gamma distribution for the Poisson.

The model introduced in this paper does not rely on the conjugate gamma prior, but rather on a normal prior for the random individual effects. Typically, there is no analytical expression for the unconditional density when using normally distributed random effects. Because of this, development of estimation methods for such cases is an active area of research (see, for example, Chib, Greenberg and Winkleman 1998). Drawing upon the contribution of Frühwirth-Schnatter and Wagner (2004), this paper contributes to this area of research by developing Gaussian random effects governed by a spatial autoregressive process that results in a Gibbs sampling scheme. By Gibbs sampler, we refer to the process where sequential sampling of all parameters in the model involves only draws from distributions having known forms.

The rest of the paper is organised as follows. Section 2 begins to set forth the context and framework for the discussion, and introduces the model proposed here. This model allows for latent regional effects parameters that take the form of a spatial autoregression. The spatial autoregressive (SAR) structure assumed to govern the origin and destination effects introduces additional sample data information in the form of an n by n spatial contiguity matrix that describes the spatial connectivity structure of the sample regions. This additional spatial structure in conjunction with the spatial autoregressive process assumption provides a parsimonious parameterisation of the regional effects parameters. This is in contrast to the typical assumption of a normal distribution with zero mean and constant scalar variance assigned as a prior for non-spatial latent effects parameters. Our approach of estimating two sets of n latent effects based on a sample of size $N = n^2$ also differs from the conventional approaches that estimate a latent effect parameter for all sample observations, which would be N in our case.

This model extends the class of Poisson spatial interaction models presented in Fischer, Scherngell and Jansenberger (2006) and relies on a hierarchical construct. We estimate the model using Markov Chain Monte Carlo (MCMC) methods and data augmentation schemes based on recent work by Frühwirth-Schnatter and Wagner (2004) to derive estimates by simulating draws from the complete set of conditional distributions for the parameters in the model. Section 3 briefly describes the data augmentation approach used and sets forth the conditional distributions for our model. Section 4 applies the methodology to the sample of high-technology patent citations from 188 European regions. Section 5 concludes the paper.

2 The Poisson Spatial Interaction Model with Spatial Effects

This section lays out the notation and conventions used in describing origin-destination flows and the modelling of these by the standard spatial interaction model (see 2.1) and then sets forth the Bayesian hierarchical structure that we suggest (see 2.2).

2.1 The Context

The spatial interaction modelling perspective shifts attention from the individual patent citations to interregional patent citations, or in other words from the dyad "cited patent – citing patent" to the dyad "cited region – citing region" within a spatial interaction system. Suppose that we have a spatial interaction system with n regions. Let Y represent the n-by-n square matrix of patent citation flows where the element Y_{ij} reflects patent citations originating in region i and cited by column region j. We therefore treat the columns as destinations of the patent citation flows and the rows as their origins. The n-by-n patent citation matrix can be vectorised into an $N = n^2$ vector that we label y which contains variation in patent citations flows across all origin-destination (OD) pairs.

A typical spatial interaction model directs attention to three types of functions to explaining the variation in the vector of OD-flows: an origin function, a destination function and a spatial separation function. There is a basic formal distinction implicit in the definitions of origins and destination functions on the one hand, and spatial separation functions on the other. Spatial separation functions are postulated to be explicit functions of numerical separation variables, while origin and destinations functions are formally only weights with origin and destination variables.

Observations on the origin and destination variables are typically organised in n-by-k variable matrices that we label X, containing k characteristics for each of the n regions. Given the origin-destination format of the vector y, where observations 1 to n reflect flows from origin 1 to all n destination regions, the matrix X would be repeated n times to produce an N-by-k matrix representing destination characteristics that we label X_d (see LeSage and Pace 2005). A second matrix can be formed to represent origin characteristics that we label X_o. This would repeat the characteristics of the first region n times to form the first n rows of X_o, the characteristics of the second region n times to for the next n rows of X_o and so on, resulting in an N-by-k matrix.

The spatial separation function constitutes the very core of spatial interaction models. Thus, a number of alternative specifications have been proposed. But the multivariate exponential function is most general (see Fischer and Reggiani 2004) and will be used in the context of this paper. Focus is laid on four distinct measures of separation: geographical distance measured in terms of the great circle distance between the regions' economic centres, a dummy variable that represents border effects measured in terms of the existence of country borders, a dummy variable that represents language barriers, and technological proximity between the regions. The corresponding data are typically summarised in form of n-by-n matrices that we label D_1, D_2, D_3 and D_4, respectively. Thus, $d_1 = \text{vec}(D_1)$, $d_2 = \text{vec}(D_2)$, $d_3 = \text{vec}(D_3)$ and $d_4 = \text{vec}(D_4)$ are N-by-1 vectors of these measures of separation from each origin region to each destination region formed by stacking the columns of the origin-destination matrices into a variable vector.

This results in the exponential spatial interaction model[1] which may be written in its equivalent log-additive version as

[1] A spatial interaction is termed exponential if the spatial separation function is specified as an exponential function.

$$y = \alpha\ \iota_N + X_d\ \beta_d + X_o\ \beta_o + \gamma_1\, d_1 + \gamma_2\, d_2 + \gamma_3\, d_3 + \gamma_4\, d_4 + \varepsilon \tag{1}$$

where ι_N is an N-by-1 vector of ones, X_d and X_o represent N-by-k matrices containing $k = k_1$ destination and $k = k_2$ origin characteristics, respectively. β_d and β_o are the associated k-by-1 parameters. d_1, d_2, d_3 and d_4 are N-by-1 vectors of spatial separation. The scalar parameters γ_1, γ_2, γ_3 and γ_4 reflect the effects of geographical distance, country borders, language barriers and technological proximity. α denotes the constant term parameter. The N-by-1 vector ε represents disturbances.

For notational convenience Equation (1) can be formally simplified by stacking the intercept term and the sample data into the vector y and matrix X as

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} 1 & x_{o1} & x_{d1} & d_{11} & d_{21} & d_{31} & d_{41} \\ 1 & x_{o2} & x_{d2} & d_{12} & d_{22} & d_{32} & d_{42} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{oN} & x_{dN} & d_{1N} & d_{2N} & d_{3N} & d_{4N} \end{pmatrix}. \tag{2}$$

This allows us to set forth the model given by Equation (1) as follows

$$y = X\ \beta + \varepsilon \tag{3}$$

where $\beta = (\alpha, \beta_o, \beta_d, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$ and X is the N-by-7 matrix including the origin, destination and spatial separation variables and the intercept term. Assuming that $E(\varepsilon) = 0$ and that the variables are measured without error, then model (3) may be estimated by means of ordinary least squares.

2.2 A Bayesian Poisson Extension of the Conventional Spatial Interaction Model

While the log-additive spatial interaction model given by Equation (3) can easily be estimated using standard least squares, it shows two major shortcomings. *First*, least squares and normality assumptions ignore the true integer nature of flows and approximate a discrete-valued by an almost certainly misrepresentative continuous distribution (Fischer and Reismann 2002). *Second*, models of type (3) neglect potential spatial dependencies of the origin-destination flows contained in the dependent variable vector y (LeSage and Pace 2005, Fischer, Reismann and Scherngell 2006). For example, neighbouring origins and destinations may exhibit estimation errors of similar magnitude if underlying latent or unobserved forces are at work or missing origin and destination variables exert a similar impact on neighbouring observations.

The following Bayesian hierarchical Poisson extension will overcome these deficiencies by assuming that the patent citation flows are independently distributed Poisson variates with finite means and by introducing two *n*-by-1 vectors of regional effects parameters, one for each region treated as an origin and another for destination regions. This model can be expressed as

$$y_i \mid \lambda_i \sim P(\lambda_i) \tag{4}$$

where λ_i is the conditional mean

$$\lambda_i = E(y_i \mid \beta, \theta, \phi) = \exp(x_i\,\beta + v_i\,\theta + w_i\,\phi) \tag{5}$$

which depends not only on the covariates with the associated parameter vector β, but also on *n*-by-1 vectors of latent regional effects parameters θ and ϕ, one for the regions treated as origin (i.e. θ) and another for destination regions (i.e. ϕ). The inclusion of these regional effects vectors allows for geographical differences or heterogeneity in the n origin and n destination regions. $v_i = (v_{i1},\ldots, v_{in})$ represents a vector that identifies region i as an origin and $w_i = (w_{i1},\ldots, w_{in})$ identifies destination regions. Given our configuration for the flow matrix of patent citations with columns as origins and rows as destinations, we could form a matrix $V = \iota_n \otimes I_n$ and $W = I_n \otimes \iota_n$ such that v_i and w_i represent the ith row of these mutually exclusive N-by-n matrices. I_n denotes the n-square identity matrix.

As with all Bayesian models, we begin by postulating suitable prior distributions for all parameters (β, θ, ϕ), and then derive the corresponding conditional posterior distributions given the observed data in the next section. We use a normal prior distribution for $\beta = (\alpha, \beta_o, \beta_d, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$ associated with the covariates in the explanatory variables matrix X centered on zero with a large standard deviation:

$$\pi(\beta\,|\,\psi) \sim N_h[0,T]\,. \tag{6}$$

The normal prior distribution is allowed to be indexed by an unknown hyperparameter that we have labelled ψ, and h denotes the number of explanatory (i.e. origin, destination and spatial separation) variables in the matrix X which includes a constant term. $T = w^2 I_q$, for some sufficiently large w, such as w = 100. We use $N_h[\mu, \Sigma]$ to represent an h-variate normal distribution with mean μ and variance-covariance Σ.

For the spatial effects parameters we rely on *spatial autoregressive* (SAR) *priors*:

$$\theta = \rho_o\, C\, \theta + u_o \qquad u_o \sim N_n[0, \sigma_o^2\, I_n] \tag{7}$$

$$\phi = \rho_d\, C\, \phi + u_d \qquad u_d \sim N_n[0, \sigma_d^2\, I_n] \tag{8}$$

where C is an n-by-n row standardised first order spatial contiguity matrix. This matrix reflects the spatial configuration of the regions in terms of common borders, with row sums of unity by virtue of row standardisation. We assign an inverse gamma (IG) prior for the parameters σ_o^2, σ_d^2, taking the form:

$$\pi(\sigma_o^{-2}), \pi(\sigma_d^{-2}) \;\sim\; G(g_1, g_2) = G(0.01,\, 0.01)\,. \tag{9}$$

It is frequently noted that flat or improper priors on variance parameters in hierarchical modeling can lead to (almost) improper posterior distributions (see Gelman et al. 1995, chapter 5). This prior implies a mean of unity, and a variance of 100. In the absence of prior information, it seems reasonable to rely on the same prior for both σ_o^2 and σ_d^2. The spatial dependence parameters are known to lie in the stationary interval: $[\kappa_{\max}^{-1}, \kappa_{\min}^{-1}]$ with $\kappa_{\min} < 0$, $\kappa_{\max} > 0$ denoting the minimum and maximum eigenvalues of the matrix C (see, for example, Lemma 2 in Sun et al. 1999). We rely on a uniform distribution over this interval as our prior for ρ_o, ρ_d, that is

$$\pi(\rho_o), \pi(\rho_d) \sim U[\kappa_{\max}^{-1}, \kappa_{\min}^{-1}] \propto 1 \tag{10}$$

Solving for θ and ϕ in terms of u_o and u_d suggests a normal prior for the origin and destination spatial effects vectors taking the form:

$$\theta \mid \rho_o, \sigma_o^2, \psi \sim N_n[0, \sigma_o^2 (B_o^T B_o)^{-1}], \tag{11}$$

$$\phi \mid \rho_d, \sigma_d^2, \psi \sim N_n[0, \sigma_d^2 (B_d^T B_d)^{-1}], \tag{12}$$

$$B_o = (I_n - \rho_o C), \tag{13}$$

$$B_d = (I_n - \rho_d C). \tag{14}$$

We note that B_o, B_d are non-singular for a conventional row-normalised first-order spatial contiguity matrix C and the spatial dependence parameters ρ_o , ρ_d in the interval $[\kappa_{\max}^{-1}, \kappa_{\min}^{-1}]$. This leads to a proper prior distribution in contrast to the well-known intrinsic conditional autoregressive (CAR) prior introduced by Besag and Kooperberg (1995).

When $\rho_o = \rho_d = 0$, our model collapses to the special case of a normal prior for the random effects vectors with means of zero for both effects and constant scalar variances σ_o^2 and σ_d^2, so our SAR prior specification subsumes this as a special case. It should be noted that estimates for these two sets of random effects parameters are identified, since a set of n mutually exclusive sample data observations are aggregated through the vectors v_i and w_i to produce each estimate θ_i , ϕ_i in the vector of parameters θ and ϕ.

3 Estimating the Model

Estimation will be achieved via Markov Chain Monte Carlo (MCMC) methods that sample sequentially from the complete set of conditional distributions for the parameters. To implement the MCMC sampling approach we need to derive the complete conditional posterior distributions for all model parameters. The model assumptions made are sufficient to derive these distributions, but the resulting posterior density does not belong to a density from a well-known distribution family. Following recent work by Frühwirth-Schnatter and Wagner (2004) Section 3.1 shows that the introduction of two sequences of artificially missing data with a data augmentation scheme eliminates both non-normality and non-linearity of the mean λ_i in the parameters and allows MCMC sampling from conditional distributions for the parameters that belong to standard distribution families. Section 3.2 introduces the conditional posterior distributions that form the basis of a Gibbs sampler for our Poisson spatial interaction model with regionally structured random effects parameters θ, ϕ.

3.1 The Data Augmentation Approach

For notational convenience in the following discussion we collect the parameters β, θ, ϕ in a vector δ, and the spatial hyperparameters ρ_o, σ_o^2, ρ_d, σ_d^2 in the vector ς, so that all parameters can be placed in a vector $\upsilon = (\delta, \varsigma) = (\beta, \theta, \phi, \rho_o, \rho_o, \sigma_o^2, \sigma_d^2)$. Furthermore we restate Equation (5) as

$$\lambda_i = \exp(z_i\, \delta) \tag{15}$$

with

$$z_i = (\ 1 \ \ x_i \ \ \ v_i \ \ w_i \ ), \tag{16}$$

$$\delta = (\ \beta, \theta, \phi\ )^T . \tag{17}$$

Then the posterior density takes the form:

$$p(\upsilon \mid \psi, y) \propto p(y \mid \upsilon) p(\upsilon \mid \psi) \tag{16}$$

$$p(y \mid \upsilon) = \prod_{i=1}^{N} \frac{\exp(z_i\, \delta)^{y_i}}{y_i\, !} \exp(-\exp(z_i\, \delta)) . \tag{17}$$

The use of a normal distribution for the random effects in place of the conjugate gamma distribution results in a posterior density that does not belong to a density from a known distribution family. Conventional MCMC methods to sample from the posterior have involved Metropolis-Hastings algorithms (see, for example, Chib et al. 1998). The contribution of Frühwirth-Schnatter and Wagner (2004) was to note that through the introduction of two sequences of artificially missing data treated using data augmentation can lead to a sequence of conditional posteriors for the parameters υ that take the same form as those that would arise if our model was a normal linear model. One of the two sequences of artificially missing data eliminates the non-linearity of the Poisson spatial interaction model and the second eliminates the non-normality of the error term.

Eliminating the non-linearity of the Poisson spatial interaction model arises from the insight that the distribution of $y_i \mid \lambda_i$ may be regarded for each observation i as the distribution of the number of jumps of an unobserved Poisson process with intensity λ_i occurring in the time interval [0,1]. This leads Frühwirth-Schnatter and Wagner (2004) to suggest a data augmentation step where for each $i = 1, \ldots, N$ inter-arrival times of the unobserved Poisson process are introduced as missing data. We label these τ_{ij}, $j = 1, \ldots, (y_i + 1)$, and note that these are known to follow an exponential, $\xi(\lambda_i)$, distribution:

$$\tau_{ij} \mid \upsilon \sim \xi(1) / \lambda_i \, . \qquad (19)$$

Using $\lambda_i = \exp(z_i\,\delta)$, Equation (19) may be reformulated as a linear model given by

$$\log \tau_{ij} \mid \upsilon \; = - z_i\,\delta + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim \log(\xi(1)) \qquad (20)$$

thereby eliminating the non-linearity, but leaving us with the non-normal error term. An important point is that the full conditional posterior for the parameters after the introduction of the inter-arrival times $\tau = \left\{ \tau_{ij}, j = 1, \ldots, (y_i + 1), \; i = 1, \ldots, N \right\}$ is independent of y, that is, $p(\upsilon \mid \psi, \tau, y) = p(\upsilon \mid \psi, \tau)$.

Given the linear model in (20) after conditioning on τ, Frühwirth-Schnatter and Wagner (2004) propose eliminating the non-linearity in the disturbance term ε_{ij} through the use of a normal mixture of five components with parameters m_r and s_r for the $r = 1, \ldots, 5$ components. This results in a conditionally Gaussian model, and is similar to approaches taken by Kim et al. (1998) and Chib et al. (2002) for the case of stochastic volatility models, where the normal mixture approximation was applied to a log χ^2-distribution. Formally we get:

$$p(\varepsilon_{ij}) = \exp(\varepsilon_{ij} - \exp(\varepsilon_{ij})) \;\approx\; \sum_{r=1}^{5} t_r \;\; f_N(\varepsilon_{ij}; m_r, s_r^2) \tag{21}$$

where values for the parameters (t_r, m_r, s_r) are provided in Table 1 of Frühwirth-Schnatter and Wagner (2004) for r = 1, …, 5. This results in a second data augmentation step where latent component indicator variables r_{ij} taking values 1, …, 5 are introduced for each observation i and arrival times $j = 1,..., y_i + 1$.

At this point, the original Poisson model from (4) reduces to a normal linear model with heteroscedastic errors having known variances:

$$\log \tau_{ij} \mid \upsilon, r_{ij} \;\; = -z_i\,\delta + m_{r_{ij}} + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N_1(0, s_{r_{ij}}^2)\,. \tag{22}$$

Frühwirth-Schnatter and Wagner (2004) provide details on how to sample the missing data τ_{ij} and r_{ij} component indicators for the normal mixture. For the parameters τ_{ij} this involves sampling at set of j order statistics for each observation i, where j = 1, …, y_i from a uniform (0,1) distribution, and a final arrival time component from an exponential (λ_i) density. The parameters r_{ij} require sampling from a five-component discrete density for each i = 1, …, N, j = 1, …, y_i+1.

3.2 The Conditional Posterior Distributions

MCMC estimation of the model requires sampling from the complete sequence of conditional distributions for the parameters. We use the normal linear model from (22) as the starting point to introduce the conditional posterior distributions that form the basis of a Gibbs sampler for our Poisson spatial interaction model with spatially structured random effects parameters θ, ϕ.

For notational convenience, we collect the component indicators into a single set: $R = \{r_{ij},\ i = 1, \ldots, N, j = 1, \ldots, y_i+1\}$. The basic scheme of the algorithm for Gibbs sampling can be viewed in terms of the following steps:

(i) Sample the parameters $\delta = (\beta, \theta, \phi)$ given τ_{ij}, R, ς

(ii) Sample the spatial hyperparameters ς given δ

(iii) Sample the inter-arrival times τ_{ij} given δ, y

(iv) Sample the component indicators r_{ij} given τ, δ.

Step (i) can be expressed as a sample from a multivariate normal distribution for δ, but for computational efficiency we adopt a component-wise multi-move approach to sampling the β parameters from a multivariate normal distribution and the spatial parameters θ from univariate conditional posteriors for each element θ_i conditional on all other elements which we denote θ_{-i}, and similarly for the elements of the vector ϕ. We provide details in Appendix A for the derivation of these univariate conditional distributions. Step (ii) involves one caveat to the strict notion of a Gibbs sampler, namely that we rely on a griddy gibbs step to sample the spatial dependence parameters ρ_o and ρ_d which involves univariate numerical integration of the conditional posteriors for these two parameters, and a draw carried out through inversion. However, unlike the case of Metropolis-Hastings steps, there is no requirement of a proposal density or tuning, and this procedure results in a draw for these parameters on every pass through the sampler. The random effects variance parameters from step (ii) take conventional inverted gamma distribution forms allowing straightforward Gibbs sampling. Steps (iii) and (iv) are elaborated in Appendix B of this paper (see also Frühwirth-Schnatter and Wagner 2004).

For notational convenience, we let $n_i = y_i+1$ and define $\tilde{y}_i$ as the n_i-by-1 vectors in Equation (23):

$$\tilde{y}_i = \begin{pmatrix} \log(\tau_{i,1}) - m_{r_{i,1}} \\ \vdots \\ \log(\tau_{i,n_i}) - m_{r_{i,n_i}} \end{pmatrix}. \tag{23}$$

We define the n_i-by-k matrix $\tilde{x}_i$ as $\iota_{n_i} x_i$, the 1-by-k vector x_i repeated n_i times for each observation i, and we stack these into the vector y and matrix X as:

$$y = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_N \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_N \end{pmatrix}. \tag{24}$$

Similarly, we define the n_i-by-n matrix $\tilde{v}_i$ as $\iota_{n_i} v_i$, the 1-by-n vector v_i repeated n_i times for each observation i, and the n_i-by-n matrix $\tilde{w}_i$ as $\iota_{n_i} w_i$, the 1-by-n vector w_i repeated n_i times for each observation i. Stacking these produces the matrices:

$$V = \begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \\ \vdots \\ \tilde{v}_N \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} \tilde{w}_1 \\ \tilde{w}_2 \\ \vdots \\ \tilde{w}_N \end{pmatrix}. \tag{25}$$

Finally we let $\tilde{\Omega}_i$ represent a diagonal matrix containing $s^2_{i,r_{ij}}$, $j = 1, \ldots, n_i$ on the diagonal, and define $m = \sum_{i=1}^{N} n_i$ the patent citation counts cumulated over all observations. We place the diagonal matrices $\tilde{\Omega}_i$ on the diagonal of an m-by-m matrix $\Omega = \mathrm{diag}(\tilde{\Omega}_1, \tilde{\Omega}_2, \ldots, \tilde{\Omega}_N)$. This allows us to set forth our model in matrix notation, where using (22) we have:

$$y \mid \upsilon, R \;\sim\; N_m(-X\beta - V\theta - W\phi, \Omega). \tag{26}$$

$$\theta \;=\; \rho_o C\theta + u_o \tag{27}$$

$$\phi \;=\; \rho_d C\phi + u_d \tag{28}$$

$$\pi(\theta \mid \rho_o, \sigma_o^2) \;\sim\; (\sigma_o^2)^{n/2} \mid B_o \mid \exp(-\tfrac{1}{2\sigma_o^2}\theta^T B_o^T B_o \theta) \tag{29}$$

$$\pi(\phi \mid \rho_d, \sigma_d^2) \;\sim\; (\sigma_d^2)^{n/2} \mid B_d \mid \exp(-\tfrac{1}{2\sigma_d^2}\phi^T B_d^T B_d \phi) \tag{30}$$

where | . | denotes the determinant. Equations (29) and (30) reflect the implied priors for the spatial effects vector θ conditional on ρ_o, σ_o^2 and that for ϕ conditional on ρ_d, σ_d^2.

Given the assumed prior independence of $\beta, \rho_o, \rho_d, \sigma_o^2, \sigma_d^2$, we have the following joint posterior density for β:

$$
\begin{aligned}
p(\beta \mid \theta, \phi, \rho_o, \rho_d, \tau, R) &\propto \pi(\beta) \exp\{-\tfrac{1}{2}(y + X\beta + V\theta + W\phi)^T \Omega^{-1}(y + X\beta + V\theta + W\phi)\} \\
&\propto \exp\{-\tfrac{1}{2}(y + X\beta + V\theta + W\phi)^T \Omega^{-1}(y + X\beta + V\theta + W\phi)\} \\
&\quad \exp\{-\tfrac{1}{2}\beta^T T^{-1}\beta\}.
\end{aligned}
\tag{31}
$$

In Appendix A we show that this results in a multivariate normal conditional posterior distribution for β taking the form:

$$
\beta \mid \theta, \phi, \rho_o, \rho_d, \sigma_o^2, \sigma_d^2, \Omega, y \;\sim\; N_h[\Sigma_\beta^{-1}\mu_\beta, \Sigma_\beta^{-1}] \tag{32}
$$

$$
\mu_\beta = X^T \Omega^{-1}(y + V\theta + W\phi) \tag{33}
$$

$$
\Sigma_\beta = (X^T \Omega^{-1} X + T^{-1}). \tag{34}
$$

Taking a similar approach to that for β, we have a joint posterior density for θ of the form:

$$
\begin{aligned}
p(\theta \mid \beta, \phi, \rho_o, \rho_d, \tau, R) &\propto \pi(\theta \mid \rho_o, \sigma_o^2) \\
&\quad \exp\{-\tfrac{1}{2}[V\theta - (y + X\beta + W\phi)]^T \Omega^{-1}[V\theta - (y + X\beta + W\phi)]\} \\
&\propto \exp\{-\tfrac{1}{2}[V\theta - (y + X\beta + W\phi)]^T \Omega^{-1}[V\theta - (y + X\beta + W\phi)]^T\} \\
&\quad \exp\{-\tfrac{1}{2\sigma_o^2}\theta^T B_o^T B_o \theta\}
\end{aligned}
\tag{35}
$$

which we show in Appendix A leads to a multivariate normal as the conditional posterior distribution for θ:

$$
\theta \mid \beta, \phi, \rho_o, \rho_d, \sigma_o^2, \sigma_d^2, \Omega, y \;\sim\; N_n[\Sigma_\theta^{-1}\mu_\theta, \Sigma_\theta^{-1}] \tag{36}
$$

$$
\mu_\theta = V^T \Omega^{-1}(y + X\beta + W\phi) \tag{37}
$$

$$
\Sigma_\theta = (\tfrac{1}{\sigma_o^2} B_o^T B_o + V^T \Omega^{-1} V), \tag{38}
$$

and similarly for the spatial effects vector ϕ we have:

$$\phi \mid \beta, \theta, \rho_o, \rho_d, \sigma_o^2, \sigma_d^2, \Omega, y \;\sim\; N_n[\Sigma_\phi^{-1}\mu_\phi, \Sigma_\phi^{-1}] \tag{39}$$

$$\mu_\phi \;=\; W^T \Omega^{-1}(y + X\beta + V\theta) \tag{40}$$

$$\Sigma_\phi \;=\; (\tfrac{1}{\sigma_d^2} B_d^T B_d + W^T \Omega^{-1} W)\,. \tag{41}$$

However, evaluation of these expressions involves inversion of matrices of size *n*-by-*n* on each pass of the MCMC sampler, which is problematical for cases where the number of regions *n* is large. A computationally efficient alternative is to sample from univariate normal conditional distribution for each parameter θ_i conditional on all other θ_{-i} where $\theta_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_n)$.

The joint posterior distributions for ρ_o, ρ_d take the forms:

$$\begin{aligned} p(\rho_o \mid \beta, \theta, \phi, \sigma_o^2, \sigma_d^2, \rho_d, \Omega, y) &\propto \pi(\theta \mid \rho_o, \sigma_o^2)\pi(\rho_o) \\ &\propto |B_o| \;\; \exp(-\tfrac{1}{2\sigma_o^2}\theta^T B_o^T B_o \theta) \end{aligned} \tag{42}$$

$$\begin{aligned} p(\rho_d \mid \beta, \theta, \phi, \sigma_o^2, \sigma_d^2, \rho_o, \Omega, y) &\propto \pi(\theta \mid \rho_d, \sigma_d^2)\pi(\rho_d) \\ &\propto |B_d| \;\; \exp(-\tfrac{1}{2\sigma_d^2}\theta^T B_d^T B_d \theta) \end{aligned} \tag{43}$$

which as noted in Smith and LeSage (2004) is not reducible to a standard distribution. We rely on the univariate numerical integration approach described in Smith and LeSage (2004) to sample from these two conditional posterior distributions. Specifically, we rely on a vectorised expression: $\theta^T B_o^T B_o \theta = \theta^T \theta - 2\rho_o \theta^T C \theta + \rho_o^2 \theta^T C^T C \theta$; for the conditional posterior computed over a grid of *h* values for ρ_o in the interval $[\kappa_{\min}^{-1}, \kappa_{\max}^{-1}]$. For the determinant term $|B_o| = |I_n - \rho_o C|$, we use the sparse matrix methods of Barry and Pace (1997) to compute and store tabled values for the determinant over this same grid of *h* values for ρ_o. This is done prior to beginning the MCMC sampling loop. Having expressed the conditional distribution over a grid of ρ_d values, we use univariate numerical integration to find the normalising constant, and then draw from this numerical approximation to the conditional posterior using inversion.

We note that Metropolis-Hastings sampling from these conditional posterior distributions would also benefit from this approach to vectorising the distribution over a grid of values for

ρ_o, allowing rapid evaluation of candidate values during sampling. However, there is the need to rely on a proposal distribution along with tuning, and the Smith and LeSage (2004) procedure produces a draw on every pass through the MCMC sampler.

The joint posterior densities for σ_o^2, σ_d^2 take the form:

$$\begin{aligned} p(\sigma_o^2 \mid \beta, \theta, \phi, \rho_o, \rho_d, \Omega, y) &\propto \pi(\theta \mid \rho_o, \sigma_o^2)\pi(\sigma_o^2) \\ &\propto (\sigma_o^2)^{-\frac{n}{2}} \exp(-\tfrac{1}{2\sigma_o^2}\theta^T B_o^T B_o \theta)\ (\sigma_o^2)^{-\frac{n}{2}+g_1-1} \end{aligned} \tag{44}$$

which results in an inverse gamma distribution for the conditional posterior. A similar result applies to σ_d^2, with details again provided in Appendix A:

$$\sigma_o^2 \mid \beta, \theta, \phi, \rho_o, \rho_d, \sigma_d^2, \Omega \sim IG(a,b) \tag{45}$$

$$a = \tfrac{n}{2} + g_1 \tag{46}$$

$$b = \theta^T B_o^T B_o \theta + 2g_2 \tag{47}$$

$$\sigma_d^2 \mid \beta, \theta, \phi, \rho_o, \rho_d, \sigma_o^2, \Omega \sim IG(c,d) \tag{48}$$

$$c = (n/2) + g_1 \tag{49}$$

$$d = \phi^T B_o^T B_o \phi + 2g_2 \tag{50}$$

4 The Application

In this section we briefly describe the data used (see 3.1) and present the main empirical findings in 3.2. These estimates are based on the model and estimation methodology described in the previous sections.

4.1 The Patent Citation Data

Our main data source is the European Patent Office (EPO) database. This is a natural choice for the purpose of our study because patents from different national patent offices are not comparable with each other. There are different patenting costs, approval requirements, citations practices, and enforcement rules across Europe. The focus is on corporate patents in the high-technology sector. High-technology is defined as the ISIC-sectors aerospace (ISIC 3845), electronics-telecommunication (ISIC 3832), computers and office equipment (ISIC 3825), and pharmaceuticals (ISIC 3522). We used MERIT's concordance table (Verspagen, Moergastel and Slabbers 1994) between these ISIC-sectors and the 628 patent subclasses of the International Patent Classification (IPC) to identify such patents from the universe of European patents.

Our patent data set includes all the high-technology patents with an application date in the years 1985-2002, totalling 177,424 patents. Data on the inventor his/her location, the assignee (i.e. the legal entity that owns the patent), the time of application, the technology of the invention as captured by IPC codes, and EPO patent citations are the main pieces of information from this file. We selected patents assigned to non-government organisations located in Europe, as our interest is on interfirm knowledge spillovers.

Patent citations is a phenomenon that derives from the relationship between two inventions or inventors as evidenced by a citing patent and cited patent. The data on this relationship come in the form of citations *made* (i.e., each patent lists references to previous patents). To identify the citation flows we need a list of cited and citing patent applications. This requires in fact access to all citation data in a way that permits efficient search and extraction of citations not by the patent number of the citing patent but by the patent number of the cited patent.

The observation of citations is subject to a truncation bias because we observe citations for only a portion of the life of an invention, with the duration of that portion varying across patent cohorts. This means that patents of different ages are subject to different degrees of truncation. To overcome this problem we have identified all the pairs of cited and citing patents where citations to a patent are counted for a window of five years following its issuance. The analysis is, thus, confined to 1985-1997 in the case of cited patents while citing patents appearing in 1990-2002 are taken into account. Although the five-year horizon appears to be short, it does capture a significant amount of a typical patent's citation life. Note that the mean citation lag of all high-technology patent citations in 1985-2002 is 4.62 years.

Given our interest on pure externalities (i.e., on interfirm knowledge spillovers), citations to patents that belong to the same assignee (so-called self-citations) were eliminated, resulting into 98,191 interfirm patent citations. The elimination of self-citations remains far from satisfactory. Although we have checked the sample for cases where company names are sufficiently similar to identify self-citations between parents and their subsidiaries, and joint ventures, this in effect can only get us so far.

The spatial interaction modelling perspective shifts attention from individual patent citations to interregional patent citations or from the dyad "cited patent – citing patent" to the dyad "cited region – citing region". We have chosen n = 188 regions, generally NUTS-2 regions for the EU-15 countries and NUTS-0 regions for the other countries. NUTS is an acronym of the French for "nomenclature of the territorial units for statistic", which is a hierarchical system of regions used by the statistical office of the European community for the production of regional statistics. At the top of the hierarchy are the NUTS-0 regions (countries), below which are NUTS-1 regions (regions within countries) and then NUTS-2 regions (subdivisions of NUTS-1 regions). In the case of cross-regional inventor teams we have used the procedure of multiple full counting which rather than fractional counting does justice to the true integer nature of patent citations, but gives the interregional cooperative inventions greater weight.

Table 1: Descriptive Statistics on the (188, 188)-Patent Citation Matrix

		Patent Citations				
	Number of Matrix Elements*	Number	Mean	Standard Deviation	Min.	Max.
All Elements	35,344	98,191	2.77	16.23	0	1,408
Intraregional Links	188	11,371	60.48	152.05	0	1,408
Interregional Links	35,156	86,820	2.46	11.14	0	351
Positive Interregional Links	11,468	86,820	7.57	18.49	1	351
National Interregional Links	3,952	25,341	6.41	20.02	0	351
International Interregional Links	31,204	61,479	1.97	9.31	0	290

* Elements of the n-by-n citation matrix

Table 1 provides some basic information about the n-by-n citation matrix (n = 188). The matrix contains $N = n^2 = 35{,}344$ elements with a total of 98,191 citations between high-technology firms. The mean number of citations between any two regions (including intraregional flows) is 2.77, but the standard deviation is rather high. Interregional citations show a highly skewed distribution. About two thirds of all pairs of regions (23,688 pairs) never cite each other's patents. The frequency of patent citations gradually declines for more intensive citation links. There are only 90 pairs of regions for which the number of citations is 100 or more. The average number of citations for all interregional pairs is 2.46 and the average for those that cite each other 7.57. Table 1 indicates, moreover, that national patent citations are more frequent than international ones.

4.2 Empirical Results

The explanatory variables matrix contained the origin variable measured in terms of the log number of patents in the knowledge producing region in the time period 1985-1997, the destination variable measured in terms of the log number of patents in the knowledge absorbing region in the time period 1990-2002, and four separation variables: geographical distance measured in terms of the great circle distance (in km) between the regions' economic centres, technological proximity between the origin and destination regions in terms of an index developed by Maurseth and Verspagen (2002), a border dummy variable measured in terms of the existence of country borders between the origin and destination regions and a language dummy variable to reflect whether origin and destination regions have different languages.

The Bayesian coefficient estimates for these explanatory variables' highest posterior density intervals are reported in Table 2, alongside the Maximum Likelihood (ML) Poisson estimates and associated *p*-values and *t*-statistics from the study by Fischer, Scherngell and Jansenberger (2006). Examining the estimates associated with the covariates in the model we see that all estimates are significantly different from the Maximum Likelihood Poisson spatial interaction estimates using the 0.05 and 0.95 Bayesian posterior density intervals.

We observe a lower elasticity of response of the flows to the origin and destination variables in the Bayesian model, with estimates around 0.76 and 0.83 versus 0.92 and 0.89 in the conventional Poisson spatial interaction model. Intuitively, the introduction of individual effects for the origins and destinations results in a model that places less reliance on the origin and destination variables. This seems plausible as the Bayesian effects model allows for spatial heterogeneity.

Table 2: Results from Bayesian and Maximum Likelihood Estimates

Variables	Bayesian MCMC			Maximum Likelihood	
	Lower 0.05 credible interval	Posterior mean	Upper 0.95 credible interval	Estimate (*p*-value)	*t*-statistic
Origin	0.7471	0.7618	0.7870	0.915 (0.000)	138.24
Destination	0.8134	0.8338	0.8516	0.885 (0.000)	140.95
Distance	-0.1993	-0.1952	-0.1912	-0.321 (0.000)	-22.88
Country border	-0.5320	-0.5125	-0.4925	-0.533 (0.000)	-11.50
Language barrier	-0.0056	0.0138	0.0337	-0.031 (0.000)	-0.72
Technological proximity	1.9129	1.9483	1.9851	1.219 (0.000)	9.32
Intercept	-9.2127	-8.9346	-8.7817	-10.881 (0.000)	-87.61
ρ_o	-0.0080	0.1786	0.3530		
ρ_d	-0.0560	0.1337	0.3210		
σ_o^2	0.4502	0.5489	0.6643		
σ_d^2	0.3918	0.4774	0.5863		

Notes: The origin variable is measured in terms of the log number of patents (1985-1997) in the knowledge producing region *i*, the destination variable in terms of the log number of patents (1990-2002) in the knowledge absorbing region *j*. Geographic distance is measured in terms of the great circle distance [in km] between the economic centres of the regions *i* and *j*, country border effects in terms of the existence of country borders between regions *i* and *j*, language barriers in terms of the existence of different languages in the regions *i* and *j*, and technological distance in terms of the technological proximity index developed by Maurseth and Verspagen (2002). The Maximum Likelihood estimate of dispersion is 0.725 and highly significant (see Fischer, Scherngell and Jansenberger 2006) which points to the presence of overdispersion that is modelled by the Bayesian Poisson spatial interaction model with spatial effects.

Distance exerts a negative impact on knowledge flows as we would expect, but this is less important in the Bayesian effects model than in the conventional Poisson model. Here too, this seems a plausible result as the introduction of spatially structured origin and destination effects should reduce the importance played by geographical distance in the non-spatial Poisson model. Borders also exert a negative impact on knowledge flows and we see a Bayesian estimate that is close to the non-Bayesian parameter magnitude. The border effect is 2.5 times as large as the distance effect. High-technology related knowledge tends to flow much more easily within countries than between countries.

The introduction of the spatial effects for origins and destinations makes language barriers not significantly different from zero. Technological similarity between the regions enhances knowledge flows as we would expect and the Bayesian estimate is about 1.5 times that from the conventional Poisson model, and about ten times larger then the distance effect. This indicates that knowledge flows are industry specific and occur most often between regions that are located close to each other in technological space. Technological proximity matters more than geographical proximity. By way of summary, the findings correspond well to the findings of the previous study, but the sensitivity of knowledge flows to the covariates is more pronounced in the Bayesian random effects model than in the standard Poisson model.

Turning attention to the parameters associated with the spatially structured origin and destination effects, we note that the origin ρ estimate (the spatial dependence parameter ρ_o) is (very nearly) different from zero using a 0.05 level, but the destination ρ is not significantly different from zero. It is important to keep in mind, that, when ρ is zero, we still have normally distributed random effects that account for heterogeneity in the sample. ρ_d equal to zero simply indicates that latent unobservable effects creating heterogeneity surrounding the destinations do not exhibit a spatial dependence character. That is, they do not necessarily look like those of the neighbours to the destinations. With regard to the origin, the patterns of heterogeneity captured by the spatially structured effects do exhibit weak spatial dependence. This means that the magnitudes of the effects tend to be similar to those from neighbouring regions. We might interpret this to mean that similar latent unobservable forces are at work, or that our model does not have covariates to explain these forces at work. The estimate for the variance of the origin effects is around 20 percent larger than that for the variance of the destination effects, suggesting more volatility in the origin effects than those assigned to the destinations.

5 Summary and Conclusions

A Bayesian hierarchical Poisson spatial interaction model that includes latent spatial effects structured to follow a spatial autoregressive process was introduced here to investigate knowledge spillovers across Europe, as captured by patent citations. The model deals with overdispersion arising from omitted origin and destination variables using structured regional or spatial effects.

Individual effects estimates are notoriously difficult to estimate with precision in conventional hierarchical linear models (see, for example, Gelfand, Sahu and Carlin 1995, Christensen, Roberts and Sköld 2005). Our approach to structuring two sets of regional/spatial effects parameters overcomes these problems in two ways. *First*, the spatial autoregressive structure placed on the latent effects parameters for the origin depend on one hyperparameter measuring the strength of spatial dependence and another representing a scalar variance parameter. These two parameters are introduced in the context of a sample of $n^2 = N$ observations, where $n = 188$ regions and N represents the sample of origin-destination pairs that arise from vectorising the origin-destination flow matrix. The $N = n^2$ sample size arises from vectorising an n-by-n origin-destination flow matrix Y, where the rows of the matrix Y reflect counts of cited patents (origins) and the columns reflect counts from regions citing the patents (destinations). We estimate only n latent regional origin effects parameters, one for each region treated as an origin, allowing us to rely on n sample data observations for *each* of the $i = 1, \ldots, n$ origin effect parameter estimates. In fact, since the n origin effects parameters are derived from the two hyperparameters that completely determine the spatial autoregressive process assigned to govern these processes, we could view this as relying on a sample of N observations to estimate the two parameters. A similar situation holds for the case of the n destination effects estimates for the patent citing regions.

Second, the spatial autoregressive structure assumed to govern the origin and destination effects introduces additional sample data information in the form of an n-by-n spatial contiguity matrix that describes the spatial connectivity structure of the sample regions. This additional spatial structure in conjunction with the spatial autoregressive process assumption provides a parsimonious parameterisation of the regional effects parameters. This is in contrast to the typical assumption of a normal distribution with zero mean and constant scalar variance assigned as a prior for non-spatial latent effects parameters. Our approach of estimating two sets of n latent effects based on a sample of size $N = n^2$ also differs from the conventional approaches that estimate a latent effect parameter for all sample observations, which would be N in our case.

Estimation of the model is via MCMC sampling based on data augmentation schemes. The results provide evidence that knowledge spillovers are geographically localised. National borders have a negative impact on knowledge flows, and this effect is very substantial. Knowledge flows are larger within countries than between regions located in different countries. The results also indicate that geographical proximity matters, while also suggesting that these effects are much smaller than the border effects. Knowledge spillovers occur most often between origin-destination regions that belong to the same country and are in geographical proximity. With regard to the origins, the patterns of heterogeneity captured by spatially structured effects do exhibit weak spatial dependence. This means that effects tend to be similar in size to those from neighbouring regions. This indicates that similar latent unobservable forces are at work or that our model does not include origin variables to explain these forces at work. Geography matters, but technological proximity tends to overcome geographical proximity. Interregional knowledge flows seem to follow particular technological trajectories, and occur most often between regions that are located in technological space not too far from each other.

Acknowledgement Two authors gratefully acknowledge the grant no. 19282 provided by the Austrian Science Fund.

Appendix A: Details regarding the MCMC sampler

First we show that the conditional posterior for β takes the multivariate from (see Equation (31) in the running text):

$$\begin{aligned} p(\beta \mid \theta, \phi, \rho_o, \rho_d, \tau, R) &\propto \pi(\beta)\exp\{-\tfrac{1}{2}(y + X\beta + V\theta + W\phi)^T \Omega^{-1}(y + X\beta + V\theta + W\phi)\} \\ &\propto \exp\{-\tfrac{1}{2}(y + X\beta + V\theta + W\phi)^T \Omega^{-1}(y + X\beta + V\theta + W\phi)\} \\ &\quad \exp\{-\tfrac{1}{2}\beta^T T^{-1}\beta\} \qquad \text{(A1)} \\ &\propto \exp(-\tfrac{1}{2}[\beta^T (X^T\Omega^{-1}X + T^{-1})\beta - 2X^T\Omega^{-1}(y + V\theta + W\phi)^T \beta) \\ &\propto \exp(-\tfrac{1}{2}[\beta - \Sigma_\beta^{-1}\mu_\beta]^T \Sigma_\beta [\beta - \Sigma_\beta^{-1}\mu_\beta]) \end{aligned}$$

where (see Equations (33) and (34) in the running text)

$$\mu_\beta = X^T\Omega^{-1}(y + V\theta + W\phi) \qquad \text{(A2)}$$

$$\Sigma_\beta = (X^T\Omega^{-1}X + T^{-1})\,. \qquad \text{(A3)}$$

In this appendix we follow Smith and LeSage (2004), and derive a sequence of univariate conditional posterior distributions for each component of θ and ϕ that allows the MCMC sampling scheme to be applied to models involving large numbers of regions *n* while avoiding matrix inversion of the *n*-by-*n* matrices required for the multivariate normals set forth in the text of the paper. For small problems involving $n < 100$ regions, it is probably faster to simply carry out the matrix inversions, but no experiments have been carried out to assess the relative computational trade-offs here.

We begin with the observation of Smith and LeSage (2004) in a similar context as in (A4), the expression for the multivariate variance-covariance Σ_θ does not involve inversion. Since the univariate conditionals for each component θ_i of θ must be proportional to this density, it follows that the mean and variance for a sequence of univariate normals for each component of θ can be calculated.

If we let $\theta_{-i} = (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_n)$, for each $i = 1, \ldots, n$, then

$$\begin{aligned} p(\theta \mid \beta, \phi, \rho_o, \rho_d, \tau, R) &\propto \pi(\theta \mid \rho_o, \sigma_o^2) \\ &\quad \exp\{-\tfrac{1}{2}[V\theta - (y + X\beta + W\phi)]^T \Omega^{-1} [V\theta - (y + X\beta + W\phi)]\} \\ &\propto \exp\{-\tfrac{1}{2}[V\theta - (y + X\beta + W\phi)]^T \Omega^{-1} [V\theta - (y + X\beta + W\phi)]^T\} \\ &\quad \exp\{-\tfrac{1}{2\sigma_o^2}\theta^T B_o^T B_o \theta\} \\ &\propto \exp\{-\tfrac{1}{2}[\theta^T (\tfrac{1}{\sigma_o^2} B_o^T B_o + V\Omega^{-1}V)\theta - 2(y + X\beta + W\phi)^T \Omega^{-1} V\theta]\}. \end{aligned} \tag{A4}$$

This expression can be reduced to terms involving only θ_i as follows. First, let $\varphi = V^T \Omega^{-1}(y + X\beta + W\phi)^T$, with φ_i representing the *i*th element of the vector φ. Rewriting the expression in the brackets in (A4) as

$$\tfrac{1}{\sigma_o^2}[\theta^T\theta - 2\rho_o\theta^T V\theta + \rho_o^2\theta^T V^T V\theta] + \theta^T V^T \Omega^{-1} V\theta - 2\varphi^T\theta\,, \tag{A5}$$

we permute indices using $\theta^T = (\theta_i, \theta_{-i})$, and then derive expressions for $\theta^T V\theta$ and $\theta^T V^T V\theta$:

$$\theta^T V\theta = \theta^T \left(\sum_{j\neq i} \theta_j v_{ij} \right) + \left(\theta_i \;\; \theta_{-i}^T \right) \begin{pmatrix} \sum_{j\neq i} \theta_j v_{ij} \\ Q \end{pmatrix} = \theta_i \sum_{j\neq i} \theta_j (v_{ji} + v_{ij}) + Q \tag{A6}$$

where Q denotes a constant not involving parameters of interest. A similar expression arises for:

$$\begin{aligned} \theta^T V^T V\theta &= (\theta_i \, v_{\bullet i} + V_{-i}\, \theta_{-i})^T (\theta_i \, v_{\bullet i} + V_{-i}\, \theta_{-i}) \\ &= \theta_i^2 \, v_{\bullet i}^T \, v_{\bullet i} + 2\, \theta_i (v_{\bullet i}^T \, V_{-i}\, \theta_{-i}) + Q \end{aligned} \tag{A7}$$

where $v_{\bullet i}$ represents the ith column of V, and V_{-i} the n-by-(n-1) matrix containing all columns of V except column i, where $V_{-i} = (\tilde{v}_1, ..., \tilde{v}_{i-1}, \tilde{v}_{i+1}, ..., \tilde{v}_n)^T$.

In addition to the above expressions, we note that:

$$\theta^T \theta = \theta_i^2 + Q \tag{A8}$$

$$\theta^T V^T \Omega^{-1} V\theta = \theta_i^2 n_i \Big/ \sum_{j=1}^{n_i} s_{i,r_{ij}} + Q \tag{A9}$$

$$-2\, \varphi^T \theta = -2\, \varphi_i \;\theta_i \; + Q. \tag{A10}$$

We can use these expressions to rewrite the conditional posterior for θ_i as

$$\begin{aligned} (\theta_i \mid \bullet) \;\; &\propto \;\; \exp(-\tfrac{1}{2}[-2\,\rho_o\, \theta_i \sum_{j\neq i}^{n} \theta_j (v_{ji} + v_{ij})\theta_i + \rho_o^2\, \varphi_i^2 \, v_{\bullet i}^T \, v_{\bullet i} + \\ &\qquad 2\,\rho_o^2\, \theta_i\, (v_{\bullet i}^T \, V_{-i}\, \theta_{-i}) \tfrac{1}{\sigma_o^2} + \theta_i^2\, \Omega_i - 2\;\varphi_i\, \theta_i\,]) \\ &\propto \;\; \exp\{-\tfrac{1}{2}(a_i\, \theta_i^2 - 2\, b_i\, \theta_i + b_i^2 \,/\, a_i)\} \\ &= \;\; \exp\{-\tfrac{1}{2(1/a_i)}\;(\theta_i - \tfrac{b_i}{a_i})\} \end{aligned} \tag{A11}$$

$$a_i = \tfrac{1}{\sigma_o^2} + \tfrac{\rho_o^2}{\sigma_o^2} v_{\bullet i}^T \, v_{\bullet i} + \Omega_i \tag{A12}$$

$$b_i = \varphi_i + \tfrac{\rho_o}{\sigma_o^2} \sum_{j\neq i}^{n} \theta_j (v_{ji} + v_{ij})\theta_j - \tfrac{\rho_o^2}{\sigma_o^2}\; v_{\bullet i}^T \, V_{-i}\;\; \theta_{-i} \tag{A13}$$

which indicates that:

$$\theta_i \mid \theta_{-i}, \beta, \rho_o, \sigma_o^2, \tau, R, y \;\sim\; N_1\left(\frac{b_i}{a_i}, \frac{1}{a_i}\right). \qquad \text{(A14)}$$

A similar derivation leads to $(\phi_i \mid \bullet)$. The conditional posteriors for σ_θ^2:

$$\begin{aligned} p(\sigma_\theta^2 \mid \upsilon, y) &\propto \pi(\theta \mid \rho_o, \sigma_o^2)\pi(\sigma_o^2) \\ &\propto (\sigma_o^2)^{-\frac{n}{2}} \exp(-\tfrac{1}{2\sigma_o^2}\theta^T B_o^T B_o \theta)(\sigma_o^2)^{g_1+1}\exp(\tfrac{g_2}{\sigma_o^2}) \qquad \text{(A15)} \\ &\propto (\sigma_o^2)^{-(\frac{n}{2}+g_1+1)} \exp[-\theta^T B_o^T B_o \theta + \tfrac{2g_2}{2\sigma_o^2}] \end{aligned}$$

which is proportional to the inverse gamma distribution reported in the text. A similar approach leads to $p(\sigma_\phi^2 \mid \upsilon, y)$.

Appendix B: Sampling schemes for τ and R

The material in this appendix is reproduced from Frühwirth-Schnatter and Wagner (2004) for the convenience of the reader. We present a two-block sampler for the inter-arrival times parameters $\tau_{ij}, j = 1, ..., y_i + 1,\ i = 1, ..., N$ and the component indicators $r_{ij}, j = 1, ..., y_i + 1,$ $i = 1, ..., N$. We assume starting values for the inter-arrival times and component indicators, which Frühwirth-Schnatter and Wagner (2004) suggest might be set in the following fashion. For the inter-arrival times, rely on the observed counts y_i and set λ_i = y_i in the procedure described below to sample inter-arrival times. For cases where y_i = 0, we set λ_i = 0.1, a small value in the procedure described below.

Starting values for the component indicators can be set to random draws from 1 to 5. The sampling steps for R and τ also presume starting values for the parameters β, θ, ϕ. Maximum Likelihood estimates from a Poisson regression are used for the parameters β. Starting values for the vectors θ, ϕ are generated from: $\theta \sim N_n[0, \sigma_o^2 (B_o^T B_o)^{-1}]$ with the value for ρ_o in $B_o = I_n - \rho_o W$ set to 0.5 and σ_o^2 = 0.5. Similarly, initial values for $\phi \sim N_n[0, \sigma_d^2 (B_d^T B_d)^{-1}]$, with identical values for ρ_d , σ_d^2 as in the case of θ .

Given these initial values, the sampler for τ and R conditional on knowing υ involves sampling inter-arrival times for $i = 1, ..., N$ setting $\lambda_i = \exp(z_i \delta)$ and $m = y_i$:

(a) if y_i > 0, sample order statistics $\varepsilon_{t,(1)}, ..., \varepsilon_{t,(m)}$ using m uniform random deviates and define the inter-arrival times τ_{ij} based on the increments: $\tau_{ij} = \varepsilon_{i,(j)} - \varepsilon_{i,(j-1)}$, $j = 1, ..., m$, where we set $\varepsilon_{i,(0)} = 0$,

(b) the final arrival time $\tau_{i,(m+1)}$ we set: $\tau_{i,(m+1)} = 1 - \sum_{j=1}^{m} \tau_{ij} + \varsigma_i$, where $\varsigma_i \sim \xi(\lambda_i)$.

The samples for each $j = 1, ..., y_i + 1$, component indicator r_{ij} conditional on τ , υ come from the following discrete density:

$$\Pr(r_{ij} = k \mid \tau_{ij}, \upsilon) \propto p(\tau_{ij} \mid r_{ij} = k, \upsilon) w_k \tag{B1}$$

$$\ln p(\tau_{ij} \mid r_{ij} = k, \psi, \upsilon) \propto -\ln s_k - \frac{1}{2}\left(\frac{\ln \tau_{ij} + z_i \delta - m_k}{s_k}\right)^2, \tag{B2}$$

where the quantities (w_k, m_k, s_k^2), $k = 1, ..., 5$ are those of the finite mixture approximation given in Table 1 of Frühwirth-Schnatter and Wagner (2004).

References

Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004): *Hierarchical Modelling and Analysis for Spatial Data.* Boca Raton , Chapman & Hall

Bernstein, J. and Nadiri, M. (1988): Interindustry R&D spillovers, rates of returns and production in high-tech industries, *American Economic Review* 78, 429-34

Besag, J.E., York, J.C. and Mollie, A. (1991): Bayesian image restoration, with two applications in spatial statistics (with discussion), *Annals of the Institute of Statistical Mathematics* 43, 1-59

Besag, J.E. and Kooperberg, C.L. (1995): On conditional and intrinsic autoregressions, *Biometrika* 82, 733-46

Cameron, A.C. and Trivedi, P.K. (1998): *Regression Analysis of Count Data*. Cambridge [UK] and New York [USA], Cambridge University Press

Chib, S., Greenberg, E. and Winkelmann, R. (1998): Posterior simulation and bayes factors in panel count data models, *Journal of Econometrics* 86, 335-44

Chib, S., Nardari, F. and Shephard, N. (2002): Markov chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics* 108, 281-316

Christensen, O., Roberts, G. and Sköld, M. (2005): Robust MCMC methods for spatial GLMM's, forthcoming in *Journal of Computational and Graphical Statistics*

Cressie, N. (1995): Bayesian smoothing of rates in small geographic areas, *Journal of Regional Science* 35, 659-73

Fischer M.M. and Reggiani, A. (2004): Spatial interaction models: From the gravity to the neural network approach. In Cappello, R. and Nijkamp, P. (eds.): *Urban Dynamics and Growth*. *Advances in Urban Economics*. Amsterdam, Elsevier, pp. 319-46

Fischer, M.M. and Reismann, M. (2002): A methodology for neural spatial interaction modelling, *Geographical Analysis* 34, 207-28

Fischer, M.M., Reismann, M. and Scherngell, T. (2006): Spatial interaction and spatial autocorrelation, Paper Presented at the International Workshop on Spatial Econometrics and Statistics, May 25-27, Rome [Italy]

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 modelling perspective, *Geographical Analysis* 38 [in press]

Frühwirth-Schnatter, S. and Wagner, H. (2004): Data augmentation and Gibbs sampling for regression models of small counts, IFAS Research Paper Series 2004-04

Gelfand, A.E., Sahu, S. and Carlin, B.P. (1995): Efficient parameterizations for normal linear mixed models, *Biometrika* 82, 479-88

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995): *Bayesian Data Analysis.* London, Chapman & Hall

Griliches, Z. (1992): The search for R&D spillovers, *Scandinavian Journal of Economics* 94, Supplement, 29-47

Jaffe, A.B., Trajtenberg, M. and Henderson, R. (1993): Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* 108, 577-98

Karlsson, C. and Manduchi, A. (2001): Knowledge spillovers in a spatial context – A critical review and assessment. In Fischer, M.M. and Fröhlich, J. (eds.): *Knowledge, Complexity and Innovation Systems*. Heidelberg, Berlin and New York, Springer, pp. 101-23

Kim, S., Shephard, N. and Chib, S. (1998): Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* 65, 361-393

LeSage, J.P. and Pace, R.K. (2005): Spatial econometric modelling of origin-destination flows, Paper Presented at the Regional Science Association International Meetings, Las Vegas, NV, November, 2005

Maurseth, P.B. and Verspagen, B. (2002): Knowledge spillovers in Europe: A patent citation analysis, *Scandinavian Journal of Economics* 104, 531-45

Sen, A. and Smith, T.E. (1995): *Gravity Models of Spatial Interaction Behavior*. Heidelberg, Berlin und New York, Springer

Smith, T.E. and LeSage, J.P. (2004): A Bayesian probit model with spatial dependencies. In LeSage, J.P. and Pace, R.K. (eds.): *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*, Oxford, Elsevier Ltd, pp. 127-60

Sun, D., Tsutakawa, R.K. and Speckman, P.L. (1999): Posterior distribution of hierarchical models using car(1) distributions, *Biometrika* 86, 341-50

Sun, D., Tsutakawa, R.K., Kim, H. and He, Z. (2000): Bayesian analysis of mortality rates with disease maps, *Statistics in Medicine* 19, 2015-35