Aurélien Fichet de Clairfontaine and Manfred M. Fischer and Rafael Lata and Manfred Paier

Barriers to cross-region research and development collaborations in Europe. Evidence from the fifth European Framework Programme

Paper

Barriers to cross-region research and development collaborations in Europe.
Evidence from the fifth European Framework Programme

Aurélien Fichet de Clairfontaine[1], Manfred M. Fischer[1], Rafael Lata[2], and Manfred Paier[3]

[1]Vienna University of Economics and Business (WU), Welthandelsplatz 1, A-1020 Vienna, Austria., E-mail: afichet@wu.ac.at, manfred.fischer@wu.ac.at
[2]The Austrian Research Promotion Agency (FFG), Sensengasse 1, A-1090 Vienna, Austria. E-mail: rafael.lata@ffg.at
[3]Austrian Institute of Technology (AIT), Donau-City-Strasse 1, A-1220 Vienna, Austria. E-mail: manfred.paier@ait.ac.at

Abstract

The focus of this paper is on cross-region R&D collaboration funded by the 5th EU Framework Programme (FP5). The objective is to measure distance, institutional, language and technological barrier effects that may hamper collaborative activities between European regions. Particular emphasis is laid on measuring discrepancies between two types of collaborative R&D activities, those generating output in terms of scientific publications and those that do not. The study area is composed of 255 NUTS-2 regions that cover the pre-2007 member states of the European Union (excluding Malta and Cyprus) as well as Norway and Switzerland. We employ a negative binomial spatial interaction model specification to address the research question, along with an eigenvector spatial filtering technique suggested by Fischer and Griffith (2008) to account for the presence of network autocorrelation in the origin-destination cooperation data. The study provides evidence that the role of geographic distance as collaborative deterrent is significantly lower if collaborations generate scientific output. Institutional barriers do not play a significant role for collaborations with scientific output. Language and technological barriers are smaller but the estimates indicate no significant discrepancies between the two types of collaborative R&D activities that are in focus of this study.

Keywords Research collaboration · EU Framework Programme · Negative binomial spatial interaction model · Spatial filter methodology · European regions
JEL Classification C31 · O39 · R15

1 Introduction

In Europe, the primary instruments to foster collaborative research and development (R&D) activities across nation states and regions are the Framework Programmes on Research and Technological Development. They are specifically designed to pool resources and promote international pre-competitive R&D collaboration by intensifying interactions among researchers and regions. By means of these instruments, the European Union has co-funded thousands of transnational, collaborative R&D projects. Implementation of the Framework Programmes started in 1984, the seventh programme did run from 2007 to 2013. Since the launch in 1984, funding has been focused on multidisciplinary research at a transnational level. Over the years, different thematic aspects have been addressed and the main emphasis has shifted towards the establishment of an integrated European Research Area.

The present study focuses on cross-region R&D collaboration networks in Europe, as captured by data on projects of the fifth Framework Programme (FP5). Within FP5 research projects were funded over a time period of five years (1998-2002), with a total budget of 13.7 billion Euro. FP5 focused on a limited number of research areas combining technological, industrial, economic, social and cultural aspects[1]. With its corresponding financial support, FP5 was open to all legal entities (individuals, industrial and commercial firms, universities, research organisations, etc) established in the member states of the European Union. Proposals could be submitted by a consortium consisting of at least two independent legal entities established in different member states or in a member state and an associated state[2] (CORDIS 2008). Proposals were funded based on a series of criteria including scientific excellence, added value for the European Community, the potential contribution to furthering the economic and social objectives of the Community, the innovative character, the prospects for disseminating/exploiting the results, and effective transnational cooperation (see European Council 1998 for more details).

We use data on joint R&D projects funded by the fifth European Framework Programme to proxy cross-region collaborative R&D activities between European regions. Cross-region collaborations between regions, say i and j, are defined as sum of collaborations between actors located i and j, respectively. The objectives of the study are threefold: *first*, to identify patterns of two types of cross-region R&D collaborations, namely those generating outcome in terms of scientific publications and those that do not; *second*, to measure effects of barriers that may hamper such collaborative R&D activities, and *third*, to explore whether there are significant differences between the two types of

[1]The thematic priorities in FP5 are the following (with the subprogramme name given in parentheses): Quality of life and management of living resources (Quality of life); user-friendly information society (IST); competitive and sustainable growth (GROWTH); energy, environment and sustainable development (EESD); confirming the international role of community research (INCO2); promotion of innovation and encouragement of SME participations (Innovation/SME); improving the human research potential and the socio-economic knowledge base (Improving) (CORDIS 2008). Moreover, it is worth noting that FP5 emphasised the protection of intellectual property rights in order to improve the efficiency of collaboration within the various types of European research projects.

[2]Associated states included the candidates for EU membership in that time period (Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuana, Malta, Poland, Romania, Slovakia, Slovenia) as well as Iceland, Israel, Liechtenstein, Norway and Switzerland (see CORDIS 2008).

collaboration. In particular, we are interested to explore the impact of barriers measured in terms of the geographical distance between regions, institutional barriers proxied by the existence of country borders between regions, language barriers between regions proxied by the existence of distinct official languages spoken, and technological barriers between regions.

With its focus on cross-region collaboration, the study shifts attention from organisations to dyads of regions as units of observation. As in Ponds et al. (2007), Hoekman et al. (2009), Maggioni and Uberti (2009), Scherngell and Barber (2009, 2011), we adopt a spatial interaction modelling perspective to estimate the barrier effects. The study area is composed of 255 NUTS-2 regions that cover the pre-2007 member states of the European Union (excluding Malta and Cyprus), as well as Norway and Switzerland.

Our study departs from previous literature in at least three major respects. *First*, the distinction in the empirical investigation between the above mentioned two types of cross-region R&D collaborations established within the fifth European Framework Programme is an original contribution of the current study. *Second*, we use dyads of regions as units of observation and analysis in order to analyse patterns of collaborative activities established within the Framework Programme, in contrast to most other studies that use organisations as observational units (see, for example, Almendral et al. 2007, Paier and Scherngell 2011, Barber et al. 2011, Reinold et al. 2014). This is an appropriate choice since the study shifts interest from organisations to cross-region collaborative research and development activities across Europe.

Third, we adopt a spatial interaction modelling perspective to estimate the impact of barriers, but in contrast to previous research (see Maggioni et al. 2007, Ponds et al. 2007, Hoekman et al. 2009, Scherngell and Barber 2009, 2011)[3] we account for network autocorrelation (also termed network dependence) present in the cross-region collaboration data with an eigenvector spatial filtering technique suggested by Fischer and Griffith (2008). A virtue of employing a spatially filtered version of the negative binomial spatial interaction model specification is that standard pseudo maximum likelihood techniques can be applied to produce consistent and unbiased parameter estimates, and to draw correct conclusions.

The remainder of the paper is structured as follows: the section that follows presents the negative binomial spatial interaction model for cross-region collaboration and sets forth the spatial filter methodology to account for network autocorrelation as it applies to the negative binomial spatial interaction model specification. Section 3 describes the empirical setting, and presents and discusses the estimation results while Section 4 closes with a summary of the main results and points to some future research.

[3]These studies fail to account for network autocorrelation. Hence the results are likely to be biased and may lead to unreliable or incorrect conclusions. A notable exception accounting for network autocorrelation in modelling collaboration flows is the study by Scherngell and Lata (2013).

2 The model for cross-region R&D collaborations

Let Y_{ij} denote R&D collaborations between regions $i = 1, \ldots, n$ and $j = 1, \ldots, n$, as measured by joint FP5 projects. For convenience, the total number of observations is denoted by N. In its simplest form, a spatial interaction model for cross-region R&D collaborations, Y_{ij} is proportional to the product of an origin factor X_i (proxied by $Y_{i\bullet}$), a destination factor X_j (proxied by $Y_{\bullet j}$)[4], and a distance deterrence function involving distance, D_{ij} between i and j, broadly construed to include all factors that might hamper cross-region collaboration activities.

$$Y_{ij} = \beta_0 X_i^{\beta_1} X_j^{\beta_2} D_{ij}^{\beta_3} \tag{1}$$

where β_0, β_1, β_2 and β_3 are unknown parameters. Typically, the stochastic version of this spatial interaction model has the form

$$Y_{ij} = \beta_0 X_i^{\beta_1} X_j^{\beta_2} D_{ij}^{\beta_3} \xi_{ij} \tag{2}$$

where ξ_{ij} is a disturbance term with $\mathrm{E}\left[\xi_{ij} \mid X_i, X_j, D_{ij}\right] = 1$ assumed to be statistically independent of the explanatory variables X_i, X_j and D_{ij}. This leads to

$$\mathrm{E}\left[Y_{ij} \mid X_i, X_j, D_{ij}\right] = \beta_0 X_i^{\beta_1} X_j^{\beta_2} D_{ij}^{\beta_3}. \tag{3}$$

The most prevalent approach to estimate the multiplicative model given by Eq. (2) is to use a log-log transformation and then to estimate the parameters of interest by ordinary least squares. But this practice is inappropriate for a number of reasons. *First*, Y_{ij} can be zero and then log-linearisation is infeasible. Indeed, the level of collaboration between any two regions is frequently zero. In this study, more than 90% of the total observations are zero flows. *Second*, even if all collaboration observations are strictly positive, it should be noted that the validity of the estimation approach critically depends on the assumption that ξ_{ij}, and hence $\ln \xi_{ij}$, are statistically independent of the explanatory variables. Santos Silva and Tenreyro (2006) show that, if we assume ξ_{ij} to follow a log-normal distribution, with $\mathrm{E}\left[\xi_{ij} \mid X_i, X_j, D_{ij}\right] = 1$ and variance-covariance $\sigma_{ij}^2 = f\left(X_i, X_j, D_{ij}\right)$, then the log-linearised version of these disturbances has $\mathrm{E}\left[\ln \xi_{ij} \mid X_i, X_j, D_{ij}\right] = -\frac{1}{2} \ln\left(1 + \sigma_{ij}^2\right)$, which exhibits dependence for consistency of ordinary least squares.

A natural solution to these problems is to estimate the spatial interaction model directly from its multiplicative form. Since this removes the need to linearise the model by using logarithms, the problem with zero collaboration observations disappears. In doing so, note that the multiplicative spatial interaction relationship can be written as the exponential function $\exp[\ln \beta_0 + \beta_1 \ln X_i + \beta_2 \ln X_j + \beta_3 \ln D_{ij}]$, interpreted as the conditional expectation of Y_{ij} given X_i,

[4]Note that $Y_{i\bullet}$ is defined as $\sum_{j=1}^{n} Y_{ij}$ and $Y_{\bullet j} = \sum_{i=1}^{n} Y_{ij}$.

X_j and D_{ij}, as shown in Eq. (4):

$$\mu_{ij} = \mathrm{E}\left[Y_{ij} \mid X_i, X_j, D_{ij}\right] = \exp\left[\ln\beta_0 + \beta_1 \ln X_i + \beta_2 \ln X_j + \beta_3 \ln D_{ij}\right]. \quad (4)$$

The advantage of this specification is that the coefficients β_1, β_2 and β_3 on the logged variables X_i, X_j and D_{ij} can be interpreted as the elasticity of the conditional expectation of Y_{ij} with respect to X_i, X_j and D_{ij}. For convenience, Eq. (4) may be written in short-form as

$$\mu_k = \mathrm{E}\left[y_k \mid \boldsymbol{z}_k\right] = \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right) \qquad k = 1, \ldots, N \quad (5)$$

where y_k denotes the k-th element of the N-by-1 vector of collaboration flows for the origin-destination pairs of regions, with $N = n^2$. The conditional mean μ_k depends on covariates $\boldsymbol{z}_k$ with associated parameter vector $\boldsymbol{\beta}$.

One way to estimate the multiplicative spatial interaction equation is based on the Poisson probability specification, with the probability density given by

$$\mathrm{Prob}\left[y_k \mid \boldsymbol{z}_k\right] = \frac{\exp\left(-\mu_k\right) \mu_k^{y_k}}{y_k!} \quad (6)$$

where μ_k is specified as $\mu_k = \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right)$. The model has the convenient property that

$$\mathrm{E}\left[y_k \mid \boldsymbol{z}_k\right] = \mu_k. \quad (7)$$

An important implicit assumption of the Poisson spatial interaction model is the equality between the conditional mean and the conditional variance, that is: $\mathrm{E}\left[y_k \mid \boldsymbol{z}_k\right] = \mathrm{var}\left[y_k, \boldsymbol{z}_k\right]$. If this assumption does not hold, then the maximum likelihood coefficient estimates are consistent but not efficient. The standard errors will be biased downward, and inferences should be based on a robust covariance matrix estimator (see Gourieroux et al. 1984 for details).

An alternative, however, that we follow in this study is to specify the variance in a more accurate way. The negative binomial spatial interaction model provides an obvious model specification (see, for example, Fischer et al. 2006) to handle the extra variance. This probability distribution can be written as

$$\mathrm{Prob}(y_k \mid \boldsymbol{z}_k) = \frac{\Gamma\left(y_k + \alpha^{-1}\right)}{\Gamma\left(y_k + 1\right)\Gamma\left(\alpha^{-1}\right)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_k}\right)^{\frac{1}{\alpha}} \left(\frac{\mu_k}{\alpha^{-1} + \mu_k}\right)^{y_k} \quad (8)$$

where $\Gamma\left(.\right)$ is the gamma function, and α is an ancillary parameter indicating the degree of overdispersion to be estimated along with $\boldsymbol{\beta}$. The larger α is, the larger is the degree of dispersion. The model converges to a Poisson spatial model if α is close to zero. A test of the Poisson distribution may be carried out by testing the hypothesis $\alpha = 0$ using the Wald or likelihood ratio test (see Greene 1997).

The negative binomial can be derived from a Poisson distribution in which the μ_k are distributed as a gamma random variable (Gourieroux et al. 1984, Greene 1997). The first two moments of the negative binomial distribution are given by

$$\mathrm{E}\left[y_k \mid \boldsymbol{z}_k\right] = \mu_k = \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right) \tag{9}$$
$$\mathrm{var}\left[y_k \mid \boldsymbol{z}_k\right] = \mu_k\left(1 + \alpha\mu_k\right) = \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right)\left(1 + \alpha \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right)\right) \tag{10}$$

so that the expected value of the observed cross-region collaborations in the negative binomial spatial interaction model is the same as in the Poisson spatial interaction model, but the variance is specified as a function of both the conditional mean and the dispersion parameter α, incorporating unobserved heterogeneity into the conditional mean (Long 1997). Since μ and $\boldsymbol{\beta}$ are positive, $\mathrm{var}\left[y_k, \boldsymbol{z}_k\right]$ is greater than $\mathrm{E}\left[y_k \mid \boldsymbol{z}_k\right]$.

The negative binomial distribution belongs to the family of linear exponential distributions. Hence, the negative binomial pseudo maximum likelihood (NBPML) estimator is defined as the solution to the following pseudo maximum likelihood equation (Gourieroux et al. 1984):

$$\sum_{k=1}^{N} \boldsymbol{z}_k \frac{y_k - \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right)}{1 + \alpha \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right)} = 0. \tag{11}$$

The Hessian matrix is

$$-\sum_{k=1}^{N} \frac{\boldsymbol{z}_k^{'} \boldsymbol{z}_k \left[1 + 2\alpha \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right)\right] \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right)}{\left(1 + \alpha \exp\left(\boldsymbol{z}_k \boldsymbol{\beta}\right)\right)^2}. \tag{12}$$

The objective function is concave and can easily be maximised by using Newton-type algorithms. From Gourieroux et al. (1984) we know that this NBPML estimator is strongly consistent and asymptotically normal. One possible estimator for α is

$$\hat{\alpha} = \frac{1}{N - R} \sum_{k=1}^{N} \frac{\left[\left(y_k - \hat{\mu}_k\right)^2 - \hat{\mu}_k\right]}{\hat{\mu}_k^2}, \tag{13}$$

proposed by Gourieroux et al. (1984), where R denotes the number of covariates. The motivation for this estimator of α implies $\mathrm{E}[(y_k - \mu_k)^2 - \mu_k] = \alpha\mu_k^2$, and thus $\alpha = \mathrm{E}[[(y_k - \mu_k)^2 - \mu_k]/\mu_k^2]$. The corresponding sample moment with degrees-of-freedom correction is Eq. (13) (see Gourieroux et al. 1984; Cameron and Trivedi 1998, p. 65).

Pseudo maximum likelihood estimation of the parameters is based on the assumption that the origin-destination collaboration flows are independent. Assuming independence between flows is heroic, since origin-destination flows are fundamentally spatial in nature, and hence not independent, but spatially dependent (see Bolduc et al. 1995, Tiefelsdorf 2003, LeSage and Pace 2009). One way to overcome this problem is by incorporating spatial dependence into the negative binomial version of the spatial interaction model. Another way to address spatial dependence in origin-destination flows involves eigenvector spatial filtering (see Chun 2008, Fischer and Griffith 2008, Chun and Griffith 2011, Griffith and Fischer 2013).

Spatial filtering used here in this paper relies on a spectral decomposition of the transformed spatial weight matrix MWM, where W is an N-by-N spatial weight matrix

$$W = W_n \otimes W_n \tag{14}$$

that captures spatial dependence between origin-destination collaboration flows from regions neighbouring both the origins and destinations, labelled origin-to-destination dependence by LeSage and Pace (2009). W_n is a row-stochastic n-by-n spatial weight matrix that describes spatial neighbourhood relationships between the n European regions. This matrix has – by convention – zeros in the main diagonal, and non-negative elements in the off-diagonal cells. Specifically the (i,j)-th element of W_n is greater than zero if i and j are neighbouring regions.[5] $\otimes$ denotes the Kronecker product, and M is the N-by-N projection matrix $M = I_N - \iota_N \iota_N' \frac{1}{N}$ where I_N is the N-by-N identity matrix, and ι_N the N-by-1 vector of ones.

The orthogonality properties of eigenvectors make the spectral decomposition useful for lower rank approximations to MWM (see Pace et al. 2013). The usual approach is to keep all the eigenvectors associated with the largest magnitude eigenvalues and discard the rest. This involves partitioning the eigenvalues and vectors into two sets, a set of eigenvectors associated with the largest Q eigenvalues and a set of eigenvectors associated with the smallest $N-Q$ eigenvalues of MWM. We follow Tiefelsdorf and Griffith (2007) to identify and optimise the subset of Q eigenvectors by stepwise integration of the eigenvectors. The Q eigenvectors identified are used as additional explanatory variables in Eq. (4) to filter or approximately destroy spatial dependences in the residuals.

3 Data description and estimation results

Based on the spatially filtered negative binomial spatial interaction model specification described in the previous section we distinguish three model versions: Model A uses, for purposes of comparison, all the cross-region R&D collaborations as dependent variable; model B uses cross-region R&D collaborations generating output in terms of scientific publications, and model C cross-region R&D collaborations producing not such an output. The dependent variables in these models describe region-by-region collaboration intensities identified as sum of individual collaborative activities (with or without publication output) between organisations located in the origin-destination pairs of regions.

Data for constructing these dependent variables come from combining two data sources: the EUPRO database that contains information on the FP5 projects and the participating organisations (including their names and addresses), and an ex-post survey[6] of FP5 projects that provides information on

[5]Neighbours may be defined using contiguity or measures of spatial proximity such as cardinal distance (for example, in terms of the great circle distance) or ordinal distance (for example, in terms of k-nearest neighbours). In this application, we use the concept of k-nearest neighbours with $k = 5$ to define W.

[6]The survey was conducted in 2007 by the Austrian Institute of Technology. Question-

output performance in terms of scientific publications. By Europe we mean the pre-2007 member states of the European Union (excluding Cyprus and Malta) as well as Norway and Switzerland, disaggregated into 255 NUTS-2 regions[7] (NUTS version 2003). A full list of the regions is provided in the Appendix. To construct the dependent variables for the three models, we use a concordance scheme between postal codes and NUTS-2 regions, to aggregate the individual collaboration activities of the organisations to the dyad level of regions, and adopt hereby the full counting rather than the fractional counting procedure to do justice to the true integer nature of R&D collaborations[8] (for details see Fischer et al. 2006). There are 5,343 intra- and interregional R&D collaborations in total; 1,858 with publication output and 3,485 without. About 95 percent of all pairs of regions do not collaborate at all.

Figure 1 about here

Figure 1 visualises the three dependent variables in form of region-by-region networks, Fig.1 (a) the dependent variable for model A, Fig.1 (b) that for model B, and Fig.1 (c) that for model C. The nodes represent the regions, and the lines the presence of R&D collaboration activities between European regions as captured by those organisations that participated in the ex-post survey. Note that only observations with an interregional collaboration intensity of more than three cooperations are displayed to circumvent the cluttering problem. Hence, the majority of the observations including numerous short-distance intra- and interregional collaborations are not visualised here. The spatial network maps reveal a quite different spatial structure of the R&D networks with and without publication output across European regions. It is notable that collaborative activities without publication output are more clustered in the centre of Europe than collaborations with publication output. Île-de-France is the central hub in both spatial networks.

We specify the variable D in Eq. (1) to include four factors that might hamper collaborative activities between regions: distance, institutional, cultural and technological barriers between NUTS-2 regions. Distance between regions is measured in terms of the great circle distance between their economic centres. Institutional barriers are proxied by a country dummy variable. The variable takes a value of zero if the two regions are located in the same country, and one otherwise. A language area dummy variable is used to proxy for cultural barriers. This variable takes a value of zero if the regions are located in the same language area[9], and one otherwise. The final variable included captures tech-

naires were sent out (via e-mail) to participating organisations of 9,107 FP5 projects with 20 or less participating organisations [that is, 59 percent of all FP5 projects]. 1,686 organisations returned the completed questionnaire, representing a response rate of 18.5 percent. The survey covers about 2.6 percent of all participating organisations in the fifth Framework Programme, and provides information on partner selection, intra-project collaboration and output performance in terms of scientific publications.

[7] NUTS-2 regions, though varying in size, are generally considered to represent an appropriate level of spatial granularity for modelling cross-region collaborations in Europe (see, for example, Scherngell and Barber 2011, Hoekman et al. 2013, Scherngell and Lata 2013).

[8] Note, for example, that for a project with three different participating organisations located in three different regions (say i, j and k), we count three links from i to j, j to k and from k to j.

[9] Language areas are defined by the region's official language. Note that Belgium has French

nological barriers between regions measured in terms of technological distance. We follow Fischer et al. (2006) to use regional patent data from the European Patent Office and construct a 630-by-1 technological vector for each region that contains its share of patenting in each of the 630 technological subclasses at the third level of the International Patent Classification System. Technological proximity between two regions is measured in terms of the uncentred correlation between their technological vectors. Two regions that patent exactly in the same proportion in each subclass have a proximity index equal to one, while two regions patenting only in different subclasses have an index equal to zero. This proximity index is appealing because it allows for a continuous measure of technological distance by simple transformation (see Fischer et al. 2006).

Table 1 about here

Table 1 reports the parameter estimates, the associated p-values and standard errors for the three model versions A, B and C. The parameters are estimated by the NBPML estimator described in Section 2. The significant estimates for the dispersion parameter α indicate that the negative binomial model specification is appropriate for controlling unobserved heterogeneity between dyads of regions leading to overdispersion. Spatial filtering relies on the eigenvectors associated with 13, 5 and 11 largest eigenvalues in the case of the three model versions A, B and C respectively.

Model A produces results for *all* the cross-region R&D collaborations considered in this study. The results provide evidence that geographical distance between regions has a negative and significant impact (parameter estimate: -0.242 with *s.e.=0.03*) on collaborative activities between European regions. The coefficient on the variable used to proxy institutional barriers is somewhat larger (parameter estimate: -0.334, *s.e.=0.09*), but the difference is not significant. The same is true for the language area variable with an estimated elasticity of -0.302 (*s.e.=0.074*). This indicates a similar role for geographical distances, institutional and language barriers in the determination of collaboration patterns. Most important, however, are technological barrier effects as evidenced by the parameter estimate -0.902 (*s.e.=0.132*). This indicates that cross-region R&D collaborative activities occur most likely between regions that are close to each other in the 630-dimensional technological space. This finding that technological barriers are more important than distance barriers is in line with previous research (see Scherngell and Barber 2009, 2011; Scherngell and Lata 2013), but also with studies using patent citations to model interregional knowledge spillovers (see Maurseth and Verspagen 2002; Fischer et al. 2006; LeSage et al. 2007).

Evidently, there are differences between the two types of collaborative R&D activities that are in the focus of this study. In the case of collaborations with publication output spatial separation effects are much less important than in the case of collaborations without, as evidenced by the results of model B and model C. The estimate for geographical distance decreases in magnitude from -0.242 (*s.e.=0.03*) in model A to a value of -0.170 (*s.e.=0.036*) in model B, and this difference is significant. In contrast, in model C the estimated elasticity

speaking and Flemish speaking regions; Switzerland has German speaking, French speaking and Italian speaking regions.

is -0.291 (*s.e.=0.03*). This significantly larger estimate suggests that the role of geographical distance as collaboration deterrent is significantly larger in the case of R&D collaborations that do not generate scientific output. Institutional barriers as captured by the country border variable are not significant in model B, but are nearly as important as geographic distance barriers in model C. This result indicates that institutional barriers have been overcome in the case of cross-region collaborations generating scientific output, however not so in the case of those without publication output. We do not find significant differences in the role of common language. Finally, it is worth noting that model B estimates a much smaller effect of technological distance for collaborations with publication output, approximately half of that indicated by model C. But the contrast in estimates is not significant.

4 Closing remarks

The negative binomial specification of the spatial interaction model along with pseudo maximum likelihood procedures has become a popular way of dealing with several econometric issues that arise when modelling cross-region flows (see, for example, Krizstin and Fischer 2014). A problem with the standard negative binomial specification, however, is that the collaboration flows are not independent in geographical space. Spatial dependence in flows involves correlation among collaboration flows between regions that are neighbouring a given origin-destination pair of regions. A failure to account for spatial dependence in the model specification may lead to biased parameter estimates and incorrect conclusions. This problem has been largely neglected so far in the empirical literature on R&D cooperations. To address this problem the paper recommends spatial filtering that provides a way of filtering spatial dependence in the sample to reduce bias in estimates of the parameters associated with the explanatory variables. A virtue of this approach is that the standard negative binomial spatial interaction model can be used to analyse barriers to cross-region collaborative R&D activities in Europe, and existing software can be applied.

The study, that uses FP5 collaboration data in combination with information on the publication output of collaborative activities provided by an ex-post survey, produced some interesting results on cross-region collaborations. The role of geographical distance as collaboration deterrent is significantly lower if collaborations generate scientific output. Institutional barriers, broadly proxied by the country border variable, do not play a significant role. Language and technological barriers are smaller, but the estimates indicate no significant discrepancies between collaborations producing scientific output or not.

A matter of future research is to extend the model formulation to a two-part model that provides a way to deal with the issue of the prevalence of zero observations (see, for example, Krizstin and Fischer 2014 for details). The first part of the model would consist of a logit or probit equation to distinguish between zero and positive outcomes, and the second part would use a negative binomial spatial interaction model specification along with spatial filtering and pseudo-maximum likelihood estimation techniques as outlined in this paper.

Acknowledgements This study reports research carried out in the framework of the Innovation Economics Vienna Program (IEV), organised by the Institute for Economic Geography and GIScience at the Vienna University of Economics and Business [WU] in cooperation with the Innovation Systems Department at the Austrian Institute of Technology [AIT].

References

Almendral JA, Oliveira J, López L, Sanjuán MA, Mendes J (2007) The interplay of universities and industry through the FP5 network. New Journal of Physics 9(6):183–198

Barber MJ, Fischer MM, Scherngell T (2011) The community structure of research and development cooperation in Europe. Evidence from a social network perspective. Geographical Analysis 43(4):415–432

Bolduc D, Laferrière R, Santarossa G (1995) Spatial autoregressive error components in travel flow models: An application to aggregate mode choice. In: Anselin L, Florax R (eds) New directions in spatial econometrics, pp. 96–108. Springer, Berlin, Heidelberg and New York

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

Chun Y (2008) Modeling network autocorrelation within migration flows by eigenvector spatial filtering. Journal of Geographical Systems 10(4):317–344

Chun Y, Griffith DA (2011) Modeling network autocorrelation in space–time migration flow data: An eigenvector spatial filtering approach. Annals of the Association of American Geographers 101(3):523–536

CORDIS (2008) FP5 - EC Programme: Maximum amounts and breakdown (1998-2002). Available from URL: http://cordis.europa.eu/fp5/about.htm

European Council (1998) Council decision of 22 December 1998 concerning the rules for the participation of undertakings, research centres and universities and for the dissemination of research results for the implementation of the fifth Framework Programme of the European Community (1998-2002). Available from URL: http:// cordis.europa.eu/fp5/-management/particip/r-ecdecpart.htm#%203%20-%20Contracts

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

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

Gourieroux C, Monfort A, Trognon A (1984) Pseudo-maximum likelihood methods: Applications to Poisson models. Econometrica 52(3):701–720

Greene WH (1997) Econometric analysis. 7th edition. Pearson Education, Essex

Griffith DA, Fischer MM (2013) Constrained variants of the gravity model and spatial dependence: Model specification and estimation issues. Journal of Geographical Systems 15(3):291–317

Hoekman J, Frenken K, van Oort F (2009) The geography of collaborative knowledge production in Europe. The Annals of Regional Science 43(3):721–738

Hoekman J, Scherngell T, Frenken K, Tijssen R (2013) Acquisition of European research funds and its effect on international scientific collaboration. Journal of Economic Geography 13(1):23–52

Kristzin T, Fischer MM (2014) The gravity model for international trade: Specification and estimation issues. Paper presented at the 54th European Congress of the Regional Science Association, St. Petersburg, 26-29 August 2014

LeSage JP, Fischer MM, Scherngell T (2007) Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction modelling perspective. Geographical Analysis 38(3):288–309

LeSage J, Pace RK (2009) Introduction to spatial econometrics. CRC Press, Boca Raton

Long JS (1997) Regression models for categorical and limited dependent variables. Sage, Thousand Oaks, London and New Delhi

Maggioni MA, Uberti TE (2009) Knowledge networks across Europe: Which distance matters? The Annals of Regional Science 43(3):691–720

Maggioni MA, Nosvelli M, Uberti TE (2007) Space versus networks in the geography of innovation: A European analysis. Papers in Regional Science 86(3):471–493

Maurseth PB, Verspagen B (2002) Knowledge spillovers in Europe: A patent citations analysis. The Scandinavian Journal of Economics 104(4): 531–545

Pace RK, LeSage JP, Zhu S (2013) Interpretation and computation of estimates from regression models using spatial filtering. Spatial Economic Analysis 8(3):352–369

Paier M, Scherngell T (2011) Determinants of collaboration in European R&D networks: Empirical evidence from a discrete choice model. Industry and Innovation 18(1):89–104

Ponds R, van Oort F, Frenken K (2007) The geographical and institutional proximity of research collaboration. Papers in Regional Science 86(3):423–443

Reinold F, Paier M, Fischer MM (2014) Joint knowledge production in European R&D networks: Results from a discrete choice modeling perspective. In: Scherngell, T (ed.) The geography of networks and R&D collaborations, pp. 201–220. Springer, Berlin, Heidelberg and New York

Scherngell T, Barber MJ (2009) Spatial interaction modelling of cross-region R&D collaborations: Empirical evidence from the 5th EU Framework Programme. Papers in Regional Science 88(3):531–546

Scherngell T, Barber MJ (2011) Distinct spatial characteristics of industrial and public research collaborations: Evidence from the fifth EU Framework Programme. The Annals of Regional Science 46(2):247–266

Scherngell T, Lata R (2013) Towards an integrated European Research Area? Findings from eigenvector spatially filtered spatial interaction models using European Framework Programme data. Papers in Regional Science 92(3):555–577

Santos Silva JMC, Tenreyro S (2006) The log of gravity. The Review of Economics and Statistics 88(4):641–658

Tiefelsdorf M (2003) Misspecification in interaction model distance decay relations: A spatial structure effect. Journal of Geographical Systems 5(1):25–50

Tiefelsdorf M, Griffith DA (2007) Semiparametric filtering of spatial autocorrelation: The eigenvector approach. Environment and Planning A 39(1):1193–1221

Wasserman S, Faust K (1997) Social network analysis: Methods and applications. Cambridge University Press, Cambridge

Appendix

NUTS is an acronym of the French for the 'nomenclature of territorial units for statistics', 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 NUTS-0 regions (countries) below which are NUTS-1 regions and then NUTS-2 regions. This study disaggregates Europe's territory into 255 NUTS-2 regions located in the EU-25 member states (excluding Cyprus and Malta) as well as Norway and Switzerland. We exclude the Spanish North African territories of Ceuta y Melilla, the Portuguese non-continental territories Açores and Madeira, and the French Departments d'Outre-Mer Guadeloupe, Martinique, Guyane Française and Réunion. Thus, we include the following NUTS-2 regions:

Austria	Burgenland, Kärnten, Niederösterreich, Oberösterreich, Salzburg, Steiermark, Tirol, Vorarlberg, Wien
Belgium	Prov. Antwerpen, Prov. Brabant-Wallon, Prov. Hainaut, Prov. Limburg (B), Prov. Liège, Prov. Luxembourg (B), Prov. Namur, Prov. Oost-Vlaanderen, Prov. Vlaams-Brabant, Prov. West-Vlaanderen, Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest
Czech Republic	Jihovýchod, Jihozápad, Moravskoslezsko, Praha, Severovýchod, Severozápad, Stredni Morava, Stredni Cechy
Denmark	Danmark
Estonia	Eesti
Finland	Aland, Etelä-Suomi, Itä-Suomi, Länsi-Suomi, Pohjois-Suomi

France	Alsace, Aquitaine, Auvergne, Basse-Normandie, Bourgogne, Bretagne, Centre, Champagne-Ardenne, Corse, Franche-Comté, Haute-Normandie, Île-de-France, Languedoc-Roussillon, Limousin, Lorraine, Midi-Pyrénées, Nord-Pas-de-Calais, Pays de la Loire, Picardie, Poitou-Charentes, Provence-Alpes-Côte d'Azur, Rhône-Alpes
Germany	Arnsberg, Berlin, Brandenburg, Braunschweig, Bremen, Chemnitz, Darmstadt, Dessau, Detmold, Dresden, Düsseldorf, Freiburg, Giessen, Halle, Hamburg, Hannover, Karlsruhe, Kassel, Koblenz, Köln, Leipzig, Lüneburg, Magdeburg, Mecklenburg-Vorpommern, Mittelfranken, Münster, Niederbayern, Oberbayern, Oberfranken, Oberpfalz, Rheinhessen-Pfalz, Saarland, Schleswig-Holstein, Schwaben, Stuttgart, Thüringen, Trier, Tübingen, Unterfranken, Weser-Ems
Greece	Anatoliki Makedonia, Thraki, Attiki, Ipeiros, Voreio Aigaio, Dytiki Ellada, Dytiki Makedonia, Thessalia, Ionia Nisia, Kentriki Makedonia, Kriti, Notio Aigaio, Peloponnisos, Sterea Ellada
Hungary	Dél-Alföld, Dél-Dunántúl, Észak-Alföld, Észak-Magyarország, Közep-Dunántúl, Közep-Magyarország, Nyugat-Dunántúl
Ireland	Border, Midland and Western, Southern and Eastern
Italy	Abruzzo, Basilicata, Calabria, Campania, Emilia-Romagna, Friuli-Venezia Giulia, Lazio, Liguria, Lombardia, Marche, Molise, Piemonte, Puglia, Sardegna, Sicilia, Toscana, Trentino-Alto Adige/Südtirol, Umbria, Valle d'Aosta/Vallée d'Aoste, Veneto
Latvia	Latvija
Lithuania	Lieteva
Luxembourg	Luxembourg (Grand-Duché)
Netherlands	Drenthe, Flevoland, Friesland, Gelderland, Groningen, Limburg (NL), Noord-Brabant, Noord-Holland, Overijssel, Utrecht, Zeeland, Zuid-Holland
Norway	Agder og Rogaland, Hedmark og Oppland, Nord-Norge, Oslo og Akershus, Sør-Østlandet, Trøndelag, Vestlandet
Poland	Dolnoślaskie, Kujawsko-Pomorskie, Lubelskie, Lubuskie, Lódzkie, Mazowieckie, Malopolskie, Opolskie, Podkarpackie, Podlaskie, Pomorskie, Ślaskie, Świetokrzyskie, Warmińsko-Mazurskie, Wielkopolskie, Zachodniopomorskie
Portugal	Alentejo, Algarve, Centro (P), Lisboa, Norte
Slovakia	Bratislavsky Kraj, Stredné Slovensko, Východné Slovensko, Západné Slovensko
Slovenia	Slovenija
Spain	Andalucía, Aragón, Cantabria, Castilla y León, Castilla-La Mancha, Cataluña, Comunidad Foral de Navarra, Comunidad Valenciana, Comunidad de Madrid, Extremadura, Galicia, Islas Baleares, La Rioja, País Vasco, Principado de Asturias, Región de Murcia
Sweden	Mellersta Norrland, Norra Mellansverige, Smaland med Öarna, Stockholm, Sydsverige, Västsverige, Östra Mellansverige, Övre Norrland
Switzerland	Espace Mittelland, Nordwestschweiz, Ostschweiz, Région Lemanique, Ticino, Zentralschweiz, Zürich
United Kingdom	Bedfordshire & Hertfordshire, Berkshire, Buckinghamshire & Oxfordshire, Cheshire, Cornwall & Isles of Scilly, Cumbria,

Derbyshire & Nottinghamshire, Devon, Dorset & Somerset, East Anglia, East Riding & North Lincolnshire, East Wales, Eastern Scotland, Essex, Gloucestershire, Wiltshire & North Somerset, Greater Manchester,
Hampshire & Isle of Wight, Herefordshire, Worcestershire & Warkwickshire, Highlands and Islands, Inner London, Kent, Lancashire, Leicestershire,
Rutland and Northamptonshire, Lincolnshire, Merseyside, North Eastern Scotland, North Yorkshire, Northern Ireland, Northumberland and Tyne and Wear, Outer London, Shropshire & Staffordshire,
South Western Scotland, South Yorkshire, Surrey, East & West Sussex, Tees Valley & Durham, West Midlands, West Wales & The Valleys, West Yorkshire

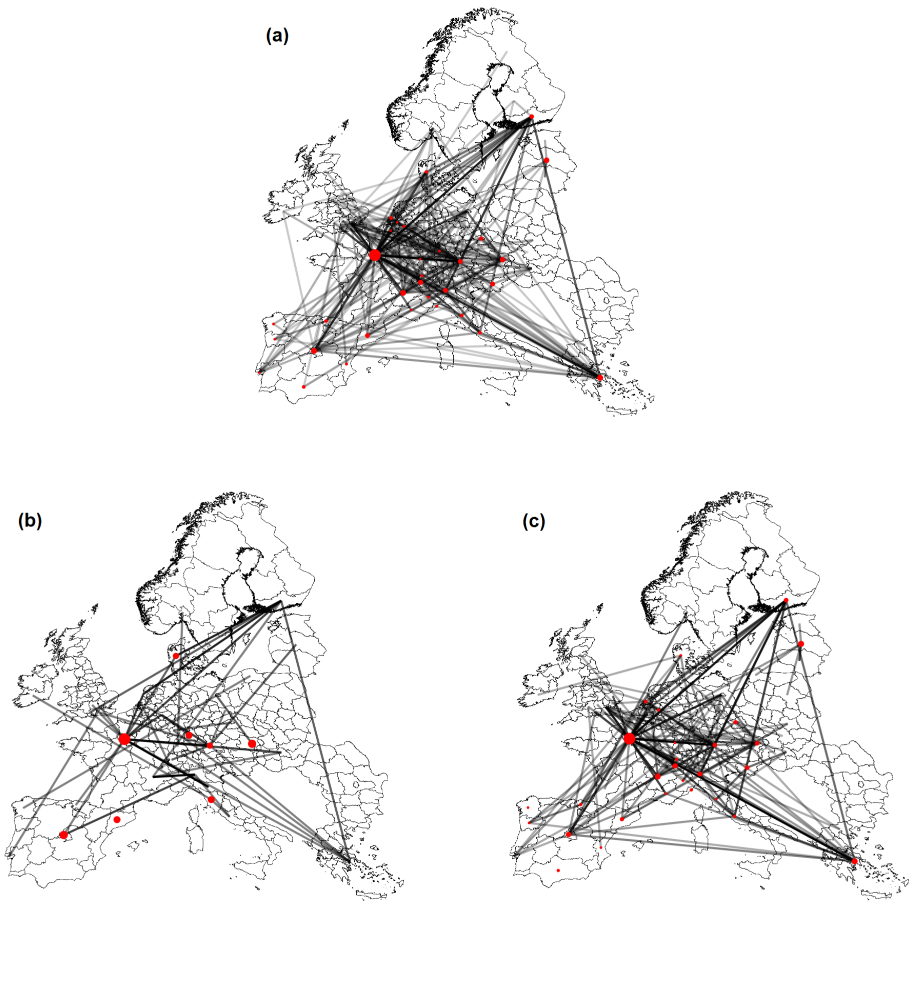

Figure 1: Cross-region R&D collaborations in Europe: (a) All cross-region R&D collaborations (b) cross-region R&D collaborations with publication output, and (c) cross-region R&D collaborations without publication output (the node size corresponds to a region's degree centrality, see Wasserman and Faust 1997)

Table 1: Pseudo ML estimates for the three model versions: Model A for all the R&D collaborations, Model B for those with and Model C for those without publication output (asymptotic standard errors in parentheses)

	Model A		Model B		Model C	
	parameter value	p-value	parameter value	p-value	parameter value	p-value
Geographical distance	-0.242 (0.030)	0.000	-0.170 (0.036)	0.000	-0.291 (0.031)	0.000
Country border	-0.334 (0.090)	0.000	0.048 (0.100)	0.593	-0.342 (0.102)	0.000
Language area	-0.302 (0.074)	0.000	-0.176 (0.083)	0.033	-0.265 (0.082)	0.001
Technological distance	-0.902 (0.132)	0.000	-0.289 (0.158)	0.067	-0.593 (0.134)	0.000
Overdispersion [α]	1.694 (0.092)	0.000	2.630 (0.335)	0.000	3.523 (0.127)	0.000
Log likelihood	-17,034.650		-8,586.992		-12,220.090	
$R^{\star}$	0.743		0.586		0.750	

Notes: $R^{\star}$ is measured as the overall fit of the model in terms of the correlation between the fitted and observed values of the dependent variable; the number of observations is 65,025, including intraregional collaborations.