ePub^WU Institutional Repository

Daniela Gumprecht

R&D Spillovers: A Non-Spatial and a Spatial Examination

Article (Published)
(Refereed)

http://epub.wu.ac.at/

R&D Spillovers:
A Non-Spatial and a Spatial Examination

Daniela Gumprecht

Vienna University of Economics and Business Administration

Abstract: In recent years there were many debates and different opinions whether R&D spillover effects exist or not. In 1995 Coe and Helpman published a study about this phenomenon, based on a panel dataset, that supports the position that such R&D spillover effects are existent. However, this survey was criticized and many different suggestions for improvement came from the scientific community. Some of them were selected and analysed and finally led to a new model. And even though this new model is well compatible with the data, it leads to different conclusions, namely that there does not exist an R&D spillover effect. These different results were the motivation to run a spatial analysis, which can be done by considering the countries as regions and using an adequate spatial link matrix. The used methods from the field of spatial econometrics are described briefly and quite general, and finally the results from the spatial models (the ones which correspond to the non-spatial ones) are compared with the results from the non-spatial analysis. The preferred model supports the position that R&D spillover effects exist.

Zusammenfassung: In den letzten Jahren gab es viele Diskussionen und unterschiedliche Meinungen darüber, ob F&E Spillover Effekte existieren oder nicht. Coe und Helpman veröffentlichten 1995 eine Studie über dieses Phänomen, die für die Existenz solcher F&E Spillover Effekte spricht. Diese Studie wurde viel diskutiert und kritisiert und viele Verbesserungsvorschläge und Alternativen wurden vorgebracht. Einige dieser Vorschläge wurden ausgewählt und z.T. kombiniert und führten schließlich zu einem neuen Modell. Obwohl dieser neue Ansatz sehr gut zu den Daten passt, ist die Schlussfolgerung eine andere, nämlich die, dass es keinen F&E Spillover Effekt gibt. Diese unterschiedlichen Ergebnisse führten zur Überlegung, den Datensatz mit Hilfe von Methoden aus der räumlichen Ökonometrie zu untersuchen, um so den Effekt der F&E Spillovers zu klären. Die verwendeten Methoden werden kurz beschrieben, und die Ergebnisse dieser räumlichen Analyse werden vorgestellt und mit jenen der nicht-räumlichen verglichen. Das bevorzugte räumliche Modell unterstützt die Hypothese der Existenz von F&E Spillover Effekten.

Keywords: Fixed Effects, Random Coefficients, Panel Data Model, Spatial Data, SAR Error Model.

1 R&D Spillover: A Non-Spatial Examination

Beside most studies about economic growth that attempt to explain the growth of an economy predominately by the amount of labor and capital spent (according to a Cobb-Douglas model) plus residual effects from some other economic and political factors,

there are theories that treat commercially oriented innovation efforts as a major engine of technological progress and productivity growth (Romer, 1990; Grossman and Helpman, 1991). Coe and Helpman (1995) defend this theory and claim that the productivity of a global economy depends on its own stock of knowledge as well as the stock of knowledge of its trade partners. This means, they believed in a spillover effect of the foreign stock of knowledge on the productivity of a commercial partner country.

To study the extent to which a country's productivity level depends on domestic and foreign stock of knowledge, Coe and Helpman analyzed a panel dataset with 22 countries (21 OECD countries plus Isreal) over a period of 20 years (from 1971 to 1990). The stock of knowledge is quantified by the amount of money spent for R&D, i.e. the domestic stock of knowledge is measured by the cumulated expenditures for R&D, and the foreign stock of knowledge is measured by an import-weighted sum of the cumulated R&D expenditures of the trade partners. This definition of the foreign R&D capital stock takes the importance of the partner country into account, the higher the imports from a country, the more important are the R&D spending of this country. The variables total factor productivity (TFP), domestic R&D spending (DRD) and foreign R&D spending (FRD) are constructed as indices with basis 1985, because TFP is originally measured in country specific currency and DRD and FRD are measured in U.S. dollars. The importance of the R&D capital stock is finally quantified by the elasticity of total factor productivity with respect to the corresponding R&D capital stock. All data are available on the webpage of Helpman (2003), which is accessible via the internet address:

        `http://post.economics.harvard.edu/faculty/helpman/data.html`

1.1 C&Hs Models and Results

In their paper Coe and Helpman (1995) used a variety of specifications to model the effects of DRD and FRD on TFP. To simplify the exposition only one of those is regarded here. The following conclusions, however, are not limited to this particular case but rather apply to all of the suggested models (for a more complete analysis see D. Gumprecht, 2003). The illustrative model contains TFP as regressand and DRD and FRD as regressors. The equation with regional index i (cross-section dimension) and temporal index t (time dimension) has the form

$$\log F_{it} = \alpha_{it}^0 + \alpha_{it}^d \log S_{it}^d + \alpha_{it}^f \log S_{it}^f + \varepsilon_{it}, \tag{1}$$

where F_{it} denotes total factor productivity (TFP), S_{it}^d domestic R&D capital stock (DRD) and S_{it}^f foreign R&D capital stock (FRD), FRD is defined as a bilateral import-share weighted average of the domestic R&D capital stocks of trade partners

$$S_{it}^f = \sum_{i \neq j} b_{ijt} S_{jt}^d$$

with b_{ijt} being the bilateral import-shares of country i from country j in period t. Note that $b_{ijt} \neq b_{jit}$ and $\sum_j b_{ijt} = 1$. And ε_{it} is an error term.

Coe and Helpman (1995) wanted to estimate the long-run relationship between TFP and DRD and FRD. Therefore and because the series exhibit non-stationarity (as confirmed by respective tests), they estimated co-integrated equations. The OLS estimate of

a co-integrated equation is said to be "super-consistent", that is, the estimate converges to the true parameter value much faster than in the case where the variables are stationary (Stock, 1987). Furthermore they assumed that the impact of domestic and foreign R&D expenditures is the same for all countries, this means the regression coefficients, which correspond to the elasticity of TFP with respect to DRD ($= \alpha_{it}^d$) and FRD ($= \alpha_{it}^f$), are constrained to be the same for all countries, and only the intercepts ($= \alpha_{it}^0$) are allowed to vary across the countries. These country specific intercepts are chosen to be considerate of country specific effects on productivity that are not captured by the variables in the model. According to standard practice in time series literature Coe and Helpman (1995) used a panel data model with fixed effects and estimated it via OLS. This leads to the following estimators of the model given in equation (1).

$$\widehat{\log F_{it}} = \alpha_{it}^0 + \underbrace{0.097}_{0.000} \log S_{it}^d + \underbrace{0.092}_{0.000} \log S_{it}^f.$$

Coe and Helpman (1995) took these estimation results, with both positive regression coefficients as a confirmation of their hypothesis that TFP of a country depends on domestic and foreign R&D capital stocks. They did not calculate t- or p-values for the parameter estimators, because using the standard method leads to biased results, and the asymptotic distribution of the t-values in the case of co-integrated panel data was not known at that time. So, this model was estimated once again, using the Least Squares Dummy Variable (LSDV) method, now including the tests for the parameters. The coefficients are the same as the ones from Coe and Helpman, both coefficients are positive and significant, the p-values are given below the coefficients. In the following, the p-values can always be found below the estimates. The fit of the model is quite fine with *pseudo* $R^2 = 0.558$, which is calculated as the squared correlation between $\hat{y}_{it}$ and y_{it}.

1.2 Critics and a New Model for the R&D Spillovers

Suggestions for improvement of Coe and Helpman's estimations came – among others – from Kao, Chiang, and Chen (1999). They criticized (among other points) that in spite of the super consistency of the time-series estimator, the bias of the estimation can be quite substantial for small samples and there is no reason to assume that this bias becomes negligible by the inclusion of a cross section dimension in panel data. Kao et al. (1999) use different estimation methods for Coe and Helpman's international R&D spillovers regression. They claim that the dynamic OLS (DOLS) estimation is the best solution for this problem because in the given setting the DOLS estimator exhibits no bias and is asymptotically normal. The DOLS estimator is based on a regression additionally including q_1 time lags and q_2 time leads of the regressors, therefore the number of time periods reduces from t to $(t - q_1 - q_2 - 1)$. For the R&D spillover model 2 lags and 1 lead were used.

As a second major issue, there are many debates in the panel data estimation literature, whether to regard the region specific or other effects as random. This poses a valuable alternative to the fixed effects model. In the present context Müller and Nettekoven (1999) suggest a random coefficients model, there the parameters are assumed to vary randomly around a common mean, see e.g. Greene (2003) to analyze the R&D spillovers model given in equation (1). They conclude that, although this alternative specification is well

compatible with the data, one astonishingly has to draw contradictory conclusions. Estimates for the random coefficients model differ decisively from the fixed effects model and especially the estimator of the foreign R&D expenditures even changes sign, although this is not statistically significant. Contrary to Coe and Helpman's conclusions, this model indicates that the foreign R&D effect is not significant. As random coefficients models are more flexible compared to fixed effects models, the fit (measured by *pseudo R^2*) improved substantially. Note however, that flexibility does not enforce a higher model ability to generalize.

After a detailed examination of the original model given in (1) and the various critics of it, the following changes and modifications are suggested by D. Gumprecht, Gumprecht, and Müller (2004). They used a random coefficients model estimated via DOLS. The DOLS random coefficient estimation yields

$$\widehat{\log F_{it}} = \alpha_{it}^0 + \underbrace{0.353}_{0.000} \log S_{it}^d - \underbrace{0.085}_{0.118} \log S_{it}^f .$$

The coefficient for DRD is significant, whereas the one for FRD is not. The fit of the model is rather good with $R^2 = 0.974$. For the calculations a special GAUSS programm (implemented by N. Gumprecht, 2003) was used.

The results of the panel co-integration model with random coefficients and dynamic regressors do not support Coe and Helpman's hypothesis, that the TFP of a country depends on domestic and foreign R&D stock of knowledge and the R&D expenditures respectively. It seems as foreign R&D do rather not affect the TFP of a country. These deviating results in the non-spatial analysis are the motivation to look at this problem from the point of view of spatial econometrics. The question to be answered is, whether the impact of FRD can be clarified under this different aspect. The impact of DRD seems to be already verified by the different non-spatial models and estimation techniques. Before the spatial analysis of the R&D data, a short and general introduction to spatial methods, especially in econometrics, is given.

2   A Short Introduction to Spatial Econometrics

One reason for using a spatial analysis is the exploitation of regional dependencies (so called information spillover) to improve statistical conclusions. Data used in an empirical analysis are a collection of measurements with attached geographical coordinates, in economics frequently called a spatial panel.

2.1   Spatial Data

In general, spatial data have the following characteristics: They contain attribute and locational information (so called georeferenced data). Spatial relationships are modelled with spatial weight matrices, also called spatial link matrices. Such matrices measure the similarities (e.g. contiguity or neighborhood matrices) or dissimilarities (distance matrices) between spatial objects. Similarity and dissimilarity matrices are inversely related – the higher the connectivity, the smaller the distance and vice versa. In general, a spatial link

matrix $W = [w_{ij}]$ is an n by n matrix (n is the number of observations) where $w_{ij} = 0$ if i and j are not spatially connected or if $i = j$ (by definition), and $w_{ij} \neq 0$ if i and j are spatially connected. The original symmetric spatial link matrices are often converted by using coding schemes to cope with the heterogeneity which is induced by the different linkage degrees of the spatial objects. Tiefelsdorf (2000) defines the linkage degree of a spatial object i by the total sum of its interconnections with all other spatial objects, that is $d_i = \sum_{j=1}^{n} w_{ij}$. There are different coding schemes used (see Tiefelsdorf, 2000, p. 29-30), the one used in this paper is the row-standardized W-coding scheme, here the sum of each row is equal to one and the elements are simply calculated by $w_{ij} / \sum_{j=1}^{n} w_{ij}$.

Note that such spatial link matrices are not necessarily restricted to the geographic space, one can also use some other kind of measures for the contiguity or distance between observations, e.g. for the R&D spillovers analysis an economic contiguity measure was defined and used for the spatial link matrix.

2.2 Spatial Dependency and Spatial Autocorrelation

Fotheringham, Brunsdon, and Charlton (2002) say about spatial dependency: "It (spatial dependency) is the extent to which the value of an attribute in one location depends on the values of the attribute in nearby locations." Griffith (2003) says about spatial autocorrelation: " It (spatial autocorrelation (...) is the correlation among values of a single variable strictly attributable to the proximity of those values in geographic space (...)." However spatial dependency is measured (by geographic distances or economic measures), positive spatial autocorrelation means that nearby values of a variable tend to be similar: high values are near high values, medium values near medium values, and low values near low values; negative spatial autocorrelation means that nearby values of a variable tend to be dissimilar: high values tend to be near low values, medium values near medium values, and low values near high values.

2.3 Moran's $\Im$

One of the first questions that raises when analysts have to deal with georeferenced data is, whether there is a spatial effect existent or not. If not, i.e. the observations are spatially independent, there is no need for using special models or methods in the analysis. There are many different possibilities to test spatial autocorrelation, the most commonly used test is based on a statistic developed by Moran (1948, 1950a, 1950b). Spatial autocorrelation can be quantified and tested with Moran's $\Im$ statistic, which is defined as scale invariant ratio of quadratic forms in the normal distributed regression residuals

$$\Im = \frac{\hat{\varepsilon}' \frac{1}{2}(W + W')\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}} \, , \tag{2}$$

where $\hat{\varepsilon}$ are the normal distributed OLS residuals and W is a spatial link matrix. Expected value and variance of Moran's $\Im$, under the assumption of spatial independence, are

$$\mathrm{E}[\Im] = \frac{1}{n-k} \, \mathrm{tr}(MW)$$

$$\mathrm{var}[\Im] = \frac{\mathrm{tr}(MWMW') + \mathrm{tr}(MW)^2 + \{\mathrm{tr}(MW)\}^2}{(n-k)(n-k+2)} - \mathrm{E}^2[\Im] \, ,$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator, and $M = I_n - X(X'X)^{-1}X'$ is the projection matrix. Inference for Moran's $\Im$ is usually based on a normal approximation, using the standardized z-value

$$z(\Im) = \frac{\Im - \mathrm{E}[\Im]}{\sqrt{\mathrm{var}[\Im]}} \, . \tag{3}$$

The z-transformed Moran's $\Im$ is for normal distributed residuals and well-behaved spatial link matrices under the assumption of spatial independence asymptotically standard normally distributed, see e.g. Tiefelsdorf (2000). With this z-value, given in equation (3), parametric hypotheses about the spatial autocorrelation level (often named ρ) can be tested. The z-values are simply compared with the well known critical values of the normal distribution.

One thing to remember, the Moran's $\Im$ is a measure or test for global spatial autocorrelation. That is, if there are different spatial structures inherent in the data, e.g. some regions have a positive spatial autocorrelation and some others have a negative spatial autocorrelation, these effects can compensate each other, and the global Moran's $\Im$ indicates spatial independence - although there is a local spatial autocorrelation included. Local effects can be detected and tested via local Moran's $\Im$, which can be calculated for each region in the dataset. Therefore, a modified spatial link matrix is used, for each region i the corresponding spatial link matrix is a star-shaped matrix W_i with the ith row and ith column of the global spatial link matrix W (all other elements are zero). For the local Moran's $\Im$ (e.g. for region i) the formula is nearly the same like the one given in equation (2), only that the local spatial link matrix W_i is used instead of the global link matrix W. The sum over all local Moran's $\Im$ gives the global Moran's $\Im$.

2.4 Spatial Regression Models

Under the assumption of a spatial effect inherent in the data, there are different possibilities to specify this spatial dependency in a linear regression model. It can be included either as an additional regressor or in the error structure. The spatial error model e.g. is appropriate when spatial data are used and the potential influence of the spatial autocorrelation should be corrected. The spatial error model depends on the specification of the spatial structure, which is expressed by the covariance matrix of the error term. A popular specification of the spatial structure is the spatial autoregressive (SAR) process, which is a functional relationship between a random variable at a given location and this same random variable at other locations. Here a spatial lag operator Wy, which is simply a weighted average of random variables at neighboring locations (also called a spatial smoother), is used, W is a $n \times n$ spatial link matrix and y a $n \times 1$ vector of random variables. If centered variables are considered ($y = y^* - \mu 1_n$, where μ is the common mean of the random variables y_i^* and 1_n is the $n \times 1$ vector of ones), the process can be defined as a simultaneous SAR process

$$y = \rho W y + \varepsilon = (I_n - \rho W)^{-1} \varepsilon \, ,$$

where I_n is the $n \times n$ identity matrix, ε are i.i.d. zero mean error terms with common variance σ^2, and ρ is the autoregressive parameter (in most cases $|\rho| \le 1$). The variance-covariance matrix of y is a function of the noise variance σ^2 and the spatial coefficient ρ,

i.e.

$$\Omega(\rho) = \text{cov}(y, y) = \text{E}[yy'] = \sigma^2[(I_n - \rho W)'(I_n - \rho W)]^{-1}.$$

For further processes and more detailed explanations see e.g. Anselin (1999). The SAR error model has the form

$$y = X\beta + u, \qquad u = \rho W u + \varepsilon. \tag{4}$$

The error variance covariance matrix is no longer $\sigma^2 I_n$ like it is in the linear regression model under standard assumptions, but

$$\text{E}[uu'] = \Omega(\rho) = \sigma^2[(I_n - \rho W)'(I_n - \rho W)]^{-1}. \tag{5}$$

For further spatial model specification see e.g. Anselin (1999).

2.5 Spatial Estimation

One problem when analyzing spatial data with standard statistical methods is the following: If the observations are spatially connected or spatially autocorrelated, the standard assumptions of uncorrelated error terms and uncorrelated observations and errors are violated. This can lead to inconsistent, inefficient and biased estimators. Therefore special estimation techniques, where the spatial dependency is adequately included in the estimation, should be used. There are different estimation methods for spatial data, one can e.g. use the Maximum Likelihood technique (first outlined by Ord, 1975), or a Spatial Two Stage Least Squares method based on Instrumental Variable estimations (see e.g. Kelejian and Robinson, 1993, or Kelejian and Prucha, 1998), or based on a Method of Moments (Kelejian and Prucha, 1999), which is described in more detail below.

Kelejian and Prucha (1999) suggest to use the following procedure for the estimation of a spatial autoregressive model, given in equation (4), with a covariance matrix, given by equation (5). The auxiliary parameters ρ and σ^2 are estimated via the generalized method of moments technique, the generalized moments (GM) estimator of ρ and σ^2 is a non-linear least squares estimator

$$(\tilde{\rho}, \tilde{\sigma}^2) = \arg\min_{\rho, \sigma^2} \left\{ \left[(\rho, \rho^2, \sigma^2)\Gamma - \gamma \right] \left[(\rho, \rho^2, \sigma^2)\Gamma - \gamma \right] \right\}, \tag{6}$$

where $\rho \in [-a, a]$ with $a \geq 1$ and $\sigma^2 \in [0, b]$. The matrix Γ and the vector γ are both functions of the OLS residuals derived from the moment conditions, and $((\rho, \rho^2, \sigma^2)\Gamma - \gamma)$ can be seen as a vector of residuals. For a detailed specification of the functions see Kelejian and Prucha (1999, p. 8). Estimates given in (6) converge in probability to the true parameters ρ and σ^2 under certain assumptions (see Kelejian and Prucha, 1999, p. 5), one of these assumptions deals with the spatial weight matrix. For row-standardized spatial weight matrices, which are used in the following R&D analysis, they expect these assumptions to hold. The parameter β of the regression model is then a feasible generalized least squares (FGLS) estimator

$$\tilde{\beta} = [X'\tilde{\Omega}^{-1}X]^{-1}X'\tilde{\Omega}^{-1}y, \tag{7}$$

where $\tilde{\Omega} = \Omega(\tilde{\rho}, \tilde{\sigma}^2)$.

3   R&D Spillover: A Spatial Examination

Looking at the R&D dataset, for a spatial analysis the countries are regarded as regions. The first question is: How to measure the distance or contiguity between the observations at different locations in an adequate way? In a global economy not the geographic distance but rather the trade intensity between two countries is relevant for R&D spillovers. To be consistent with Coe and Helpman (1995), the bilateral import shares (of the year 1990), also available on the webpage of Helpman (2003), are used as a row-standardized spatial link matrix, denoted by V. This asymmetry in the spatial link matrix is a problem if we want to define some kind of economic distances. Therefore a symmetric trade intensity was specified and used to measure the contiguity and consequently the distance between economies. In this context the symmetric trade intensity between two countries is defined as the average of the bilateral import-shares of these countries, the elements are simply calculated by

$$w_{ij} = \frac{b_{ij} + b_{ji}}{2}$$

if $i \neq j$, and b_{ij} are the bilateral import-shares of country i from country j in period 1990, and by definition $w_{ij} = 0$ for $i = j$. It was assumed that the trade intensity is the same for all periods, this means the same spatial link matrix is used for all years. The distances between two countries are simply the inverse connectivity

$$d_{ij} = \frac{1}{w_{ij}}$$

and by definition $d_{ii} = 0$. These distances can be used to produce a "trade-intensity" landscape by projecting the distances from the 21-dimensional space to the two-dimensional space. For this projection a Multidimensional Scaling method is used: the squared sums of the distances between the original and the projected points (the points represent the countries) are minimized. This gives an approximation of all 231 distances between the 22 countries in the two-dimensional space, and provides a quite good survey of the relationships in the data set (see Figure 1). Here the countries are quite evenly scattered, nevertheless some clusters can be identified, e.g. Australia, New Zealand and Israel are quite far apart from the rest of the countries, this means they have a small trade intensity with other countries and a relative high trade intensity within their group. The U.S. are settled in the center, it can be interpreted in the way that the U.S. are an important trade partner for all countries. One thing to remember when looking at this landscape is, it is only an approximation and it can never show the true and exact distances.

3.1   A Spatial Approach for the Analysis of R&D Spillover

The spatial link matrix for the spatial regression model is the original row-standardized bilateral import-shares matrix V from Coe and Helpman's dataset. The first steps in the spatial analysis are the estimation of a fixed effects model without any foreign R&D spending and without any spatial structure assumed

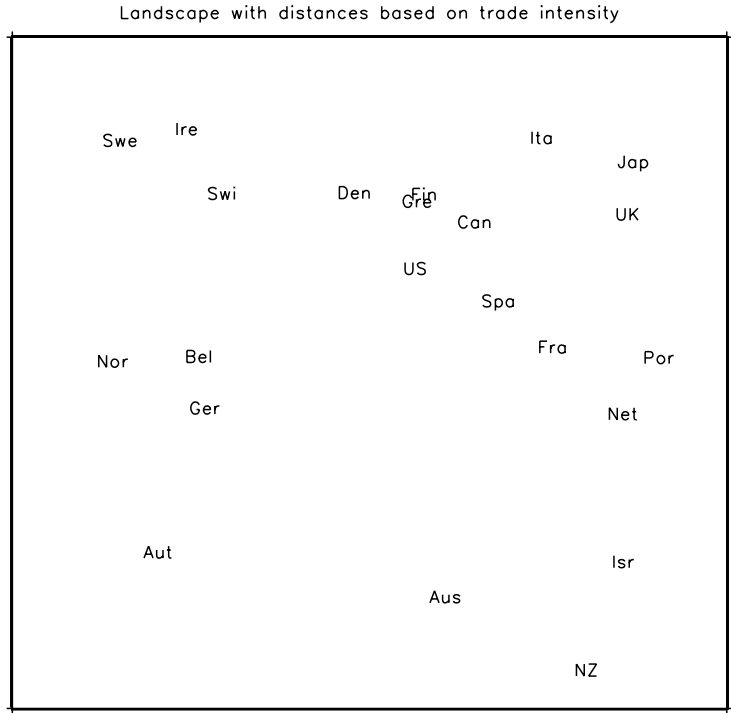$$\widehat{\log F_{it}} = \alpha_{it}^0 + \underbrace{0.136}_{0.000} \log S_{it}^d \tag{8}$$

Landscape with distances based on trade intensity

Swe   Ire   Ita
            Jap
         Swi   Den   Fin
              Gre
                        Can   UK
                    US
                         Spa
Nor   Bel   Fra   Por
       Ger   Net

Aut   Isr
                    Aus
                         NZ

Figure 1: Landscape based on trade-intensities between the countries.

and to calculate and test Moran's $\Im$ for the residuals of this model for each period separately, see equations (2) and (3). Again, as spatial link matrix the bilateral import shares (matrix V) are used. Nearly all values are not significant (see Table 1), this means there seems to be no global spatial effect in the error term. Nevertheless, some local spatial effects can be detected. Moreover, as there are only 22 countries in the dataset one should not put too much weight on the Moran's test because $z(\Im)$, given in equation (3) is only approximately normally distributed. Tiefelsdorf (2000, p. 97) recommends to use this test for datasets with at least 100 observations for exploratory statistical analysis and at least 200 observations for confirmatory statistical analysis. The assumption of some spatial effect is legitimate because the effect of FRD, which measures some kind of spatial dependency, is significant in the original model (1). Under the assumption of a spatial effect included in the error term, one should use an adequate estimation technique for the SAR error regression model, given in equation (4), e.g. the FGLS estimation from Kelejian and Prucha (1998), see equations (6) and (7). This leads to similar results as the non-spatial analysis, namely $\hat{\alpha}_{it}^d = 0.138$ with p-value 0.000 and the auxiliary parameters, estimated with the GM method, are $\hat{\rho} = 0.137$ and $\hat{\sigma}^2 = 0.003$.

A fixed effects SAR error model including the foreign R&D spending is estimated to compare the results with the ones from Coe and Helpman (1995). This spatial model yields

$$\widehat{\log F_{it}} = \alpha_{it}^0 + \underbrace{0.096}_{0.000} \log S_{it}^d + \underbrace{0.096}_{0.000} \log S_{it}^f$$

Table 1: Moran's $\Im$ for residuals of Fixed Effect model, independent variable $\log S_{it}$.

Period	Moran's $\Im$	$z(\Im)$	Period	Moran's $\Im$	$z(\Im)$
1990	0.0532	1.3818	1980	−0.0323	0.2101
1989	−0.1777	−1.7822	1979	−0.0860	−0.5263
1988	−0.1184	−0.9693	1978	−0.0426	−0.0694
1987	−0.0607	−0.1789	1977	0.0623	1.5064
1986	−0.0330	−0.1789	1976	−0.1337	−1.1788
1985	−0.0258	0.2993	1975	−0.0584	−0.1480
1984	−0.0702	−0.3088	1974	−0.0473	0.0042
1983	−0.0101	0.5144	1973	0.0198	0.9241
1982	0.0506	1.3451	1972	−0.0053	0.5795
1981	0.0910	1.8989	1971	−0.0328	0.2026

$\mathrm{E}(\Im) = -0.0476$ and $\mathrm{var}(\Im) = 0.0053$ for all periods

with the auxiliary parameters $\hat{\rho} = 0.164$ and $\hat{\sigma}^2 = 0.002$. The results are quite similar, both parameter estimators are positive and significant, in the spatial analysis as well as in the non-spatial one. The fit of the SAR error model is with $R^2 = 0.580$ a bit better than the non-spatial one with $R^2 = 0.558$. The standardized Moran's $\Im$ of the residuals, which indicates the magnitude of the spatial dependency not captured by the variables in the model, is higher for the non-spatial model ($z(\Im) = 0.361$) than for the spatial model ($z(\Im) = 0.141$), as expected.

Another alternative to analyze the R&D dataset spatially is the following: The foreign R&D spending can be regarded as spatially lagged domestic R&D spending, i.e. $S_{ij}^f = \sum_{i \neq j} b_{ijt} S_{jt}^d$. To avoid the logarithms of the independent variables and as all of the values of S_{it}^d are around one, a Taylor Series approximation can be employed for the logarithm, i.e. $\log S = \log(1) + \frac{1}{1}(S - 1) + \dots$. Therefore, $\log S_{it}^d$ can be substituted by the approximation $\log S_{it}^d \simeq S_{it}^d - 1$, and $\log(\sum_{i \neq j} b_{ijt} S_{jt}^d)$ can be substituted by $\sum_{i \neq j} b_{ijt} S_{jt}^d - 1$. This leads to the model

$$\log F_{it} = \tilde{\alpha}_{it}^0 + \alpha_{it}^d S_{it}^d + \alpha_{it}^f \sum_{i \neq j} b_{ijt} S_{jt}^d + \varepsilon_{it}, \qquad (9)$$

where the fixed effects change to $\tilde{\alpha}_{it}^0 = \alpha_{it}^0 - \alpha_{it}^d - \alpha_{it}^f$. In a first approach the fixed effects panel regression, given in equation (9) is estimated by LSDV, which gives positive and significant parameter estimators for the effect of DRD as well as FRD.

$$\widehat{\log F_{it}} = \tilde{\alpha}_{it}^0 + \underbrace{0.067}_{0.000} S_{it}^d + \underbrace{0.179}_{0.000} \sum_{i \neq j} b_{ijt} S_{jt}^d. \qquad (10)$$

The fit of this model is with $R^2 = 0.624$ a bit better than the one for the model (1) without a spatial lag. The standardized Moran's $\Im$ of the residuals is $z(\Im) = 0.255$ which is smaller than the one for the residuals of model (1) which is $z(\Im) = 0.361$.

Under the assumption of a SAR error model, where a spatial effect is included in the error term, see equation (4), a FGLS estimation based on GM estimators of the autoregressive parameter $\hat{\rho} = 0.228$ and the noise variance $\hat{\sigma}^2 = 0.002$ (using equation (6) and

(7)) leads to

$$\widehat{\log F_{it}} = \tilde{\alpha}_{it}^0 + \underbrace{0.141}_{0.000} S_{it}^d - \underbrace{0.050}_{0.031} \sum_{i \neq j} b_{ijt} S_{jt}^d.$$

The result diverges from the one of the non-spatial analysis, the effect of DRD on TFP is again positive and significant but the effect of FRD on TFP is negative and significant. On the other hand, the fit of this model yields a worse $R^2 = 0.269$, and gives a negative $z(\Im) = -0.506$, even though it is not significant. Nevertheless, these values indicate an overcompensation of the spatial effect, due to the fact that the spatial dependency is included twice, once as the spatially lagged variable DRD and once in the error term.

However, as all of the critics of the original, non-spatial R&D spillovers analysis are also legitimate in the spatial context, all different more sophisticated models (namely the dynamic fixed effects, the static random coefficients and finally the dynamic random coefficients one) were estimated via OLS and FGLS and the results can be found in D. Gumprecht (2005).

The method of choice should again be the DOLS estimation of the random coefficients model. For the original variables DRD and FRD, the SAR error model should be used to correct for a spatial effect. The FGLS estimation yields

$$\widehat{\log F_{it}} = \tilde{\alpha}_{it}^0 + \underbrace{0.252}_{0.000} \log S_{it}^d - \underbrace{0.016}_{0.393} \log S_{jt}^f$$

with $R^2 = 0.956$, and estimates of the auxiliary parameter $\hat{\rho} = 0.375$ and $\hat{\sigma}^2 = 0.007$; $z(\Im)$ of the residuals is -0.060. Concerning the parameters, we have the same result as in the non-spatial case: A positive effect of DRD and no spillover effect of FRD.

Now, using the approximated variables instead of the original ones and running the FGLS estimation yields $\hat{\rho} = 0.720$ and $\hat{\sigma}^2 = 0.003$. This leads to non-significant parameter estimates

$$\widehat{\log F_{it}} = \tilde{\alpha}_{it}^0 + \underbrace{0.081}_{0.076} S_{it}^d + \underbrace{0.016}_{0.401} \sum_{i \neq j} b_{ijt} S_{jt}^d$$

with $R^2 = 0.960$ and $z(\Im) = -0.184$. Neither the effect of DRD nor the effect of FRD is significant. The unusual high value of $\hat{\rho}$ indicates overcompensation. This is caused by the fact, that the spatial effect is already included as spatially lagged independent variable and an additional spatial effect in the error term leads to an overcompensation (like in the case of the fixed effects model).

Thus, the preferred method is the DOLS estimation of the random coefficients model with approximated variables, which yields

$$\widehat{\log F_{it}} = \tilde{\alpha}_{it}^0 + \underbrace{0.125}_{0.011} S_{it}^d + \underbrace{0.166}_{0.015} \sum_{i \neq j} b_{ijt} S_{jt}^d$$

with $R^2 = 0.976$ and $z(\Im) = -0.191$. This model has the best fit of all examined models and the result is in consensus with the original conclusions from Coe and Helpman (1995).

4 Conclusions

In general, one of the advantages of using spatial models and methods is, that a spatial dependency that might be inherent in empirical data, can be taken into account and treated

correctly. And even if there is already a spatial dependency assumed, one can correct further spatial relationships that might no be captured by the variables in the model, by using a spatial error model. Especially when there is a spatial link matrix available, that describes the relationship between the observations, it is no problem to use adequate models and estimation techniques. The price one pays for running a spatial analysis is much less than the benefit one can earn by getting unbiased and consistent estimates.

The aim of the analysis of the R&D spillover data set was to answer the question, whether domestic and foreign R&D spending have an effect on the total factor productivity of a country. Concerning domestic R&D spending the answer is quite obvious, all different estimation techniques (static and dynamic fixed effects- and random coefficients model) and both non-spatial and spatial approach lead to the conclusion that domestic R&D spending have a positive effect on the total factor productivity of a country. Concerning the foreign R&D spending the answer is not that clear, because different estimation techniques lead to different conclusions. Some results support the conclusion in Coe and Helpman (1995) of an R&D spillover effect, some do not. Nevertheless if one takes the dynamic random coefficients model with a spatially lagged exogenous variable as the superior specification, the effect of foreign R&D expenditures seems to be existent.

Acknowledgments

References

Anselin, L. (1999). *Spatial Econometrics* (Tech. Rep.). University of Texas at Dallas, Richardson, TX: Bruton Center, School of Social Siences.

Coe, D., and Helpman, E. (1995). International R&D spillovers. *European Economic Review*, *39*, 859-887.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression*. New York: Wiley.

Greene, W. H. (2003). *Econometric Analysis*. New Jersey: Prentice Hall.

Griffith, D. A. (2003). *Spatial Autocorrelation and Spatial Filtering*. Berlin: Springer.

Grossman, G. M., and Helpman, E. (1991). *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.

Gumprecht, D. (2003). *Ein Panel Kointegrationsmodell für internationale Forschungs- und Entwicklungs Spillovers*. Unpublished master's thesis, University of Vienna.

Gumprecht, D. (2005). *Spatial Methods in Econometrics: An Application to R&D Spillovers* (Tech. Rep. No. 26). Vienna University of Economics and Business Administration.

Gumprecht, D., Gumprecht, N., and Müller, W. G. (2004). Some Current Issues in the Statistical Analysis of Spillovers. In G. Meier and S. Sedlacek (Eds.), *Spillovers and Innovations: City, Environment, and the Economy* (Interdisciplinary Studies in Economics and Management ed., Vol. 4, p. 51-70). Wien: Springer.

Gumprecht, N. (2003). *Regression mit zufälligen Koeffizienten: Software Implementierung*. Unpublished master's thesis, University of Vienna.

Helpman, E. (2003). *Professor Elhanan Helpman's data on the web.* Retrieved January 24, 2003, from Harvard University, Economics Department Web site: http://post.economics.harvard.edu/faculty/helpman/data.html.

Kao, C., Chiang, M., and Chen, B. (1999). International R&D spillovers: An application of estimation and inference in panel cointegration. *Oxford Bulletin of Economics and Statistics*, *61*, 693-711.

Kelejian, H., and Prucha, I. (1998). A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances: A serious problem. *International Regional Science Review*, *20*, 103-111.

Kelejian, H., and Prucha, I. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40*, 509-533.

Kelejian, H., and Robinson, D. P. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a country expenditure model. *Papers in Regional Science*, *72*, 297-312.

Müller, W. G., and Nettekoven, M. (1999). A panel data analysis: Research and development spillover. *Economics Letters*, *64*, 37-41.

Moran, P. A. P. (1948). The interpretation of satistical maps. *Biometrika*, *35*, 255-260.

Moran, P. A. P. (1950a). Notes on continuous stochastic phenomena. *Biometrika*, *37*, 17-23.

Moran, P. A. P. (1950b). A test for the serial dependence of residuals. *Biometrika*, *37*, 178-181.

Ord, J. K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*, 120-126.

Romer, P. M. (1990). Endogenous technical change. *Jounal of Political Economy*, *98*, 71-102.

Stock, J. H. (1987). Asymptotic properties of least squares estimations of co-integrating vectors. *Econometrica*, *55*, 1035-1056.

Tiefelsdorf, M. (2000). *Modelling Spatial Processes*. Berlin: Springer.

Author's address:

Daniela Gumprecht
Department of Statistics and Mathematics
Vienna University of Economics and Business Administration
Augasse 2-6
A-1090 Vienna, Austria

Tel. +43 1 31336 / 5032
Fax +43 1 31336 / 774
E-mail: `daniela.gumprecht@wu-wien.ac.at`