# A spatial panel data version of the knowledge capital model

Daniel Bekesi, Christoph Hammer, Matthias Koch*, Christian Sommeregger

*Vienna University of Economics and Business; *mailto: Matthias.Koch@wu.ac.at*

**Abstract**

This paper attempts to analyze the impact of knowledge and knowledge spillovers on regional total factor productivity (TFP) in Europe. Regional patent stocks are used as a proxy for knowledge, and TFP is measured in terms of a superlative index. We follow Fischer et. al (2008) by using a MAR-spillover model and a data set covering 203 regions for six time periods. In order to estimate the impact of knowledge stocks we use a spatial autoregressive model with random effects, which allows for three kinds of spatial dependence: Spatial correlation in the innovations, the exogenous and the endogenous variables. The results suggest that there is a significant positive impact of knowledge on regional TFP levels, and that knowledge spills over to neighboring regions. These spillovers decay exponentially with distance at a rate of 8%. Using Monte Carlo simulations we calculate the distribution of direct and indirect effects. The average elasticity of a region's TFP with respect to its own knowledge stock is 0.1 and highly significant. The average effect of all other regions' TFP is about 50% higher, which confirms that the cross-country externalities are important in the measuring of the impact.

This paper will be presented at the European Regional Summit in Barcelona 2011 and the Annual Meeting of the Austrian Economic Association in Graz 2011.

*JEL Classifications:* O40, C59

*Key words:* knowledge spillover, total factor productivity, nonlinear spatial panels

## 1 Introduction

This paper attempts to analyze the impact of knowledge on regional total factor productivity (TFP) in Europe. Regional TFP can be seen as a proxy for technological progress and hence is one key determinant in regional economic growth. This study focuses on the empirical relationship between knowledge and TFP by allowing for three kinds of spatial dependence: Spatial correlation in the innovations, the exogenous and the endogenous variables.

The model is based on the approach presented by Fischer, Scherngell and Reisman (2008), which will be labeled as FSR- model. The FSR- model is itself based on the knowledge capital model introduced by Griliches (1979), that "has been applied in hundreds of empirical studies on firm-level productivity and also extended to macroeconomic growth models" (Doraszelski and Jaumandreu 2006: p2). To ensure that the various agents, individuals, networks and enterprises, which are involved in the innovative process are integrated into a coherent unit, the FSR- Model treats regions as the unit of observation. One innovation of the FSR model was to add regional external knowledge stocks to the Griliches model, recognizing that spillover are in close relation with the production of knowledge. FSR 2008 shows that these spillovers are strongly localized.

The current paper explores the model provided by Fischer et al (2008), by including a spatial lag of the TFP. This is motivated by Le Sage and Pace (2009) where they argue that a spatial lag can be caused be by time dependence, can reduce omitted variable bias and model uncertainty. Hence, ignoring the possible spatial lag can cause biased estimations. Incorporating a spatial lag in the TFP- variable makes it a little bit more complicated to distinguish between direct and indirect effects. We find out that the average elasticity of a region's TFP with respect to a change to the other regions' knowledge stocks is almost two times higher that the average elasticity with respect to its own knowledge stock.

We use the same data as in Fischer et al. (2008) [1], which consists of 203 European NUTS-2 regions covering the 15 pre-2004 European Union members during 1997-2002. Since the regional policy of the European Union is essentially based on the NUTS-2 region level, NUTS-2 represents for our model a useful geographic scale.

The remainder of this paper is organized as follows: In the second section we will briefly review the FSR model and introduce the extension mentioned earlier. The third section is dedicated to the estimation and interpretation of the model. Here the econometric specification and estimation strategy are described in more detail. Furthermore the calculation of direct and indirect effects will be discussed. The fourth section describes the used data, explains how relevant variables are constructed and ends by presenting our empirical results. Finally section five summarizes our findings and concludes the article.

[1] The data set was supplied by the Institute for Economic Geography and GIScience (Vienna University of Economics and Business)

## 2 The model approach

This section provides the background for our empirical model. First it starts by revisiting the empirical formulation of the FSR model and provides some conceptual background. For more details regrading the FSR model see Fischer et al (2008). The second subsection adds a spatial lag of the dependent variable to the empirical form of the FSR- model.

### *2.1 Basic Model*

We use the FSR- model as our model basis and start with Eq. (8) from FSR (2008) in log-additive form:

$$\log(f_{it}) = \log(k_{it})\beta_1 + \log(\sum_{j\neq i}^{n} w_{ij}(\delta)k_{j-1})\beta_2 + \varepsilon_{it} \tag{1}$$

where $1 \leq i \leq n$ and $1 \leq t \leq T$ or in matrix notation:

$$\log(\mathbf{f}) = \log(\mathbf{k})\beta_1 + \log(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1})\beta_2 + \boldsymbol{\varepsilon}.$$

Here $f_{it}$ denotes the TFP of region $i$ at time $t$, which is assumed to be generated by a Cobb-Douglas styled function. This function depends on the region-internal $k_{it}$ and spatially weighted region-external knowledge stocks $\sum_{j\neq i}^{n} w_{ij}(\delta)k_{j-1}$, which may be also interpreted as the pool of extra-regional knowledge productive in region $i$. As the literature suggests, external knowledge does not become effective in production immediately [ref]. Here, we assume a one period time lag. We further note that this specific functional form implies that $\beta_1$ and $\beta_2$ are the elasticities corresponding to internal and external knowledge capital.

Another noteworthy property of the FSR- model regards the definition of $w_{ij}$.

$$w_{ij}(\delta) = \exp(-\delta d_{ij}) \text{ where } \delta \geqq 0 \tag{2}$$

As can be seen from Eq. (2), it is assumed that $w_{ij}$ is an exponential function of the distance between $i$ and $j$ and the decay constant $\delta$. $d_{ij}$ is the distance between region $i$ and $j$ measured as great circle distance between their economic centers. If $\delta = 0$ then $w_{ij} = 1$ and distance does not matter for the ability to internalize external knowledge. However, if $\delta > 0$ then distance matters: The greater the distance between $i$ and $j$, the smaller $w_{ij}$ and therefore region $j's$ contribution to region $i's$ external knowledge pool. Consequently, this setup implies that knowledge is exposed to spatial depreciation with a constant rate $\delta$.

### 2.2 Extended Model

The novel feature of our paper is the incorporation of a first-order spatial lag in dependent variable. Therefore the regional TFP is no longer exclusively driven by the knowledge stocks of the regions and the spatially lagged extra-regional knowledge stocks, but also by the spatially lagged TFP. LeSage and Pace (2009) give three possible motivations for this spatial lag: time dependence, omitted variable bias and model uncertainty. Each of these motivations can be seen as relevant for the FSR model. Hence it seems necessary to incorporate a spatially lagged TFP because a failure to do so would lead to biased estimations. Therefore one should treat the estimates of the FSR model with caution.

Adding a spatial lag of the dependent variable to the FSR- model yields Eq. (3).

$$\log(\mathbf{f}) = \rho\overline{\mathbf{M}}\log(\mathbf{f}) + \log(\mathbf{k})\beta_1 + \log(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1})\beta_2 + \boldsymbol{\varepsilon} \tag{3}$$
$$\text{where } \overline{\mathbf{M}} = 1/\overline{m}\,(\mathbf{I}_T \otimes \mathbf{M})\,, \overline{\mathbf{W}(\delta)} = \mathbf{I}_T \otimes \mathbf{W}(\delta).$$

$\log(\mathbf{f})$ denotes the stacked vector of the logarithmic TFPs cross-sections with dimensions $nT \times 1$. Likewise $\log(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1})$ and $\log(\mathbf{k})$ are the $nT \times 1$ vectors of the spatially lagged and non-lagged capital-stocks. $\mathbf{I}_T$ denotes the identity matrix of dimension $T$. We use two different $n \times n$ weight matrices: $\mathbf{W}(\delta)$ and $\mathbf{M}$. The typical element of $\mathbf{W}(\delta)$ is defined by (2). Note that $\mathbf{W}(\delta)$ is a function of the decay parameter $\delta$ and therefore Eq. (3) has a no linear representation. $\mathbf{M}$ denotes a binary spatial weight matrix, where regions with a common border take the value of 1 and 0 otherwise. Assuming constant weights throughout time we are able to define the matrices $\overline{\mathbf{W}(\delta)}$ and $\overline{\mathbf{M}}$ respectively, where we make use of the convenient Kronecker operator $\otimes$. $\rho$ is the autoregressive parameter and it is assumed that $|\rho| < 1$. The rows of $\overline{\mathbf{M}}$ are normalized by the largest absolute characteristic root of $\mathbf{M}$ denoted by $\overline{m}$. Due to our parameter space restriction we can solve Eq. (3) for $\log(\mathbf{f})$:

$$\log(\mathbf{f}) = (\mathbf{I}_T \otimes \mathbf{A}(\rho)^{-1})\left[\log(\mathbf{k})\beta_1 + \log(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1})\beta_2 + \boldsymbol{\varepsilon}\right]$$

$$\text{where } \mathbf{A}(\rho) = \mathbf{I}_n - \rho/\overline{m}\mathbf{M}$$

Hence if $\rho \neq 0$ we can no longer interpret $\beta_1$ as elasticity like we do in Eq. (1), since $\frac{\partial \log(f_{it})}{\partial \log(K_{it})} \neq \beta_1$. This problem will be addressed in Section 3.3. Additionally if $\rho = 0$, then $(\mathbf{I}_T \otimes \mathbf{A}^{-1}) = \mathbf{I}_{Tn}$ and the model reduces to the basic model.

## 3 Model estimation and interpretation

This section provides the necessary background for the model estimation and interpretation. Due to the nonlinearity of our model, caused by $\overline{\mathbf{W}(\delta)}$ we had to program the Maximum Likelihood estimator ourself. If faced with this task the programming of the log-determinant becomes crucial. Therefore, we first focus in the next subsection on the specified error term and calculate the variance- covariance matrix of the TFP given the data and then discuss the computational issues implied by the error structure. The last subsection discusses the necessary background for the model interpretation.

### *3.1 Specification of the Error Term*

We now incorporate the third kind of spatial dependence into our model, reflecting possible spatial dependence in the error term. Note that the data has a panel structure. Therefore, we use a random effects model. Additionally remark that unlike in the case of the linear regression model, ignoring the structure of the variance- covariance matrix will lead in spatially autocorrelated models to biased estimations if Maximum Likelihood is used. Hence we use the error structure given by Eqs. (4) and (5)

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\mu} + \boldsymbol{\zeta}_t \text{ where } \boldsymbol{\mu} \sim N(0, \mathbf{I}_n \sigma_\mu^2) \tag{4}$$

$$\boldsymbol{\zeta}_t = \lambda / \overline{m} \mathbf{M} \boldsymbol{\zeta}_t + \boldsymbol{\eta}_t \text{ where } \boldsymbol{\eta}_t \sim N(0, \mathbf{I}_n \sigma_\eta^2). \tag{5}$$

The vectors in Eqs. (4) and (5) now denote one cross-section at time $t$ where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, ...\varepsilon_{n,t})'$, $\boldsymbol{\eta}_t = (\eta_{1,t}, ...\eta_{n,t})'$ and $\boldsymbol{\mu}_t = (\mu_{1,t}, ...\mu_{n,t})'$. Both innovations $\boldsymbol{\mu}_{it}$ and $\boldsymbol{\eta}_{it}$ are independently and normal distributed with the corresponding variances $\sigma_\mu^2$ and $\sigma_\eta^2$. The used weight matrix $\mathbf{M}$ is the same[2] as in Eq. (3). Since we assume that $|\lambda| < 1$ we can solve Eq. (5), put it into Eq. (4) and get Eq. (6):

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\mu} + \mathbf{A}(\lambda)^{-1} \boldsymbol{\eta}_t \tag{6}$$

Note that for $\lambda = 0$ the specification reduces to a standard random effects model. Consequently, the stacked vector $\boldsymbol{\varepsilon}$ can be written as:

$$\boldsymbol{\varepsilon} = \boldsymbol{\iota}_T \otimes \boldsymbol{\mu} + \left( \mathbf{I}_T \otimes \mathbf{A}(\lambda)^{-1} \right) \boldsymbol{\eta}$$

where $\boldsymbol{\iota}_t$ denotes a row vector of ones with the length $T$. Under these assumptions the variance- covariance matrix of the TFP given the knowledge stocks is:

$$\boldsymbol{\Omega} := E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) = \left( \mathbf{I}_T \otimes \mathbf{A}(\rho)^{-1} \right) \left[ (\boldsymbol{\iota}_T' \boldsymbol{\iota}_T \otimes \mathbf{I}_n \sigma_\mu^2 + (\mathbf{I}_T \otimes (\mathbf{A}(\lambda)' \mathbf{A}(\lambda))^{-1} \sigma_\eta^2 \right] \tag{7}$$

[2] There has been pointed out that for models like...with MC- Studies that ML results in biased estimates. For example Bivand [SEA 2010] pointed out... Observe that the bias only occurs if both weigth maritices are the same and no pannel structure is assumed. In order to verify our programing we conducted also MCs and it seems as long as $\sigma_\mu$ and $\sigma_\eta$ are different, that the Bias fanishes. Since the hypothesis $\sigma_\mu = \sigma_\eta$ has a p- value of... we conclude that our estimation method is unbiased.

### 3.2 Computational Issues: the Log-Determinant

Due to the fact that the calculation of the log-determinant of a 1213x1213 matrix can be considerably resource-consuming we carried out some simplifications: Following the suggestions of Wansbeek and Kapteyn (1983) the determinant of our variance-covariance matrix can be written as

$$|\mathbf{\Omega}| = \sigma_\eta^2 \left|(\mathbf{A}(\rho)'\mathbf{A}(\rho))^{-1}\right|^T \left|\frac{T\sigma_\mu^2}{\sigma_\eta^2}\mathbf{I}_n + (\mathbf{A}(\lambda)'\mathbf{A}(\lambda))^{-1}\right| \left|(\mathbf{A}(\lambda)'\mathbf{A}(\lambda))^{-1}\right|^{T-1}. \tag{8}$$

Following Griffith (1988) we further simplify this expression by making use of:

$$|\mathbf{A}(x)| = \prod_{i=1}^{n}\left(1 - x\frac{m_i}{\overline{m}}\right) \tag{9}$$

$$\left|\frac{T\mu}{\eta}\mathbf{I}_n + (\mathbf{A}(\lambda)'\mathbf{A}(\lambda))^{-1}\right| = \prod_{i=1}^{n}\left(\frac{T\sigma_\mu^2}{\sigma_\eta^2} + \left(1 - \lambda\frac{m_i}{\overline{m}}\right)^{-2}\right)$$

where $m_i$ denotes the eigenvalues of $\mathbf{M}$. In a final step we put Eq. (9) and Eq. (8) together and take logs to end up with:

$$\log(|\Omega|) = n\log(\sigma_\eta^2) + (-2T)\log(\boldsymbol{\iota}_n'\mathbf{r}(\rho)) + \log\left(\frac{T\sigma_\mu^2}{\sigma_\eta^2} + 1/\mathbf{r}(\lambda)'\mathbf{r}(\lambda)\right) - 2(T-1)\log(\mathbf{r}(\lambda)'\mathbf{r}(\lambda)) \tag{10}$$

where $\mathbf{r}(x) = \boldsymbol{\iota}_n - \frac{x}{\overline{m}}\mathbf{m}$. Here $\mathbf{r}$ is a $n$ by 1 vector and $\mathbf{m}$ is the vector containing the eigenvalues of $\mathbf{M}$. Since the innovations are assumed to be drawn from a normal distribution, we now can write the log likelihood function as:

$$LL(\rho, \lambda, \beta_1, \beta_2, \delta, \sigma_\eta^2, \sigma_\mu^2) = \frac{-nT}{2}\log(2\pi) - \frac{1}{2}\log(|\Omega|) - \frac{1}{2}\left(\log(\mathbf{f}) - \widehat{\log(\mathbf{f})}\right)'\Omega^{-1}\left(\log(\mathbf{f}) - \widehat{\log(\mathbf{f})}\right) \tag{11}$$

where $\widehat{\log(\mathbf{f})}$ are the fitted values of $\log(\mathbf{f})$ for given values of $\rho, \lambda, \beta_1, \beta_2, \delta, \sigma_\eta^2, \sigma_\mu^2$. To further improve computational efficiency we concentrated out the Parameters $\beta_1$ and $\beta_2$. As can be seen from Eq. (7) the property of homoscedasticity does not hold, consequently the usual OLS estimator would be inefficient. In order to estimate $\beta_1$ and $\beta_2$ efficiently we use in our optimization routine the standard GLS estimator.[3]

$$\hat{\beta}(\rho, \lambda, \delta, \sigma_\eta^2, \sigma_\mu^2) = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\log(\mathbf{f}) \tag{12}$$

[3] All estimations were carried out within the Matlab Software Package. For the maximization of the loglikelihood-function we used the fminsearch procedure, the Hessian was computed with help of the DERIVEST package. For further details on the utilized methods have a look at the appendix.

$$\text{where } \mathbf{X} = [\log(\mathbf{k}), \log(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1})]$$

### *3.3 Interpretation of the Model*

In this section we provide a (possible) interpretation of the model given in Eq. (3). Like we briefly mentioned in 2.2, unlike to the basic model the coefficients $\beta_1$ and $\beta_2$ for $\rho \neq 0$ can not longer be interpreted as elasticity, since for Eq. (3): $\frac{\partial \log(f_{it})}{\partial \log(k_{it})} \neq \beta_1$ and $\frac{\partial \log(f_{it})}{\partial \log(Wk_{it-1})} \neq \beta_2$. Unlike to a simple linear model estimated with OLS, we have to consider two important issues regarding the interpretation of the model coefficients:

*First*, due to the spatial dependence of the dependent variable, each region's TFP has different elasticities with respect to a change in any region's knowledge stock. Hence our model provides 41.209 different elasticities regarding a change in the knowledge stock of period $t$. This interpretation issue is pointed out by LeSage and Penn (2009: p34). We follow their suggestion and provide average direct and indirect effects.

*Second*, in order to get comparable interpretations for $\beta_1$ and $\beta_2$ we have to take into account, that the lagged knowledge stocks represent a weighted average. This is particular important, since it makes a difference if the weighted average is increased due to a change of a near or distanced region. This effect is a result of the different spatial dependencies represented by the terms $\log\left(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1}\right)$ and $\rho\overline{\mathbf{M}}\log(\mathbf{f})$.

We focus now on the mathematical treatment of these two effects. Stating our regression model once again:

$$\log(\mathbf{f}) = \rho\overline{\mathbf{M}}\log(\mathbf{f}) + \log(\mathbf{k})\beta_1 + \log(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1})\beta_2 + \boldsymbol{\varepsilon}$$

Following and Penn, (2009: p34) we calculate the matrices of the partial derivatives $\mathbf{S}_1$ and $\mathbf{S}_2$ which are generally called effect matrices.

$$\mathbf{S}_1 = (\mathbf{I}_T \otimes \mathbf{A}(\rho)^{-1})\beta_1 \tag{13}$$

$$\mathbf{S}_2 = (\mathbf{I}_T \otimes \mathbf{A}(\rho)^{-1})\mathbf{J}_{\mathbf{Z}}\beta_2 \text{ where } \mathbf{z} = \log(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1}) \tag{14}$$

$\mathbf{S}_1$ and $\mathbf{S}_2$ are $nT$ by $nT$ matrices with the typical element $s_{ij}$ which denotes the partial derivative of $\log(\mathbf{f})$ with respect to $\log(\mathbf{k})$. $\mathbf{J}_{\mathbf{Z}}$ denotes the Jacobian of $\mathbf{z}$ with respect to $\log(\mathbf{k}_{-1})$. The Appendix shows that the typical element of $\mathbf{J}_{\mathbf{Z}}$ can be written as Eq. (15):

$$j_{i,l} = \frac{w_{i,l}}{\sum_{p=1}^{n} w_{i,p}k_p}k_l \tag{15}$$

Once $\mathbf{S}_1$ and $\mathbf{S}_2$ are computed we can establish direct and indirect effects, which are conventionally defined as the mean of the diagonal elements and the mean of off-diagonal elements of $\mathbf{S}_1$ and $\mathbf{S}_2$ respectively (LeSage and Penn, 2008: p37).

$$E_{d,j} = \frac{trace(\mathbf{S}_j)}{nT} \tag{16}$$

$$E_{id,j} = \frac{\boldsymbol{\iota}'_{nT}\mathbf{S}_j\boldsymbol{\iota}_{nT}}{nT} - E_{d,j} \tag{17}$$

Although the calculation of $E_{d,j}$ and $E_{id,j}$ are rather simple, it is neither straightforward how to calculate their standard deviation or their distribution. Therefore, we follow Elhorst (2010) and simulate the distribution of the direct and indirect effects by using the variance-covariance matrix implied by our maximum likelihood estimates:

$$\left(\widetilde{\beta}_1,\ \widetilde{\beta}_2,\ \widetilde{\delta},\ \widetilde{\lambda},\ \widetilde{\rho},\ \widetilde{\sigma}^2_\mu,\ \widetilde{\sigma}^2_\eta\right)' = \overline{\Omega}'\boldsymbol{\phi} + \left(\widehat{\beta}_1,\ \widehat{\beta}_2,\ \widehat{\delta},\ \widehat{\lambda},\ \widehat{\rho},\ \widehat{\sigma}^2_\mu,\ \widehat{\sigma}^2_\eta\right)' \tag{18}$$

where $\overline{\Omega}'$ denotes the upper-triangular Cholesky decomposition of the variance-covariance matrix resulting from our maximum likelihood estimates, $\boldsymbol{\phi}$ is a random vector drawn from $N(\mathbf{0}, \mathbf{I}_7)$, the decoration $\widehat{}$ represents the ML estimators and $\widetilde{}$ the corresponding Monte Carlo values.

## 4 Data, variables and empirical results

### *4.1 Data and Variables*

In the compilation of the data set used by FSR two major sources played a leading role: The Cambridge Econometrics and the European Patent Office database. The Cambridge Econometrics database is used to construct the region-level relative TFP index. The knowledge stocks are proxied by patent data out of European Patent Office patent database. Our overall data set observes 203 European NUTS-2 regions from the 15 pre-2004 European Union members during 1997-2002. Since the regional policy of the European Union is essentially based on the NUTS-2 regions and our empirical results can be of importance for policy makers, NUTS-2 represents for our model a useful geographic scale.

To overcome index-number-problems inherent in TFP- comparisons a TFP-index is preferred over ordinary TFP- levels. So it is ensured that the productivity measure is commensurable across regions and time, which is vital for panel data analysis. In their analysis FSR (2008) used the index introduced by Caves, Christensen, and Diewert (1982) which is transitive and superlative as well:

$$\log(f_{it}) = (\log(q_{it}) - \log(\overline{q_{it}})) - s_{it} \cdot (\log(l_{it}) - \log(\overline{l_{it}})) - (1 - s_{it}) \cdot (\log(c_{it}) - \log(\overline{c_{it}})) \tag{19}$$

In Eq. (19) $q_{it}, l_{it}$ and $k_{it}$ denote output, labour input and capital stock and a bar denotes their corresponding mean. $s_{it}$ denotes the share of labor in the total production costs, which is a more robust specification in case of imperfect competition. For $q_{it}$ data on gross value-added at constant prices of 1995 was used. In order to measure $l_{it,}$ the Cambridge Econometrics Database's labour input data was modified so that it captures the differences in the average annual hours worked and $c_{it}$, was constructed by summing up gross investment using the perpetual inventory method.

In this cost-based setup TFP is invariant to exogenous changes in output, hence it is "uncorrelated with all variables known to be neither causes of productivity shifts nor to be caused by productivity shifts." [Hall (1990)].

The second variable of crucial importance is the knowledge capital stock $k_{it}$. In order to proxy these stocks for every period and region, EPO patents with an application date of 1990-2002, were regionally matched and sorted annually to yield $p_{it}$. Consequently these flows are transformed to stocks using:

$$c_{it} = (1 - r_c)c_{it-1} + p_{it-1} \tag{20}$$

where $r_k$ was set to be 12%,.In case of international patents a fraction was determined, instead of a full-counting. FSR (2008) note that using patent data also has some drawbacks: Patenting is inherently a strategic decision, it causes costs (money, energy) and therefore it is beyond any reasonable doubt that inventions are merely partly patented (codified). Another problem with this counting approach is the fact that patents are differently useful: there are large innovations and patents with virtually no practical use at all. Nevertheless, to the extended that patents document inventions, an aggregation of them is very closely related to knowledge capital stocks.

### *4.2 Empirical Results*

In this section we proceed with presenting the estimates from the autoregressive model and comparing them with the results reported by FSR . For this purpose the ordinary nonlinear GLS model, which neglects spatial correlation in the error term, the standard FSR model, and our autoregressive FSR model are shown side by side in table 1.

[INSERT TABLE 1 here]

The GLS estimates indicate that $\beta_1$, the elasticity with respect to region-internal knowledge, is positive and significant. The same is true for $\beta_2$ which advocates, in consistence with the spillover literature, that region-external knowledge also exercises a positive influence on a regional TFP. The positive estimate of $\delta$ suggests that this influence is declining exponentially with distance at a rate of 8%. In comparison with the FSR- model there are some notable differences: The GLS model seems to overstate the importance of knowledge, both the estimates of $\beta_1$ and $\beta_2$ are smaller in column two. The opposite is true for the decay parameter $\delta$ which is about one percent point higher in the FSR setting. These findings are emphasized by the positive and highly significant estimate of $\lambda$, which indicates that spatial autocorrelation due to neighboring regions is indeed an issue which has to be addressed.

In column 3 we finally present the estimates of the extended FSR, which incorporate the spatial autoregressive lag. As can be seen readily, signs and significance levels of the parameter estimates are similar to the FSR- model however their magnitude differs notably: $\beta_1$ ($\beta_2$) is about 10% (20%) smaller than in the FSR- model. On the contrary the estimate of $\delta$ increases by about 10%, and is close to the estimate achieved with GLS. The parameter on the AR-term $\rho$ takes on a value of 0.536 and is highly significant, whereas $\lambda$, which describes the spatial correlation in the error term, is reduced drastically. This reinforces the concerns expressed in section 3.1: Although the estimates of $\sigma_n$ and $\sigma_\mu$ do not differ significantly in comparison with FSR, $\lambda$ and $\rho$ do. Consequently, FSR's $\Omega$ is different which causes a bias in all the other estimates. Viewed from a different angle one could also say that the lagged TFP values in this setting, have marginal explanatory power, because they include additional factors which contribute to knowledge and therefore influence productivity positively. Once this issue is accounted for, patent stocks loose somewhat of their importance, as can be seen from the smaller estimates of the coefficients $\beta_2$ and $\beta_1$.The results put forward our presumption that the standard FSR model suffers from omitted variable bias, which is at least partly corrected by the introduction of the autoregressive term. This is in line with the recent work of Fingleton (2010) who states that "... the presence of the endogenous lag should help mitigate omitted variable bias in spatial regressions"

As already stressed out in section 3.1, in the presence of an autoregressive term $\beta_2$ and $\beta_1$ are no longer equivalent to elasticities. To make our results comparable to those of FSR we therefore have to calculate the implied mean total effects with respect to internal (MTI) and spatially discounted external knowledge (MTE) capital stocks separately and simulate their distribution[4]. The results are contrasted in Table 2.

[4] If we use the definitions given in section 3 and derive the mean total effects with respect to $\log(\mathbf{k})$ and $\log(\overline{\mathbf{W}(\delta)}\mathbf{k}_{-1})$ we end up with $\boldsymbol{\iota}'_{nT}(\mathbf{I}_T \otimes \mathbf{A}(\rho)^{-1})\beta_1\boldsymbol{\iota}_{nT}/(nT)$ and $\boldsymbol{\iota}'_{nT}(\mathbf{I}_T \otimes \mathbf{A}(\rho)^{-1})\beta_2\boldsymbol{\iota}_{nT}/(nT)$ respectivly.

| | extended FSR | | | FSR | |
|---|---|---|---|---|---|
| | mean | std | | mean | std |
| mean total effect of intra-regional knowledge | 0.333 | 0.056 | $\beta_1$ | 0.209 | 0.026 |
| mean total effect of extra-regional knowledge | 0.171 | 0.039 | $\beta_2$ | 0.126 | 0.020 |

Table 2: Total Effect Estimates

Obviously the elasticities in the extended model are higher than the corresponding estimates presented by FSR. Particularly the MTE and MTI exceed their FSR counterparts ($\beta_1$ and $\beta_2$) by 60% and 35% respectively. To complete our analysis we continue by stating the estimates of the average direct (DE) and indirect effects (IDE) implied by the autoregressive Model:

| | mean | std | p-value |
|---|---|---|---|
| mean direct effect | 0.206 | 0.0248 | 0.0001 |
| mean indirect effect | 0.304 | 0.0554 | 0.0001 |

Table 3: Direct and indirect Effect Estimates

As Table 3 shows, both the DE and IDE seem to be positive and highly significant. This result suggests that, accounted for spatial autocorrelation and given our estimates of the spatial decay parameter $\delta$, on average augmenting knowledge stocks, exercises a positive effect on TFP directly and indirectly: Not only that a region benefits from rising its stock of knowledge in the first place, it also benefits indirectly through feedback from neighboring regions to which knowledge was spilled over beforehand. This second-order effect is about 50% higher than the first-order-effect. One remarkable consequence is that it seems more desirable now to raise TFPs by raising extra-regional than intra-regional knowledge capital.

## 5 Conclusion

In this paper we augmented the FSR model with a first order spatial lag in order to address OMB-issues. As the results show, both the region internal and the region external knowledge capital stock have a positive influence on TFP. The importance of region external knowledge suggests that spillovers play a significant role: regions, that have a common border with high-knowledge regions, have significantly higher TFPs themselves. These spillovers, however, are localized, they decay exponentially with rate of about 8%.

Our estimate of the spatial lag parameter was positive and highly significant. This suggests that the extra-regional TFP may act as a proxy for multiple influences which are relevant for a regions knowledge, but not covered in the number of patents. Consistent with this view the coefficient of the intra (extra) regional patent-stocks was 10% (20%) smaller then reported in the FSR model.

In a third step we tried to assess the direct and indirect effects implied by our estimates. Our simulations showed that both direct and indirect effects are positive and significant. Notably the indirect effect was about 50% higher than the direct effect. This piece of evidence stresses out, once more, that cross-region knowledge spillovers reinforce the impact of knowledge production activities.

# Acknowledgment

The study reported in this paper was integral part of the Seminar Course "Spatial Economics" [Instructors: Professor Dr. Manfred M. Fischer and Mag. Matthias Koch] in the Economics Masters Program of the Vienna University of Economics and Business. We are very grateful for the valuable comments and discussions of Professor Fischer, and the data and excellent research environment provided by the Institute for Economic Geography and GIScience.

**References**

[1] Caballero, J. R., and J. B. Jaffe. (1993). "How High are the Giants Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth." In NBER Macroeconomics Annual 1993, Vol. 8, 15–74, edited by U. Blanchard and S. Fischer. Cambridge, MA: MIT Press.

[2] Caves, D., L. Christensen, and W. Diewert. (1982). "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers." Economic Journal 92, 73–86.

[3] Doraszerlski U., and J. Jaumandreu. (2008). "R&D and Productivity: Estimating Production Functions when Productivity is Endogenous." Discussion Paper No. 2147, Harvard University, Cambridge, MA.

[4] Elhorst JP (2010) "Matlab Software for Spatial Panels." Presented at the IVth World Conference of the Spatial Econometrics Association (SEA), Chicago, June 9-12, 2010

[5] Fingleton, B. (2001). “Equilibrium and Economic Growth: Spatial Econometric Models and Simulations.” Journal of Regional Science 41(1), 117–47.

[6] Fingleton B, Le Gallo J. (2008). "Estimating Spatial Models with Endogenous Variables, a Spatial Lag and Spatially Dependent Disturbances: Finite Sample Properties," Papers in Regional Science, Blackwell Publishing, vol. 87(3), pages 319-339, 08

[7] Fischer, M. M., T. Scherngell, and E. Jansenberger. (2006). “The Geography of Knowledge Spillovers Between High-Technology Firms in Europe. Evidence from a Spatial Interaction Modelling Perspective.” Geographical Analysis 38(3), 288–309.

[8] Fischer, Manfred M., Scherngell, Thomas, Reismann, Martin. (2009). "Knowledge spillovers and total factor productivity. Evidence using a spatial panel data model." Geographical Analysis. An International Journal of Theoretical Geography 41 (2): 204-220.

[9] Griffith, D. A. (1988). Advanced Spatial Statistics. Dordrecht, The Netherlands: Kluwer.

[10] Griliches, Z. (1979). “Issues in Assessing the Contribution of Research and Development to Productivity Growth.” The Bell Journal of Economics 10(1), 92–116.

[11] Keller, W. (2002). “Geographic Localization of International Technology Diffusion.” American Economic Review 92(1), 120–42.

[12] LeSage, J.P., and R.K. Pace. (2009). "Introduction to Spatial Econometrics." Boca Raton, US: CRC Press Taylor & Francis Group.

[13] Solow, R. M. (1957). “Technical Progress and the Aggregate Production Function.” Review of Economics and Statistics 39, 312–20.

[14] Wansbeek, T. J., and A. Kapteyn. (1983). “A Note on Spatial Decomposition and Maximum Likelihood Estimation of ANOVA Models with Balanced Data.” Statistics and Probability Letters 1, 213–5.

# Appendix

## The analytical Jacobian

Given our set $\mathbf{z}$ of $n$ equations in $n$ variables $k_1, ...k_n$

$$\mathbf{z} \equiv \ln(W^{n\times n}\mathbf{k})$$

or more explicitly

$$\begin{cases} \ln(\sum\limits_{i=1}^{n} w_{1i} * k_i) \\ \ln(\sum\limits_{i=1}^{n} w_{2i} * k_i) \\ \qquad ... \\ \ln(\sum\limits_{i=1}^{n} w_{ni} * k_i) \end{cases}$$

the ijth Element of the Jacobian Matrix with respect to $ln(\mathbf{k})$ is defined as:

$$J_{ij} \equiv \frac{dz_i}{d\ln(k_j)} = \frac{\frac{dz_j}{d(k_j)}}{\frac{d\ln(k_j)}{d(k_j)}} = \frac{w_{ij}}{\sum\limits_{s=1}^{n} w_{js} * k_s} k_j.$$

Therefore:

$$\mathbf{z}_{\ln(\mathbf{k})} \equiv J = \begin{bmatrix} \frac{w_{11}}{\sum\limits_{i=1}^{n} w_{1i}*k_i} k_1 & \frac{w_{12}}{\sum\limits_{i=1}^{n} w_{1i}*k_i} k_2 & ... & \frac{w_{1n}}{\sum\limits_{i=1}^{n} w_{1i}*k_i} k_n \\ \frac{w_{21}}{\sum\limits_{i=1}^{n} w_{2i}*k_i} k_1 & \frac{w_{22}}{\sum\limits_{i=1}^{n} w_{2i}*k_i} k_2 & ... & \frac{w_{2n}}{\sum\limits_{i=1}^{n} w_{2i}*k_i} k_n \\ ... & ... & & ... \\ \frac{w_{n1}}{\sum\limits_{i=1}^{n} w_{ni}*k_i} k_1 & \frac{w_{n2}}{\sum\limits_{i=1}^{n} w_{ni}*k_i} k_2 & ... & \frac{w_{nn}}{\sum\limits_{i=1}^{n} w_{ni}*k_i} k_n \end{bmatrix}$$

| | GLS | | | FSR | | | extended FSR | | |
|---|---|---|---|---|---|---|---|---|---|
| | mean | std | p-value | mean | std | p-value | mean | std | p-value |
| region-internal stock of knowledge ($\beta_1$) | 0.200 | 0.026 | 0.000 | 0.209 | 0.026 | 0.000 | 0.187 | 0.022 | 0.000 |
| region-external stock of knowledge ($\beta_2$) | 0.120 | 0.021 | 0.000 | 0.126 | 0.020 | 0.000 | 0.096 | 0.018 | 0.000 |
| Distance sensitivity parameter ($\delta$) | 0.080 | 0.036 | 0.027 | 0.072 | 0.027 | 0.001 | 0.084 | 0.021 | 0.000 |
| errorterm autocorrelation parameter ($\lambda$) | | | | 0.640 | 0.040 | 0.000 | 0.180 | 0.058 | 0.001 |
| endogenous spatial lag parameter ($\rho$) | | | | | | | 0.536 | 0.039 | 0.000 |
| Variance $\left(\sigma^2_\eta\right)$ | 0.004 | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 | 0.004 | 0.000 | 0.000 |
| Variance $\left(\sigma^2_\mu\right)$ | 0.173 | 0.024 | 0.000 | 0.155 | 0.009 | 0.000 | 0.128 | 0.013 | 0.000 |
| Log likelihood | 1016.363 | | | 1146.29 | | | 1136.08 | | |
| AIC | −2026.726 | | | −2232.58 | | | −2258.16 | | |

Table 1. Comparison of the nonlinear GLS, the FSR and the extended FSR Model