Spatial knowledge spillovers: a regional view of Germany

Torben Klarl
Department of Economics, University of Augsburg,
Universitätsstraße 16, D-86159 Augsburg, Germany
e-mail: torben.alexander.klarl@wiwi.uni-augsburg.de
Tel.: ++49(0)821-598-4177
Fax: ++49(0)821-598-4229

April 8, 2009

Preliminary Version

Abstract

The aim of this paper is to give an answer to the question, whether local spatial knowledge spillovers can be found in a cross section analysis for German NUTS-2 data. The analysis is based on regional production functions embedded in a general spatial model context. In addition, the paper expands the analysis to a Bayesian econometric view to allow for the existence of spatial heterogeneity in the data. Further, both using Bayesian and Non Bayesian methods, it should be more likely to obtain a more reliable model selection mechanism. Finally, using spatial filtering methods, own and neigbhouring effects of regions regarding their innovative potential are separated. Particularly, the paper find evidence, first, that the output per capita of a region follows a spatial process, driven mainly by patents and human capital and second, that, consulting the knowledge production function theory, spatial knowledge spillovers clusters mainly exists with some exceptions in West German regions. Third, employing eigenvector based filter methods it was found that West German high productive regions are productive mainly due to their own innovative potential and mainly East German regions are confronted with negative neighbouring effects.[1]

Keywords: Spatial econometrics, Spatial filtering, Bayesian spatial econometrics
JEL Classification Number: C11, C31, C49

[1] Preceding versions of the paper have been accepted at the following conferences: Knowledge for Growth Conference Toulouse, 07.-09. July 2008, 9th. Conference on Patents & Innovations, Applied Econometrics Association (AEA), Tokyo, 19.-20. December 2008, 6th. European Meeting on Applied Evolutionary Economics (EMAEE), Jena, 21.-23. May 2009, III World Conference of Spatial Econometrics (SEC), Barcelona, 08.-10. July 2009.

1 Introduction

It is an undisputable fact that knowledge and technological change are the driving forces for long run economic growth. Additionally, endogenous growth theory tells us that knowledge spillovers are necessary for long term growth of high-income regions. Several contributions regarding this topic have been published during the last years. (Lucas, 1988), (Krugman, 1991) and (Romer, 1986), for instance have explicitly focused on the accumulation of new knowledge in context of new growth theory. Their key finding is, that endogenous accumulation of knowledge is the surety of per capita income growth. These approaches have in common that they focus on convexities in production process[2]. For instance, convexities in production can arise from positive externalities caused by learning-by-doing, human capital accumulation and the supply of governmental goods.

As argued by (Keilbach, 2000), knowledge spillovers can be treated as a special type of positive externalities and, moreover, is one motivation for positive returns to scale in an aggregate production function approach which was first used by (Griliches, 1979).

At the latest as European leaders met in Lisbon 2003 and defined the goal of becoming "the most dynamic and competitive knowledge-based economy in the world" by 2010 the term it can be said without any limitations that the knowledge-based economy has gained much attraction, not only in research but also in politics. Today, the creation and diffusion process of knowledge is the focal point of research, because "knowledge is the most important strategic resource and learning the most important process"[3]. But what is knowledge? Well, the term knowledge is often used in scientific publications, but it is sometimes confounded with the term "information". It must be clear that knowledge comprises the individual specific abilities which can be used to solve more or less strategic problems underpined with a pool of information. As pointed out by (Krugman, 1991) "[k]nowledge flows are invisible; they leave no paper trail by which they may be measured and tracked[...]". Information

[2]Refer on (Krugman, 1991) for this topic for instance.

[3](Morgan, 1997), p. 493.

instead, is more or less visible. It can interpreted as the collection of knowledge, for instance the collection of data. Hence, when talking about knowledge, we often don't know what we know. Thus, knowledge cannot be measured directly, as other production inputs such as the stock of capital, for instance. The consequence is, that we have to find proxies for this knowledge, for instance human capital or data of patent citations. But doing so, we have to macerate the strict distinction between information and knowledge. That should be kept in mind when talking about the outstanding role of knowledge for economic growth.

It is rather intuitive, that spatial barriers of knowledge diffusion can be used as an argument for income and production differentials between regions. That should be considered as one reason why we observe cluster and agglomeration in economic long run growth. Regions (take cities for example) which are more productive and supply a higher life quality are more attractive for innovative companies. Consequently, these regions become more attractive again and this process leads to a more and more decreasing productive differential. It is not a surprising fact, that economic growth and agglomeration are positive correlated (Baldwin and Martin, 2003). Hence, growth differentials are enforced by knowledge capital concentration. As mentioned by (Fujita and Thisse, 2002), knowledge spillovers can be interpreted as a source for sustainable regional growth, given decreasing returns of learning are excluded.

If we argue that spatial patterns are worth investigating, it is necessary to ask the question how knowledge spillovers affect agglomeration. To answer this question we could argue that cities or densely populated regions may have positive effects on their productivity due to so called Marshallian externalities. (Marshall, 1920) mentioned, that so called externalities are necessary for economic agglomeration and therefore create a so called look-in effect[4]: "When an industry has thus chosen a location for itself, it is likely to stay there long: so great are the advantages which people

[4](Fujita and Thisse, 2002), p. 7.

following the same skilled trade get from near neighbourhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. Good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas."[5] Of course, the justification of agglomeration by Marshall is primarily based on trade arguments but can easily be expanded to other factors, which influence the decision of where to situate a location, as mentioned above. (Kahnert, 1998) found that knowledge intensive processes are agglomerated in dense regions, while less knowledge intensive processes are situated in more peripheral regions. Thus, knowledge spillovers cause externalities and force agglomeration and as a consequence, as pointed out by (Scotchmer and Thisse, 1992) leads to uneven geographical distribution of economic activity.

Hence, from a theoretically driven view, increasing returns to scale, agglomeration and distribution of economic numbers, for instance per capita productivity are closely linked with space. Although, the link of technological innovations and knowledge diffusion for technological growth is acknowledged in growth literature[6], the role of knowledge diffusion is only partly considered. Some of the North-South trade literature on diffusion and technological progress[7] consider feedback effects between the North and the South in the steady state, but an analysis of the transitional dynamics for either region is missed. (Barro and Sala-I-Martin, 1997) indeed derived transitional dynamics for the South but feedback effects are excluded due to the effect of no trade of intermediate goods. Thus, a transition path for the North cannot be derived. The communality of this strand of literature is only focused on two country or two region models, which consists of a rich North and a poor South

[5](Marshall, 1920), p. 225.

[6]Refer to (Romer, 1986), (Romer, 1990) and (Krugman, 1991) for instance.

[7]Refer to (Krugman, 1979), (Dollar, 1986), (Grossman and Helpman, 1991b), (Grossman and Helpman, 1991a), (Rivera-Batiz and Romer, 1991), (Barro and Sala-I-Martin, 1997) and (Glass, 1997).

or a core and a peripheral country. From this perspective, those types of models are less suitable to investigate the link of increasing returns to scale, agglomeration and distribution of economic numbers because of the simple reason: in a two country framework, it is not reasonable to investigate agglomeration effects when referring to regions. One of the factors, why multiple country or regional focused growth models are less attractive or gained less attention could be the fact that such growth models become very complex and cannot solved analytically and only numerically solutions remain.

For this reason, the relevant literature which investigates the link between increasing returns to scale, agglomeration and distribution of economic numbers is heavily empirical orientated and is sometimes more or less ad hoc. To investigate spatial agglomeration effects empirically, one has to refer to tools from a toolbox which can be summarized with "spatial econometrics", a term widely used in New Economic Geography (NEG)[8]. (Anselin, 1988)'s book can be described as the first comprehensive introduction to spatial econometrics. In contrast to spatial statisticians, where pure data or data based approaches are in the front, the spatial econometricians deal with model-funded approaches, based upon a theoretical model. However, the commonality of the two perspectives is the acceptance of the existence of spatial stochastic processes.

Although, from an empirical view, there has been made much progress in explaining the link between increasing returns to scale, agglomeration and distribution of economic numbers. But there are still limitations especially when talking about the grasp of knowledge spillovers and knowledge diffusion.

First, less attention is concentrated on the fact, that knowledge diffusion is not a constant process over space. Often it is assumed that only the nearest neighbour has a significant influence on economic growth, whereas farther away neighbours do not exert any economic influence, or more technically spoken, often it is assumed that knowledge diffusion follows a spatial AR(1) or spatial MA(1) process and second or higher order effects or a combination of both are neglected. This assumption

[8]For an overview of NEG refer to (Krugman, 1998a) and (Krugman, 1998b) for example.

seems to be to strict. Instead of ignoring higher order effects of spatial influence, one should insert them into a model framework, because neglecting them could lead to an underestimating of spatial influence. Further, this second or higher order processes should not be treated as a constant extrapolation, but rather as non constant function over space. Hence, it is reasonable to assume that more contiguous neighbours have a direct and stronger influence than less contiguous neighbours.

In most of the existing empirical studies the grasp of knowledge spillovers has only gained limited attention. (Anselin et al., 1997) and (Anselin et al., 1997) are two of the few studies how mentioned concrete numbers of knowledge spillover scope. (Anselin et al., 1997) found by investigating the influence of university related research and private research and development (R&D) effort on of knowledge transfer that a significant positive effect can be detected within a 50 mile radius of metropolitan statistical areas (MSAs) only for the university research. For private R&D such an significant effect could not be detected. (Anselin et al., 1997), with a similar setup as (Anselin et al., 1997) additionally have shown, that not only spillovers within MSA but also between MSA can be found. The key cognition of the latter mentioned study is, that without exact geographical distance measures, it can be shown that spatial influence is bounded locally. (Audretsch and Mahmood, 1994) have shown on patent basis for 59 US metropolises, that knowledge spillovers are limited towards the metropolises' boarders. They come to this conclusion because they found that only for research institutes which are settled within a metropolis, significant knowledge spillovers can be detected, whereas for research institutes, settled in each metropolis related country, no such effects could be found.

Second, within the specification of spatial models, spatial heterogeneity is mostly missed. It is sometimes ignored, that spatial effects can appear as two types: the one type is spatial dependence, the other is spatial heterogeneity. Spatial dependence, which is consistently assumed in the above mentioned studies, is mainly caused by problems of measuring that are caused by spatial spillovers and spatial externalities. In contrast to spatial dependence, spatial heterogeneity means that spatial effects are not uniformly distributed across space and outliers could exist. From a standard

econometricians toolbox, this could be seen as a spatial kind of heteroscedasticity. Although several arguments militate in favour that spatial heterogeneity matters[9], this aspect is not "seen as a serious problem in spatial regression"[10]. One reason could be, that spatial econometrics, if we refer to theoretical econometrics, is still a developing discipline.

But what should be done, if spatial dependence, spatial heterogeneity or a combination of both types is relevant and further a set of possible AR(p), MA(q) or ARMA(p, q) processes with order $p$and q respectively, are suitable in model context? Given, our model is correctly specified, than standard econometrics tells us, that parameter estimates are insufficient if spatial heterogeneity is ignored, although it is relevant. But given, the model is based on a wrong choice of AR(p), MA(q) or ARMA(p, q) terms, then our model is wrong specified. Of course, the latter problem is the more serious one.

Although, model selection should be taken seriously, we frequently find that empirical based studies using tools from spatial econometrics, based on ex ante conceptions of a spatial model. This means, a model selection is often defaulted or, if done, it is based mainly on a limited class of spatial processes, which commonly include the decision of relying on a spatial AR(1) or spatial MA(1) process based on the assumption of spatial homogeneity. There are, to best of my knowledge only a few papers which cover the aspect of spatial model choice.[11]

Thus, traditional or frequentest econometrics approach suffers from two reason in context of spatial econometrics: first, the models and the underlying estimation methods assume spatial homogeneity, and second, model selection is rather heuristic. For that reasons, Bayesian methods have been prevailed and proved in spatial econometric application. The key difference between frequentest and Bayesian

[9](Anselin, 1988) for instance comment on page 13 with respect to importance of spatial heterogeneity in econometricians work, that "several factors, such as central place hierarchies, the existence of leading and lagging regions, vintage effects in urban growth [...] would argue for modeling strategies that take into account the particular features of each location (or spatial unit)."

[10](Keilbach, 2000), p. 122.

[11]For instance refer to (Hendry, 1979), (Florax et al., 2003) and (Hendry, 2006) for an intensive discussion regarding model selection methods.

methods are that the latter treat the coefficient vector of estimators itself as random, whereas frequentest say that the resulting estimates of the coefficient vector is random. Bayesian methods hold a great deal for several reasons: for instance, first, it is possible to model hierarchy of place or regions, second, one can integrate a more or less systematic change of variance over space, and thus spatial heterogeneity and third it is possible to acknowledge a hierarchy of regions or places. Bayesian methods can incorporate these ideas because of their underlying concept as prior information complements existent sample data information, whereas frequentest methods can solely rely on latter mentioned. As mentioned before, although Bayesian methods seem to be very attractive, their usage in application is very limited. On the other side, frequentest methods are, if they only limited to the spatial dependence case, and therefore assume spatial homogeneity, lead to insufficient parameter estimates. Anyway, a more or less large research agenda for both, spatial econometrics and spatial statistics remains.

From the discussion above, we see that two different arguments regarding productivity growth are discussed in the relevant literature: on the one hand, the (theoretically) role of technological innovations and knowledge diffusion for technological growth[12], and on the other hand the (empirical) role of spatial agglomeration on long run productivity growth[13]. The point is, that the first mentioned strand does discuss growth implications of knowledge diffusion in a less suitable frame when focusing on distribution questions and agglomeration, while the latter strand suffers more or less from theoretical fortification.

Hence, these two approaches are more or less discussed in isolation rather to be combined and to investigate the relationship between knowledge diffusion, agglomeration and growth. This topic has gained less attention in relevant literature, although (Fujita and Thisse, 2002) mentioned that "increasing returns to scale (IRS) are essential for explaining geographical distributions of economic activities"[14].

[12]Refer to (Romer, 1986), (Romer, 1990) and (Krugman, 1991) for instance.

[13]Refer to (Keilbach, 2000), (Bottazzi and Peri, 2003), (Greif, 1998) and (Frauenhofer, 2000) for instance.

[14](Fujita and Thisse, 2002), p. 342.

There is to best of my knowledge only one study, which tries to bridge the two approaches: (Keilbach, 2000) has investigated the role of knowledge for German "Kreise"[15] both empirically and theoretically within a (Romer, 1986) context. He found, that increasing returns to scale lead to significant cluster effects. Further, he found on basis of several production functions estimations, that spatial dependence has a significant influence on labour productivity. But it has to be mentioned, that (Keilbach, 2000) assumes explicitly spatial homogeneity and only first order spatial effects, both in his theoretical and empirical studies. Further, using "Kreise" as regions could lead to spatial dependence per definition, due to the fact that "Kreise" are the smallest entity of regions for the case of Germany, and thus stream of commuters can lead to biased estimations of spatial effects by construction.

Thus, the main intention of this paper is to identify knowledge spillovers in a spatial context. Particularly, it is assumed, that spatial effects per se are heterogeneous, an assumption which seems to be plausible. Hence, this paper combines spatial heterogeneity and spatial dependence which are the two main aspects of spatial econometrics.[16] Especially, in the most existing studies dealing with spatial knowledge spillovers, the aspect of spatial heterogeneity has been either totally neglected or it has been assumed that spatial effects are uniform across regions. In this paper both aspect can be ideally integrated and discussed in a Bayesian framework.

The empirical study per se is based on a spatial cross section production function approach, proposed by (Griliches, 1979) which should measure the effects of innovativeness, measured by knowledge capital, such as human capital, patents or $R\&D$ and spatial spillovers on output for German NUTS-2 regions. NUTS-2 regions are used to exclude spatial dependence by construction.

Further, a new model choice mechanism is introduced which on the one hand is based on traditional econometric tools and on the other hand integrates Bayesian model choice criteria. This mechanism also controls for spatial heterogeneity, as mentioned right before. Finally, under the condition that spatial processes can be detected in the data, a filter method is applied to remove spatial influence and thus

[15] "Kreise" is a German administration unit which is equivalent to NUTS-3 level.

[16] Refer to (Anselin, 1988), p. 11.

to identify own and neighbour productivity effects of regions and to discuss political implications against the background of obtained results.

2 Motivation

The basic cross section regression model stems from a simple production function approach and can be written as follows:

$$y = X\beta + \epsilon, \tag{1}$$

where y is a non stochastic $N \times 1$ vector of observation, X is a full rank $N \times K$ matrix of K non stochastic independent variables, β is a $K \times 1$ vector of regression coefficients and ϵ is treated as a normally and independently distributed $N \times 1$ vector of errors. The drawback of a formulation like equation 1 is, that it does not acknowledge spatial dependence. But if spatial dependence, especially spatial autocorrelation, exist in the data, and if they are neglected within the estimation setup 1, an estimation based on OLS may not be consistent[17]. This argumentation is familiar when talking about estimation problems within a pure time series approach. Therefore, equation 1 has to be altered and expanded for spatial processes. Generally, spatial events appear in three forms: first, spatial dependence is only observed in the y vector. As a consequence of that, a spatial lag model or a spatial AR(1) model has to be estimated. Second, spatial dependence is only observed in the error term vector ϵ, which means that one has to model a spatial error or a spatial MA(1) model. Or third, a combination of both spatial events occur in the data. Then a mixture of a spatial lag and a spatial error model has to be used. Given the latter is true, then we can rewrite equation 1 as a spatial ARMA(1,1) model as follows:

$$y = \rho W y + X\beta^X + \tilde{X}\beta^{\tilde{X}} + \lambda W\epsilon + \kappa, \tag{2}$$

with $X = [x_1, x_2, ..., x_K]$, $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_M]$ and the $K \times 1$ vector $\beta^X = [\beta_1^X, \beta_2^X, ..., \beta_K^X]$, and the $K \times 1$ vector $\beta^{\tilde{X}} = [\beta_1, \beta_2, ..., \beta_M]$.

[17] Refer to (Anselin, 1988) and (Anselin and Rey, 1991).

The parameter ρ is the so called spatial autoregression coefficient, W is a $N \times N$ matrix containing spatial weights, and κ is a $N \times 1$ vector containing errors. Often it is assumed that $M = K$. Thus, a close relationship between time series and spatial econometrics modeling can be observed. But it is worth to note, that the analogy regarding the labeling of such a process to time series is misleading sometimes because spatial spillovers are often described by feedback-processes, as mentioned before.

The $N \times K$ matrix X contains non spatial exogenous variables, whereas the $N \times M$ matrix $\tilde{X}$ contains the spatial lagged exogenous variables. Of course we can write $\tilde{X} = WX$. Stacking Wy, X, $\tilde{X}$ and $W\epsilon$ in $\tilde{X}^+ = [Wy, X, \tilde{X}, W\epsilon]$ and $\tilde{\beta} = [\rho, \beta^X, \beta^{\tilde{X}}, \lambda]'$ leads to

$$y = \tilde{X}^+\tilde{\beta} + \kappa. \tag{3}$$

Although it is common to assume that $\kappa \sim N(0, \sigma^2 I)$, it is more plausible to assume that $\kappa \sim N(0, \sigma^2\Omega)$ with $\sigma_i = \tilde{h}(f_i'\alpha)$ and $h(\cdot) > 0$ as unknown, continuous function which are treated as the diagonal elements of the error covariance matrix $\sigma^2\Omega$.

Although (Keilbach, 2000) and (Klotz, 1996) argue that spatial heterogeneity is not seen as a serious problem in spatial econometrics context it should be in fact treated as a serious problem ex ante. Remember for instance that some regions do not follow the same spatial relationships as other regions. This "enclave effects" or in an econometric notation, these "outliers" could cause severe problems such as fat-tailed errors which are not normal of course. A t-distribution is more appropriate then. In such cases it seems more appropriate to acknowledge these outliers and use Bayesian methods for instance.

Only for the fact that $\tilde{h} = \sigma^2$ it follows that $\kappa \sim N(0, \sigma^2 I)$ which implies spatial homogeneity. The big problem estimating a heterogeneous spatial model is that allowing for heteroscedasticity we have to estimate N additional parameter for each σ_i. Of course, this leads to the so called "degree of freedom" problem, because we do not have simply spoken enough observations to compute an estimate for every point located in space. Therefore, we are confronted with a problem using the "traditional" econometricians toolbox. One way to deal with this problem is

to refer to Bayesian econometrics. Bayesian methods in regression context do not encounter the similar degree of freedom problems, because informative priors are available. As seen later, the prior distribution for our N diagonal elements of Ω are independently $\frac{\chi^2(s)}{s}$ distributed. Note, that the χ^2-distribution is a single parameter distribution where we can represent this parameter as s. This allows us to estimate N additional parameter of the diagonal elements of Ω by adding a single parameter r to our regression procedure.

Hence, the estimation strategy is defined as follows: one should start with an estimation of a spatial ARMA-model with homogeneous errors based on equation 2. Of course, expression 2 can be considered also as a spatial ARIMA-model, if $|\rho| = 1$. If we do observe a significant coefficient of ρ close to one[18], one should estimate a spatial ARIMA-model to avoid results based on spurious regressions. Equation 2 can be consistently estimated via Maximum-Likelihood (ML) as mentioned by (Anselin and Rey, 1991). Please note again, that (ML) based models are not suitable to model spatial heterogeneity. For this reason, (ML) estimations implicitly assume spatial homogeneity. For this reason, Bayesian models with the additional assumption of heterogeneous errors should be estimated. After performing model selection mechanism, a direct model comparison of the (ML) based and the Bayesian model should be used, to find the model which best fits to the data generating process. If one detects dissimilarities between the two approaches, then one of course should rely on the Bayesian model than on the (ML) approach.

3 Spatial weight

Until today, there is no theory about how to find the "correct" spatial weight matrix W. Therefore, the choice of the spatial weights should be done on the basis of the specific research topic. The first question one has to ask is how to proxy spatial proximity. One approach is to say, that spatial proximity is best proxied by geographical distances. Another way is to say, that geographical boarders are less

[18]Naturally, ex ante it is difficult to decide, whether one is confronted with a highly persistent or an unit root process with respect to space.

important for spatial proximity and for this reason one should better rely on non geographical data, such as trade shares[19] or data on FDI[20].

The latter strategy has two major drawbacks in this context: First, in this work, it is primarily focused on knowledge diffusion. When talking about this issue it is rather not intuitive to proxy spatial proximity by trade shares or FDI data for instance. Second, there is a methodological problem: using these weights it is very likely, that they can be endogenous and therefore lead to biased estimators if not using an IV or GMM approach.

Hence, the majority of the literature is refering to more geographical weights. It is common using geographical distances (Keller, 2001) or more precisely using great circle distances between regions' centroids (Anselin, 1988). But this has the inherent assumption that knowledge spillover sources are located in region´s centroids. Another way, which is also consulted in this study, is simply to refer to binary weighting schemes [21]. If a region i is a neighbour of another region j, then the i-th element of W, w_{ij} takes a 1, otherwise a 0.

Thus, we can write for the symmetric $N \times N$ matrix W with weights w_{ij}:

$$w_{ij} = \begin{cases} 1, & \text{if i and j have a common border and i}\neq\text{j} \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

Often, this matrix is weighted or standardized because this facilitates the interpretation of the estimated coefficients[22] and guarantees that the Moran's $\mathcal{I}$ is situated in the interval $[-1; 1]$[23]. Using the weighting scheme, proposed by (Anselin, 1988), we write for the standardized elements w_{ij}^{+} of W^{+}:

$$w_{ij}^{+} = \frac{w_{ij}}{\sum_{j=1}^{N_j} w_{ij}}. \tag{5}$$

In this way we have created a row standardized spatial weighting matrix W^{+} which is used in the preceding estimation exercise.

[19] Refer to (Coe and Helpman, 1995).

[20] Refer to (Lichtenberg and van Pottelsberghe de la Potterie, 1996).

[21] Refer to (Tappeiner et al., 2008).

[22] (Anselin, 1988), p. 23.

[23] Refer to (Ord, 1975) and (Griffith, 1996).

4 Higher order spatial influence specification

One major drawback of model 2 is, that higher order spatial dependencies are not included. To obtain a higher order weighting matrix W^{+r} for $r = \{1, ..., R\}$ we should increase the power of the simple contiguity matrix [24]. Labelling the order of the spatial dependency with $r = \{1, 2, 3, ..., R\}$ then $\tilde{X}$ can be expanded as follows:

$$\tilde{X}^{++} = \begin{pmatrix} \begin{pmatrix} \tilde{x}^1_{11} & \tilde{x}^1_{12} & \cdots & \tilde{x}^1_{1M} \\ \tilde{x}^1_{21} & \tilde{x}^1_{22} & \cdots & \tilde{x}^1_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}^1_{N1} & \tilde{x}^1_{N2} & \cdots & \tilde{x}^1_{NM} \end{pmatrix} & 0 & \cdots & 0 \\ 0 & \begin{pmatrix} \tilde{x}^2_{11} & \tilde{x}^2_{12} & \cdots & \tilde{x}^2_{1M} \\ \tilde{x}^2_{21} & \tilde{x}^2_{22} & \cdots & \tilde{x}^2_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}^2_{N1} & \tilde{x}^2_{N2} & \cdots & \tilde{x}^2_{NM} \end{pmatrix} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \begin{pmatrix} \tilde{x}^R_{11} & \tilde{x}^R_{12} & \cdots & \tilde{x}^R_{1M} \\ \tilde{x}^R_{21} & \tilde{x}^R_{22} & \cdots & \tilde{x}^R_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}^R_{N1} & \tilde{x}^R_{N2} & \cdots & \tilde{x}^R_{NM} \end{pmatrix} \end{pmatrix}, \tag{6}$$

or in short hand notation:

$$\tilde{X}^{++} = [\tilde{X}^1, \tilde{X}^2, ..., \tilde{X}^R]. \tag{7}$$

Defining $P = [\rho^1, \rho^2, ..., \rho^R]'$, $\tilde{y} = [W^{+1}y, ..., W^{+R}y]$, $\Lambda = [\lambda^1, \lambda^2, ..., \lambda^R]'$ and over the more $\tilde{\epsilon} = [W^{+1}\epsilon, ..., W^{+R}\epsilon]$, and $\beta^{++} = [\beta^{\tilde{X}^1}, \beta^{\tilde{X}^2}, ..., \beta^{\tilde{X}^R}]'$ with $\beta^{\tilde{X}^r} = [\beta_1^{\tilde{X}^r}, ..., \beta_M^{\tilde{X}^r}]$ we can rewrite our model 2 as:

$$y = \tilde{y}P + X\beta^X + \tilde{X}^{++}\beta^{++} + \tilde{\epsilon}\Lambda + \kappa \tag{8}$$

with $\kappa \sim N(0, \sigma^2\Omega)$. For $R = 1$ model 2 follows directly. From the general model 8 we can derive three major submodels for $r = \{1, ..., R\}$: the spatial lag (SAR(r)) and spatial error (SEM(r)) and a spatial model with exogenous spatial variables (SEV(r)). For the (SAR(r)) we can write:

[24] Refer to (Anselin, 1992).

$$y = \tilde{y}P + \kappa \tag{9}$$

with $\kappa \sim N(0, \sigma^2\Omega)$, for the (SEM(r)) we can write

$$y = X\beta^X + \tilde{\epsilon}\Lambda + \kappa \tag{10}$$

with $\epsilon = \tilde{\epsilon}\Lambda + \kappa$ and $\kappa \sim N(0, \sigma^2\Omega)$ and for the (SEV(r)) we notate:

$$y = X\beta^X + \tilde{X}^{++}\beta^{++} + \kappa \tag{11}$$

with $\kappa \sim N(0, \sigma^2\Omega)$.

It has to be pointed out, that the estimation of 8 and its submodels 9, 10 and 11 could lead to biased and inconsistent OLS estimates. Take submodel 9 for instance: $P\tilde{y}$ is correlated not only with κ but also with neighbourings κ. If all elements of $\tilde{y}P$ are zero OLS estimates are unbiased but inefficient. If submodel 11 is chosen, then the model contains only exogenous spatial lagged variables besides non spatial lagged exogenous variables. In this case OLS is only BLUE if $\kappa \sim N(0, \sigma^2 I)$. OLS is even more unbiased if estimating a spatial error model, thus referring on submodel 10.

To test this spatial model, we regress the regional output, measured as gross value added on regional R&D-effort, human capital, regional number of patent applications, regional capital stock, regional number of low qualified labour force, regional infrastructure, spatial weighted gross value added, spatial weighted dependent variables and a West-East dummy, which covers the fact that East German regions are less productive than West German regions. Additionally, the number of patent applications are regressed on regional R&D output, as proposed by (Griliches, 1979). In this way it is possible to cover "articulated knowledge" and "tacit knowledge".[25] The question which remained unanswered is, how to choose the order R. If one refers to the literature there is no hint how to choose the order R. Regarding this subject,

[25] Refer to (Maurseth and Verspagen, 2002).

(Anselin, 1992) argues that especially for small samples the order of the weighting matrix W should be chosen small. As mentioned above, in this investigation we base the order of R on the data, especially on Moran's $\mathcal{I}$. But before checking the data concerning spatial dependencies, we should throw a first glance at the data.

5 Data and variables

Before testing the model, which has been introduced in the preceding chapter, one has to give a short description of the data. As mentioned before, NUTS-2 data for all German regions for the year 2003 have been used. The reason why one should decide to base the empirical study upon NUTS-2 data is, that referring on so called "Kreisdaten" could result in spurious spatial dependence, which could be caused by streams of commuters, for example.[26] This problem is boosted by the empirical fact of suburbanization, which has increasingly appeared in the last years.[27] That is why most similar research field studies refer to so called "land use planning units", such as NUTS-regions, particularly for European studies or "Arbeitsmarktregionen" for German investigations. Whatever of the latter mentioned spatial unit one decides to use, the worth mentioning communality is, that a "land use planning unit" subsumes smaller subgroups, such as "Kreise". Thus, referring to "land use planning units", the spurious spatial dependence problem is from less importance or even canceled out. The year 2003 was selected because of reliability and accessibility of European patent data. Particularly the problem of missing data is serious for NUTS-2 data. Of course, if data would have been available for a longer period of time, then regression based on time averages would be the appropriate approach. For Germany, 39 NUTS-2 regions are available for regression analysis.

The data stem from the online database provided by Eurostat, from the online support of the German statistical office in Wiesbaden (genesis online), from the online representation of the "Arbeitskreis "Volkswirtschaftliche Gesamtrechnungen der

[26](Keilbach, 2000), p. 120-121.

[27]Refer to (Kühn, 2001) and (Kaltenbrunner, 2003) for a discussion.

Länder"" as well as from the INKAR-database CD-Rom published by the "Bundesamt für Bauwesen und Raumordnung".

6 A first hint for spatial knowledge diffusion: a descriptive view

After describing the data set, this section should provide us a first guess concerning the existence of knowledge diffusion phenomena in the data. The traditional way detecting spatial phenomena in the data is to compute the so called Moran's $\mathcal{I}$, which is defacto "the" standard instrument in spatial econometrics for detecting spatial correlation[28] coefficient.[29]

The interpretation of the spatial correlation coefficient based on Moran´s $\mathcal{I}$ is a priori similar to time series analysis context. But it is not the same: Autocorrelation in time series means proximity of variables in time. Autocorrelation in space instead means geographic proximity of variables which is often two-dimensional. The important difference between the time series and the spatial econometric context is that spatial correlation has the attribute that a spatial event can be described via feedback loops, whereas time series correlation goes only in one direction, that is time. The interpretation of spatial correlation is quiet easy: if negative spatial correlation is observed, then regions are dissimilar with respect to their economic performance, whereas if positive spatial correlation is observed, then regions are similar with respect to their economic performance. The aim of the Moran´s $\mathcal{I}$ analysis is to measure the strength of spatial correlation and to find a hint how far spatial correlation spreads.

The Moran´s $\mathcal{I}$ is defined as follows:

$$\mathcal{I} = \frac{N}{O}\frac{e'W^{+r}e}{e'e}, \tag{12}$$

[28] Refer to (Moran, 1948) and (Moran, 1950).

[29] This is most used indice for detecting spatial phenomena. Despite Moran´s $\mathcal{I}$, other indices such as Geary´s $\mathcal{C}$ and Ripley´s $\mathcal{K}$. But the two latter are seldom used.

with O as the sum over all elements in W^{+r} and N as the number of observations. Of course if $\frac{N}{O} = 1$ we have a row standardized weighting scheme. e are the residuals obtained from an OLS estimation of a variable V on its spatial counterpart VW^{+r}. To center 12 around zero we follow (Ord, 1975) and standardize 12:

$$\tilde{\mathcal{I}} = \frac{\mathcal{I} - E(\mathcal{I})}{\sqrt{Var(\mathcal{I})}}, \tag{13}$$

with

$$E(\mathcal{I}) = \frac{N}{O}\frac{tr(V^{++}W^{+r})}{N-K},$$

and

$$Var(\mathcal{I}) = \left\{\frac{N}{O}\right\}^2 \frac{\left\{tr(V^{++}W^{+r}V^{++}W^{+r\prime}) + tr(V^{++}W^{+r})^2 + [tr(V^{++}W^{+r})]^2\right\}}{(N-K)(N-K+2)} - [E(\mathcal{I})]^2,$$

with $V^{++} = I - V(V'V)^{-1}X'$ as the projection matrix. In this way $\tilde{\mathcal{I}}$ is normal distributed.

Before computing Moran´s $\mathcal{I}$ for the desired variables, we should first have a look at the data. As mentioned above, we try to estimate a standard production technique to investigate the effects of spatial knowledge spillovers on labour productivity. Table 1 and 2 provide an overview of the data used in the analysis.

	Y	K	L	H
Mean	48795.52	266105.80	627841.30	8.49
Modus	–	–	–	–
Median	41022.01	228133.0	544004.00	8.38
Max	140902.40	895491.10	1603418.00	14.01
Min	9963.63	66538.54	135678.00	4.26
Std. Dev.	33057.33	177768.70	350356.20	2.66
Skewness	1.44	1.63	1.17	0.38
Kurtosis	4.23	5.66	3.62	2.15
Observations	39	39	39	39

Table 1: Table of descriptive statistics (I) of variables used for the analysis

	P	I	R&D	Den	Dummy
Mean	330.06	43891.18	1333.87	432.76	–
Modus	–	–	–	–	1.00
Median	223.28	43586.00	612.69	211.60	–
Max	1486.63	76028.00	7035.16	3803.00	1.00
Min	26.16	4785.00	67.64	74.99	0.00
Std. Dev.	352.11	17383.47	1592.97	698.53	–
Observations	39	39	39	39	39

Table 2: Table of descriptive statistics (II) of variables used for the analysis

From table 1 and 2 we can see that all variables exhibit positive skewness, what means that the distribution has a long right tail. This is especially true for the variable density (Den) but not astonishing, because we have a few high densely populated areas such as Berlin, Hamburg and Bremen. Additionally the distributions are peaked, which means they are leptokurtic relative to the normal distribution.
The computation of Moran's $\mathcal{I}$ for $v = \{y, h, p, r\&d, i\}$ is done with a program written in `R`, version 2.6.2.[30]. After completing the computation with `R`, figure 1 gives a graphical interpretation of spatial dependence between the v and its spatial lagged counterpart W^+v. Note that the variables are mean standardized, as mentioned before. Thus, besides a regression line the standardization allows us to plot one and two standard deviations areas. The interpretation of figure 1 and figure 2 is as follows: every subgraph is divided into four areas: the first area is located in North-East direction, the second in North-West direction, the third in South-West direction and the last in South-East. The first area contains positive standard deviation from a region i and its corresponding neighbour j. On contrary the third area contains negative standard deviation from region i and its corresponding neighbour j. All other areas contain couples of negative standard deviation of region i and positive standard deviation of region j and vice versa. Thus, we have a positive spatial correlation if regions are located in the first and in the third area. Otherwise we have a negative spatial correlation. With other words: If the slope in the scatter

[30]The source code is available on request.

plot is negative that means that we have a sort of checkerboard pattern or a sort of spatial competition in which high standard deviation regions are clustered with low standard deviation regions. Alternatively, if the slope is positive, we find the contrary.

If we now have a look at figure 1 we see first that positive spatial correlation is significant on a 5% significance level for the output y and for the patents p. Despite the fact that $r\&d$, human capital h and infrastructre i exhibit positive spatial correlation as expected, the Moran's $\mathcal{I}$ is not significant on a 10% significance niveau for $r = 1$. Next, the degree of spillover is boosted to $r = 2$ and again the Moran's $\mathcal{I}$ coefficient for each variable is computed.

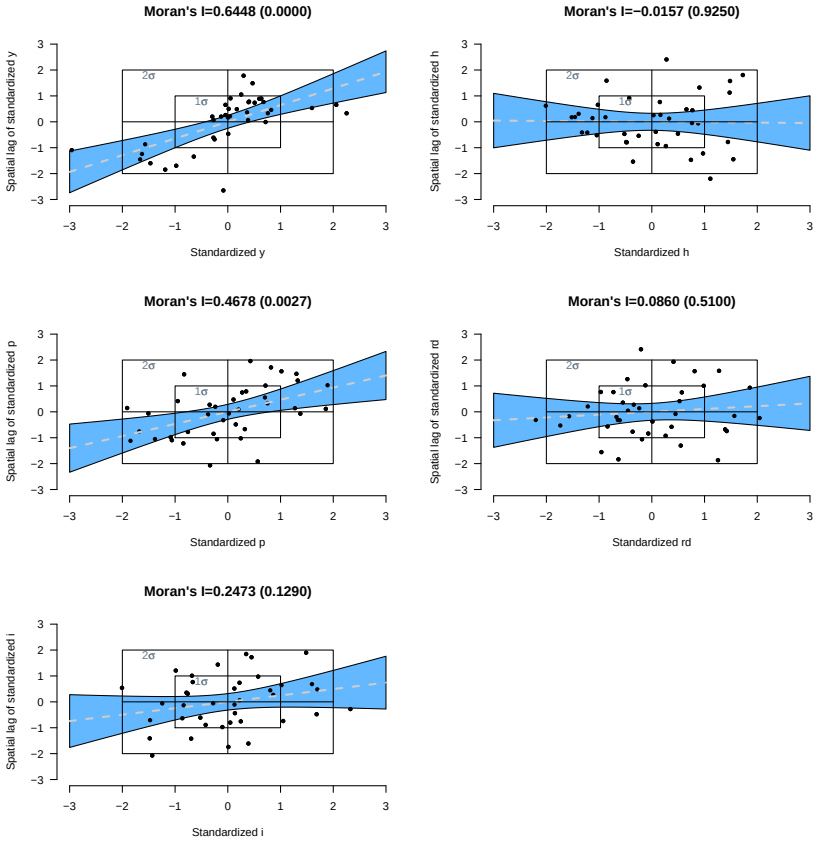

Figure 1: Computation of Moran's $\mathcal{I}$ with corresponding p-values for dependent and independent variable for $r = 1$

On the next step we take the weighting scheme to the power of two and additionally

compute the Moran's $\mathcal{I}$ for every variable. The result of this computation can be found in figure 2. The interpretation is equal to the preceding analysis.

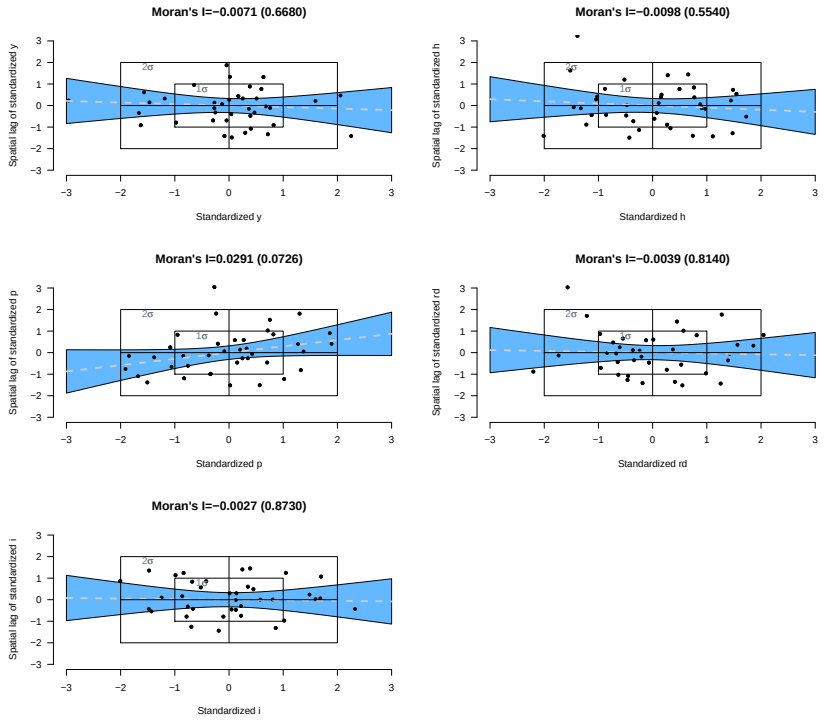

Figure 2: Computation of Moran's $\mathcal{I}$ with corresponding p-values for dependent and independent variable for $r = 2$

If we look at the sub pictures of figure 2 we find that only the spatial correlation of patents p is significant on a 10% significance niveau. All other variables do not exhibit significant spatial correlation. Therefore, we have to conclude that knwoledge spillovers, proxied by p, h and $r\&d$ are limited regarding space and in consequence more or less local and restricted to the nearest neighbours. Hence, we should acknowledge first order and second order degree of knwoledge spillover in the regression analysis. Additionally, we see some evidence from figures 1 and 2 that spatial outliers exits[31], which implies that spatial heterogeneity matters.

[31] Outliers are defined as data points which are situated outside the 2σ area.

7 Spatial model estimation

In this section a spatial model estimation strategy is introduced, which is an expansion of the proposed strategy by (Florax et al., 2003). Before introducing the new estimation method, the classic method of (Florax et al., 2003) for cross section analysis is briefly sketched. First, one has to start by estimating an initial model $y = X\beta + \epsilon$. Second, on the basis of the estimated model, Lagrange Multiplier tests are used to test for for spatial lag or spatial error model. If the null hypothesis is rejected, than spatial dependence matters and an appropriate spatial error or spatial lag model should be estimated. If we further acknowledge higher order spatial effects, the test statistic under the null hypothesis $H := \rho^r = 0$, $\forall r$ for $\mathcal{LM}_{\rho^r}$ can be written in the following way $r = \{1, ..., R\}$:

$$\mathcal{LM}_{\rho^r} = \frac{\left(\frac{e'W^{+r}e}{s^2}\right)^2}{T}, \tag{14}$$

with T as the trace of $(W^{+r\prime} + W^{+r})W^{+r}$, $e = My$ the residuals of regression, $M = I - X(X'X)^{-1}X'$ as the projection matrix and $s^2 = \frac{e'e}{N}$ as the estimated variance of the error term and N the number of observations. On contrary, the test statistic for $\mathcal{LM}_{\lambda^r}$ is, given $r = \{1, ..., R\}$, under $H := \lambda^r = 0$, $\forall r$ can be written as:

$$\mathcal{LM}_{\lambda^r} = \frac{\left(\frac{e'W^{+r}y}{s^2}\right)^2}{NJ}, \tag{15}$$

with $J = \frac{1}{Ns^2}\left[(W^{+r}Xb^{+++})'M(W^{+r}Xb^{+++}) + Ts^2\right]$ and b^{+++} as the OLS estimator of model 7.

Third, if for $\mathcal{LM}_{\rho^r}$ and $\mathcal{LM}_{\lambda^r}$ each the null hypothesis cannot be rejected, then the initial model should be used. Otherwise one should compare both test statistics. If they are both significant, one has to compute additionally the robust versions of $\mathcal{LM}_{\rho^r}$ and $\mathcal{LM}_{\lambda^r}$ to come to a final decision. If only one test is significant, then one has to adopt the initial model with respect to the significant test statistic.

The robust variant of $\mathcal{LM}_{\rho^r}$ read as:

$$\tilde{\mathcal{LM}}_{\rho^r} = \frac{\left(e'W^{+r}y - \frac{e'W^{+r}e}{s^2}\right)^2}{NJ - T}, \tag{16}$$

For the robust variant of $\mathcal{LM}_{\lambda^r}$ we can write:

$$\mathcal{L}\tilde{\mathcal{M}}_{\lambda^r} = \frac{\left(\frac{e'W^{+r}e}{s^2} - T(NJ)^{-1}\frac{e'W^{+r}y}{s^2}\right)^2}{T[1-T(NJ)]^{-1}}. \tag{17}$$

If $\mathcal{L}\tilde{\mathcal{M}}_{\rho^r} > \mathcal{L}\tilde{\mathcal{M}}_{\lambda^r}$, then one should decide to estimate a spatial lag model otherwise if $\mathcal{L}\tilde{\mathcal{M}}_{\rho^r} < \mathcal{L}\tilde{\mathcal{M}}_{\lambda^r}$ then one should refer to a spatial error model. Given, only $\mathcal{LM}_{\rho^r}$ is significant but $\mathcal{LM}_{\epsilon}$ is not, then one should use a spatial lag model, otherwise, if $\mathcal{LM}_{\lambda^r}$ is significant, then a spatial error model should be chosen. Further, it should be kept in mind, that experimental based simulations by (Anselin and Florax, 1995b) and (Anselin et al., 1996) found evidence, that robust counterparts of the $\mathcal{LM}$-tests have more power in pointing out the appropriate alternative than the non robust $\mathcal{LM}$ versions. But as shown by (Florax et al., 2003), the classical top down approach, that means relying on the non robust $\mathcal{LM}$ test, outperforms the robust strategy in means of performance and accuracy. Thus, the same authors emphasise, that one should use the classic approach when testing for spatial effects. It should be further noted that, although this strategy is not theoretically justified yet, it is the only systematic approach of model selection in literature and used in empirical studies.[32] The estimation strategy proposed by authors such as(Anselin, 2005) has three main drawbacks: first, the strategy lacks regarding their underlying tests hypothesis. For both tests, the LM_{ρ^r} and $\mathcal{LM}_{\epsilon}$ or in their robust form $\mathcal{L}\tilde{\mathcal{M}}_{\rho^r}$ and $\mathcal{L}\tilde{\mathcal{M}}_{\epsilon}$ the null hypothesis is either $H_0 := \rho^r = 0$ for $\mathcal{LM}_{\rho^r}$ or $\mathcal{L}\tilde{\mathcal{M}}_{\rho^r}$ and $H_0 := \lambda^r = 0$ for $\mathcal{LM}_{\lambda^r}$ or $\mathcal{L}\tilde{\mathcal{M}}_{\lambda^r}$. The null hypothesis $H_0 := \lambda^r = 0$ is realized in presence of ρ^r for the spatial error and $H_0 := \rho^r = 0$ in presence of λ^r for the spatial lag model. Although, robust $\mathcal{LM}$ tests are available, only one test is available, to compare the two models directly. This test, developed by (Mur, 1999) and (Trivezg, 2004) allows us to differentiate between spatial lag and spatial error models. But a drawback of the test proposed by (Trivezg, 2004) is, that it is only applicable for small samples, because it requires the computation of Eigenvalues and Eigenvectors of the underlying spatial weight matrix, which is cumbersome or even not possible for large data sets as noted by (Kelejian and Prucha, 1998).

[32] Refer to (Kim et al., 2003) for instance.

Second, the strategy is exclusive in the way, that this strategy does not allow for a ARMA(p,q) model specification, which is as mentioned above, a combination of spatial lag and spatial error model. There is no reason, why one should exclude this combination ex ante. This could create a serious problem, because even if λ^r differs significant from zero but the robust $\mathcal{LM}_{\rho^r}$ test, which exceeds the value of the robust $\mathcal{LM}_{\lambda^r}$ statistic, suggests to model a spatial lag model, we should choose, going in line with (Florax et al., 2003), a spatial lag model. It is obvious, that there is an inherent potential of misspecification using the strategy proposed by (Florax et al., 2003).[33]

Third, both tests, if robust or not do not sufficiently control for heterogeneity of the error term nor do they cover the aspect of outliers. In other words, this methods neglect spatial heterogeneity entirely. Fortunately, spatial heterogeneity can be elegantly considered in an Bayesian approach.

Until today, Bayesian model selection criteria are seldom used in empirical applications. This might be due to three reasons: first, normally, spatial Bayesian model techniques are not included in standard econometricians tools, such as *EViews*. Second, these methods require extended programming techniques. In addition, their use for large sample applications is problematic, because then one is often confronted with numerical problems, especially in calculating the determinant of spatial weight matrix[34]. Third, Bayesian methods are often rejected or disregarded by the class of frequentest or "main stream" econometricians, mainly because of the Bayesian assumption that the vector of coefficients is treated as random, whereas the frequentest treat the vector of coefficients estimate as random.[35]

In this application, both views should be acknowledged, the frequentest based Maximum-Likelihood estimation techniques and Bayesian methods. It should be clear that both methods exhibit advantages and disadvantages, but to acknowledge them within the

[33]For example, assume $\mathcal{L}\tilde{\mathcal{M}}_{\rho^r}$ statistic takes the significant value x and $\mathcal{L}\tilde{\mathcal{M}}_{\lambda^r}$ statistic takes the significant value $x + \epsilon$, with a very small but positive value $\epsilon > 0$. In this case we conclude to use the spatial error model, because $\mathcal{L}\tilde{\mathcal{M}}_{\lambda^r} > \mathcal{L}\tilde{\mathcal{M}}_{\rho^r}$.

[34]To avoid this problem either rely on Bayesian methods or use the Monte Carlo based method proposed by (Barry and Kelley, 1999).

[35]See (Koop, 2003) for an excellent introduction to Bayesian Econometrics.

interpretation the strategy should improve the strategy of (Anselin, 2005), because of the above mentioned advantages of the Bayesian methods, especially their heteroscedastic formulation. The strategy can be formulated as follow:

1. *First, estimate the initial model via OLS.*

2. *Use Moran's $\mathcal{I}$ and $\mathcal{LM}$-test for detecting potential spatial dependence. If the proposed tests cannot reject the null hypothesis of no spatial correlation, then select the model estimated via OLS in step 1. Otherwise, proceed with step 3.*

3. *If the null hypothesis of no spatial correlation is rejected, then expand the model estimated in step 1 by adding spatial counterparts of the independent variables. Perform an OLS estimation of this model.*

4. *Given the model setup in step 3, use Moran's $\mathcal{I}$ and $\mathcal{LM}$-test for detecting potential spatial dependence. If tests cannot reject the null hypothesis of no spatial correlation, then select the model estimated via OLS in step 3. Otherwise, proceed with step 5.*

5. *Expand the model of step 3 with spatial error and spatial lag components. Again, Perform an OLS estimation of this model.*

6. *Use Moran's $\mathcal{I}$ and $\mathcal{LM}$-test for detecting potential spatial dependence. If the tests cannot reject the null hypothesis of no spatial correlation, then select the model estimated via OLS in step 5. Otherwise, proceed with step 7.*

7. *Estimate a general spatial model (SAC) and separate spatial lag (SAR) and spatial error models (SEM) with MLE. OLS would yield in this case inconsistent parameter estimates even if spatial homogeneity is assumed.*

8. *Use the $\mathcal{LM}$ power comparison mentioned by (Florax et al., 2003) to select the optimal model from the set of models estimated in step 7. Note, this model assumes spatial homogeneity.*

9. *Given the optimal model found with step 8, estimate the Bayesian counterpart of the optimal model selected in step 8 to control for spatial heterogeneity. If*

both models exhibits similar results and spatial heterogeneity is rejected, then take the optimal model found in step 8 as optimal. Otherwise, if spatial heterogeneity matters, take the Bayesian model as the optimal one.

It is worth to mention, that Moran's $\mathcal{I}$ is valid, as long as heteroscedasticity is not spatial correlated. This is a very new insight, but until today no appropriate method is developed to test for spatial correlated heteroscedasticity. There is only one test proposed by (Kelejian and Robinson, 2004), which cover the aspect of spatial correlated heteroscedasticity, but it is only valid for large samples and small samples properties are not known.

8 Initial model estimation

Let us start with the first step of the laid out strategy. First, we estimate the initial model with ordinary least square procedure.[36] The initial model, based on a per head Cobb Douglas production technique, with $ln\left(\frac{Y}{L}\right)$ as the dependent variable, can be written in log-log form as follows:

$$\ln(y) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + d\gamma + \kappa \quad (18)$$

or in a more compact manner as

$$y = X\beta^X + d\gamma + \kappa \quad (19)$$

with $\beta^X = [\beta^c, \beta^k, \beta^l, \beta^h, \beta^p, \beta^i]$ and $X = [1, k, l, h, p, i]$ with $\kappa \sim (0, \sigma^2\Omega)$, $\sigma^2\Omega \neq \sigma^2 I$, $\Omega = diag(v_1, ..., v_N)$ and d as West-East dummy. Two remarks regarding the specification of equation 18 or equation 19: First, as usual, the coefficient vector β^X contains constant production elasticities of the respective values stacked in X. Because we estimate a production technique per capita, the depended variable is $y = ln\left(\frac{Y}{L}\right)$. Thus the elasticity of production for labour l in this context is defined as $\beta^l + 1$. Therefore, we expect a negative sign of β^l . Second please note, that the

[36] All estimations have been performed with *Matlab* on the basis of the package provided by LeSage with some adoptions. $\mathcal{LM}$ program for spatial lags as other programs are available on request. If appropriate, results have been checked with `R 2.6.2` and `EViews 5.0`.

inclusion of both $R\&D$ expenditures and P leads to a serious endogenity problem, because patents are produced with $R\&D$ expenditures or $P = u(R\&D)$ with $u(\cdot)$ as continuous function. It is worth to mention that patents generally outperforms $R\&D$ expenditures regarding their interpretation as a quality measure of innovativeness.[37]

In table 3 one can find four different specifications. For every specification $\mathcal{LM}$ tests have been conducted, both for spatial lag and spatial error. Additionally, the test statistics for first order and second order spatial influence have been computed.[38] Further, Moran's $\mathcal{I}$ test has been performed, also for first and second order spatial influence.

Column (1) of table 3 reports a simple estimation of y on k and l and a West-East dummy d. The values of the elasticity of production for capital and labour indicate the expected positive sign and have the expected dimension.[39] and have the correct dimension regarding their influence on per capita production. Furthermore, the dummy is positive as expected and highly significant which indicates that West German regions are more productive on average than East German regions. As we can see from column (1) of 3, both Moran's $\mathcal{I}$ tests cannot reject the null hypothesis of no spatial correlation. Also the $\mathcal{LM}$ lag for $r = 1$ and $r = 2$ are not significant. This is again the case for the $\mathcal{LM}$ error test for $r = 1$. For $r = 2$ the $\mathcal{LM}$ error test of no spatial correlation under the null hypothesis can be rejected at a 5% significance level.

Although, we find a contradiction regarding the evaluation of Moran's $\mathcal{I}$ for $r = 2$ and the $\mathcal{LM}$ error test for $r = 2$ with respect to spatial influence we should expand the estimation and include the knowledge variables human capital h and patents p. Further infrastructure i as additional regressor has been included. The estimation results of this expanded specification can be found in column (2) of table 3. For all three additional included coefficient regressors we should expect a positive sign. This is true for the estimated coefficients of human capital and infrastructure, but

[37] Refer (Lechevalier et al., 2007) for instance.

[38] For example $\mathcal{LM}_\lambda^2$ stands for a test of no spatial correlation up to order $r = 2$ for spatial error component.

[39] The value for the elasticity of production for labour is 1-0.19=0.81.

not for patents, which is contra intuitive at first glance. But looking at significance we find, that patents are not significant, not even at a 10% significance level. This is also true for infrastructure which is not significant at a 10% significance level. Additionally, looking again on the coefficent for patents the influence of own patents on own labour productivity is at least zero. Refering to the test statistics, it should be noted, that the $\mathcal{LM}$ test for spatial lag is significant at a 5% significance level. Moran's $\mathcal{I}$ for $r = 1$ suggests, that a spatial error model should be estimated which is underpined by the siginifcant $\mathcal{LM}$ test for the spatial error component for $r = 2$. Given our estimation strategy, we should expand our model by exogenous spatial lagged variables. The advantage of this formulation is straightforward: the estimators of this estimation are unbiased using OLS. Keeping in mind our results obtained from picture 1 and 2 we include first order spatial lags of human capital $\ln(H^{+1})$, of patents $\ln(P^{+1})$ and of infrastructure $\ln(I^{+1})$ and in addition the second order lag of patents $\ln(P^{+2})$. Stacking this values in $\tilde{X}^1 = [h^{+1}, p^{+1}, i^{+1}]$ and $\tilde{X}^2 = [p^{+2}]$ defining $\tilde{X}^{++} := [\tilde{X}^1, \tilde{X}^2]$ and letting $\beta^{++} = [\beta^{\tilde{X}^1}, \beta^{\tilde{X}^2}]'$ with $\beta^{\tilde{X}^1} = [\beta^{h^{+1}}, \beta^{p^{+1}}, \beta^{i^{+1}}]$ and $\beta^{\tilde{X}^2} = [\beta^{p^{+1}}]$, this leads to the following expansion of equation 18:

$$\ln(y) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + \tag{20}$$

$$+ \ln(H^{+1})\beta^{H^{+1}} + \ln(P^{+1})\beta^{P^{+1}} + \ln(I^{+1})\beta^{I^{+1}} + \ln(P^{+2})\beta^{P^{+2}} + d\gamma + \kappa,$$

or again in compact notation:

$$y = X\beta^X + \tilde{X}^{++}\beta^{++} + d\gamma + \kappa \tag{21}$$

with $\beta^X = [\beta^c, \beta^k, \beta^l, \beta^h, \beta^p, \beta^i]$, d as West-East dummy and $X = [1, k, l, h, p, i]$ with $\kappa \sim (0, \sigma^2 I)$.

The estimation results for 21 can be found in table 3 in column (3). Once again, we would expect positive effects from neighbouring regions. But with the exception of patents, we find negative signs of coefficients for neighbouring human capital and neighbouring infrastructure. Over the more the latter two coefficients are highly

non significant. The negative second order spillover coefficient of patents is highly insignificant, too. As the coefficient for the own patents, this second order coefficient of neighbouring patents is close to zero. But what can we see is, that the first order neighbouring patent activity has a significant positive effect on own productivity. If we look at our test statistics in column (3) we find that the $\mathcal{LM}$ test for spatial lag is, on contrary to column (2), not significant anymore. This could be due to the inclusion of the spatial lagged patent activity. Furthermore, the second order $\mathcal{LM}$ error test is still significant at a 10% significance level, whereas the first order $\mathcal{LM}$ error test is now significant at a 5% significance level. Also the first order Moran's $\mathcal{I}$ test is significant at a 5% significance level. This lead us to conclude that a first order spatial error model should be modeled, because of the fact that $\mathcal{LM}_{\lambda^1} > \mathcal{LM}_{\lambda^2}$. The last column of table 3 shows the same regression as in column (3) but with the exclusion of the highly non significant spatial second order patent activity. If we compare column (3) and column (4) we can assert, that the exclusion of spatial second order patent activity does not change the sign and significance of the regression. Therefore, we should proceed with the specification which can be found in column (4) in 3.

In summary, we can conclude from 3 that spatial processes can be detected in the data. In consequence, we have to acknowledge them in our regression equation and in an adequate estimation procedure. From column (4) in table 3 we further know, that spatial dependence in the error term should be acknowledged. What we do not know up to this stage is, if spatial heterogeneity matters. This topic is treated in the next section.

9 Expansion of the initial model

From the discussion before we know that we have to expand our regression equation in 3 by an spatial lagged error term. Therefore, we have to reformulate our regression model 20 or 21 as a spatial error model (SEM). This is done with equation 22:

dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$				
independent variables $\mathbf{x} \in \mathbf{X}$	OLS	OLS	OLS	OLS
Column	(1)	(2)	(3)	(4))
Constant	10.50177	10.72613	10.39325	10.38857
	(0.0000)$^{\diamond}$	(0.0000)	(0.0000)	(0.0000)
ln(K)	0.251596	0.237242	0.282934	0.282665
	(0.0124)	(0.0139)	(0.0109)	(0.0088)
ln(L)	-0.193904	-0.219739	-0.243889	-0.243392
	(0.0483)	(0.0264)	(0.0186)	(0.0143)
ln(I)	—	0.014814	0.016332	0.016466
	(—)	(0.5218)	(0.4653)	(0.4673)
ln(H)	—	0.149377	0.161219	0.161679
	(—)	(0.0290)	(0.0076)	(0.0046)
ln(P)	—	-0.002382	-0.023189	-0.023487
	(—)	(0.9373)	(0.5216)	(0.5011)
$\ln(H^{+1})$	—	—	-0.066795	-0.066647
	(—)	(—)	(0.6357)	(0.6286)
$\ln(P^{+1})$	—	—	0.054656	0.054706
	(—)	(—)	(0.0453)	(0.0391)
$\ln(I^{+1})$	—	—	-0.019147	-0.019317
	(—)	(—)	(0.8033)	(0.8022)
$\ln(P^{+2})$	—	—	-0.000197	—
	(—)	(—)	(0.9726)	(—)
d	0.218824	0.273835	0.226299	0.226651
	(0.0000)	(0.0002)	(0.0061)	(0.0056)
Moran-$\mathcal{I}_1$	0.96	2.34	3.19	3.29
	(0.2506)	(0.0253)	(0.0024)	(0.0018)
Moran-$\mathcal{I}_2$	-0.11	0.26	0.21	0.27
	(0.3967)	(0.3860)	(3879)	(0.3850)
$\mathcal{LM}_{\lambda^1}$	0.21	2.42	4.90	4.89
	(0.6483)	(0.1201)	(0.0268)	(0.0270)
$\mathcal{LM}_{\lambda^2}$	5.25	3.74	3.52	3.53
	(0.0219)	(0.0532)	(0.0601)	(0.0601)
$\mathcal{LM}_{\rho^1}$	1.21	5.58	1.82	1.81
	(0.2800)	(0.0184)	(0.1775)	(0.1782)
$\mathcal{LM}_{\rho^2}$	1.82	0.11	0.97	0.01
	(0.1774)	(0.7350)	(0.3236)	(0.9202)
Observations	39	39	39	39
adjusted R^2	0.69	0.74	0.75	0.76

$^{\diamond}$White heteroscedasticity-consistent p-values in ().

Table 3: Results of OLS estimation for German NUTS-2 regions

$$\ln(y) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + \tag{22}$$

$$+ \ln(H^{+1})\beta^{H^{+1}} + \ln(P^{+1})\beta^{P^{+1}} + \ln(I^{+1})\beta^{I^{+1}} + d\gamma + \epsilon,$$

with $\epsilon = \lambda_1 W^{+1}\epsilon + \kappa$ or again in compact notation:

$$y = X^{+++}\beta^{+++} + d\gamma + \tilde{\epsilon}\Lambda + \kappa, \tag{23}$$

with $\beta^X = [\beta^c, \beta^k, \beta^l, \beta^h, \beta^p, \beta^i]$, $X = [1, k, l, h, p, i]$, $\Lambda = [\lambda^1]$, $W^{++} = [W^{+1}]$, $X^{+++} = [X, \tilde{X}]$, $\beta^{+++} = [\beta^X, \beta^{++}]$, with $\kappa \sim (0, \sigma^2 I)$ and d as West-East dummy. Model 23 should be estimated via two different ways:

- *The first approach is to estimate this model with the assumption of $\sigma^2\Omega = \sigma^2 I$, implying spatial homogeneity, which is a common assumption in the relevant studies in this subject*[40]*. As mentioned above, model 22 should be estimated via ML.*

- *The second approach is to estimate this model with the assumption of $\sigma^2\Omega \neq \sigma^2 I$, implying spatial heterogeneity with a Bayesian approach which is laid out latter.*

If we go back to the first approach, first we have to set up our Likelihood function. This is:

$$\mathcal{L} = \frac{\left|\tilde{N}\right|}{(2\pi\sigma^2)^{\frac{N}{2}}} exp\left\{\frac{1}{2\sigma^2}(y - X^{+++}\beta^{+++})'\Theta^{-1}(y - X^{+++}\beta^{+++})\right\}, \tag{24}$$

with $\Theta^{-1} = \tilde{N}'\tilde{N}$ and $|\Theta|^{\frac{1}{2}} = |\tilde{N}|$ and N the numbers of observations.
The corresponding log-likelihood for 24 is

$$\ln \mathcal{L} = -\frac{N}{2}\ln 2\pi - \frac{N}{2}(\sigma^2) + \ln|\tilde{N}| - \frac{1}{2}\xi'\xi, \tag{25}$$

[40] Refer for instance to (Olejnik, 2008) or (Santolini, 2008).

with $\tilde{N} = (I - \lambda^1 W^{+1})$ and $\xi = \tilde{N}(y - X^{+++}\beta^{+++})$. This expression 25 can be written in concentrated form as

$$\ln \mathcal{L}_c \propto \ln |\tilde{N}| - \frac{N}{2}\tilde{\xi}'\tilde{\xi}, \tag{26}$$

with $\tilde{\xi} = \frac{1}{\sigma}\tilde{N}(y - X^{+++}\tilde{\beta}^{+++}_{ML})$. The obtained Maximum-Likelihood based estimators can be written as

$$\tilde{\beta}^{+++}_{ML} = (X^{+++\prime}\tilde{N}'\tilde{N}X^{+++})^{-1}X^{+++\prime}\tilde{N}'\tilde{N}y \tag{27}$$

and

$$\hat{\sigma}^2_{ML} = \frac{1}{N}(\tilde{\xi}'\tilde{\xi}), \tag{28}$$

obtained from maximizing 25. As we can see, equation 26 is highly non linear in the parameter λ^1. Because both β^{+++} and κ are a function of λ we should use an iterative method to estimate λ^1. An approach is to first, estimate β^{+++} via OLS, then find with the associated estimated residuals a value of λ^1 which maximizes the concentrated likelihood function 26, third update the OLS values of β^{+++}. With the new updated values of β^{+++} then estimate new λ^1, based on the updated estimated residuals. Convergence is achieved, if values for both residuals and for β^{+++} do not change anymore from one to the next iteration step, which means the difference between $\beta^{+++}_t - \beta^{+++}_{t-1} < \vartheta$ for a small value of ϑ near zero.[41]

It is worth to note, that refering on Maximum-Likelihood, we have to impose a restriction on the parameter λ^1. Referring to (Anselin and Florax, 1995a), p. 34, this parameter takes on feasible parameter values in the range of:

$$\frac{1}{\tilde{\lambda}^1_{min}} < \lambda^1 < \frac{1}{\tilde{\lambda}^1_{max}}. \tag{29}$$

$\tilde{\lambda}^1_{min}$ is the minimum Eigenvalue of the matrix W^{+r}, whereas $\tilde{\lambda}^1_{max}$ represents the maximum Eigenvalue of W^{+r}. This suggest a constrained Maximum-Likelihood maximization. If W^{+r} is row standardized, as it should be, then of course $\lambda^1_{max} = 1$.

[41] In this application ϑ is set to $\vartheta =$ `1e-8`. Further t is set to a maximum value of `500`.

Please note, that this procedure could become extremely laborouis with respect to computational issues. More precisely, the computational costs increase with the dimension of the weighting scheme matrix W^{+r}. Alternatively, one can set ex ante values for λ, such as $\lambda^1 \in (0, 1)$ which implies only positive spatial error dependence. In this work ex ante values for λ^1 ranging from $\lambda^1 \in (-1, 1)$ have been imposed, although a direct computation via Eigenvalues would be passable.

The second approach dealing with the estimation of model 23 is to refer on a Bayesian approach but with the additional assumption of spatial heterogeneity, which means that $\sigma^2\Omega \neq \sigma^2 I$. If the model yields the same results and spatial heterogeneity is insignificant, we can conclude, that spatial heterogeneity can be ignored, otherwise, there is at least little evidence that spatial heterogeneity a justified assumption and we have to control for it.

Based on the likelihood function expressed by equation 24 a spatial Bayesian heteroscedastic model is set up. The core of Bayesian econometrics is the Theorem of (Bayes, 1763) which is needed in this context for parameter estimation. Assume for a moment that θ is a vector of unknown parameters which should be estimated. Before any data are observed, we have beliefs and some uncertainty with respect to our vector of parameter θ. These beliefs are called "a priori" probabilities which are fully represented by the probability function $p(\theta)$. The entire probability model itself is totally defined by the likelihood $p(y\,|\theta)$. $p(y\,|\theta)$ can be described as the core of Bayesian econometrics, because it contains the entire set of information from the data. Given, we have observed y, then we should update our beliefs regarding θ. By using the theorem of Bayes we obtain the so called "a posteriori" distribution of θ, given y, which is

$$p(\theta|y) = \frac{p(y\,|\theta)p(\theta)}{p(y)}, \tag{30}$$

with $p(y) = (y|\theta)p(\theta)$, defined by the law of total probability. Because $p(y)$ do not contain any information regarding θ and, over the more, we only interesting in θ, we can ignore $p(y)$. Thus the "a posteriori" probability is proportional to the likelihood times the "a priori" probability:

$$p(\theta|y) \propto p(y\,|\theta)p(\theta). \tag{31}$$

Although the dimensionality of $p(\theta|y)$ depends on the number of unknown parameters, we can often focus on individual parameters such as $\theta_1 \in \theta$ by numerically or analytically integrating out other components[42]. For instance we can write:

$$p(\theta_1|y) = \int p(\theta|y)d\theta_2 d\theta_3 ... \tag{32}$$

The entire information needed for inference about θ_1 is contained in the marginal distribution of θ_1. What we have to do now is to specify our exogenous given priors and the likelihood function.

In this context, we assume normal priors for β^{+++} and a diffuse prior for σ. The relative variance terms $v_i \in \Omega$ are fixed but unknown and therefore we have to estimate them. We have to treat the v_i as informative priors. The distribution of all elements of Ω are assumed to be independently $\frac{\chi^2}{s}$ distributed, with $s \sim \Gamma(a, b)$. As mentioned we are confronted with a degree of freedom problem, if the number of estimated coefficients exceeds the number of observations. Considering the fact, that the χ^2 distribution is a single parameter distribution we are able to compute N additional parameters v_i by adding only one single parameter s to our model. This idea goes back to (Geweke, 1993) who uses this type of prior to model heteroscedasticity and outliers in a linear regression framework. The idea becomes more clear if one knows that the mean of this priors is unity, whereas the variance of this prior is $\frac{s}{2}$. Thus, if s takes a large value, then all terms of Ω tend to unity, yielding a homoscedastic scenario, because σ is weighted equally for every observation, hence we obtain a constant variance over space. An assumption, which is made within the traditional spatial Maximum-Likelihood approach. On contrary, small values of s lead to a skewed distribution. The role of v_i therefore is, as in a traditional GLS approach, to down weight observations with large variances. For this reason, the degrees of freedom s plays a crucial role when robustifying against

[42] Refer to (Geweke, 1993).

outliers. For $s \to \infty$ the limiting normal and therefore a homoscedastic "scenario" is realized. One option could be to assign a improper value to s. The other possibility is to use a proper prior for s which is Gamma distributed:

$$s \sim \Gamma(a, b), \tag{33}$$

with hyperparameter a and b. It has to point out, that the virtue of the first option is that less draws compared to the second option are required for parameter estimations and moreover convergence is quicker.
If $\Gamma(a = \frac{s}{2}, b = 2)$ this is equivalent to $\chi^2(s)$, hence we obtain a so called mixing distribution controlled by s. As shown by (Geweke, 1993) we can write

$$\pi\left(\frac{s}{v_i}\right) \sim \text{iid}\, \chi^2(s),\ \forall\, i, \tag{34}$$

with $\pi(\cdot)$ denoting the prior from now. This implies, that the normal mixture model with 34 is equivalent to a model based on independently distributed Student-t values with s degrees of freedom, known as the (Theil and Goldberger, 1961) Model. The spatial error parameter is assumed to follow an uniform, but proper distribution with the range $\hat{N}$ as $\pi(\lambda^1) = \frac{1}{\hat{N}} = \frac{1}{\tilde{\lambda}^1_{min} < \lambda^1 < \tilde{\lambda}^1_{max}} \sim \mathcal{U}[-1, 1]$.
Let us summarize our assumptions regarding the priors as follows:

$$\pi(\beta^{+++}) \sim \mathcal{N}(c, T), \tag{35}$$

$$\pi(\frac{s}{v_i}) \sim \text{ iid } \frac{\chi^2(s)}{s}, \tag{36}$$

$$\pi(\lambda^1) \sim \mathcal{U}[-1, 1]. \tag{37}$$

Given the priors defined above, we need the conditional posterior distributions for each parameter $\beta^{+++}, \sigma, \lambda^1, \Omega$ to estimate them. Using the priors, assuming that they are independent from each other, we can define the joint posterior as:

$$p(\beta, \sigma, \lambda^1) = p(\beta)p(\sigma)p(\lambda^1)$$

$$\propto \left|I - \lambda^1 W^{+1}\right| \sigma^{-N} exp\left\{-\frac{1}{2\sigma^2}(\xi'\Omega^{-1}\xi)\right\} \sigma^{-1} \exp\left\{-\frac{1}{2\sigma^2}(\beta - c)'T^{-1}(\beta - c)\right\}. \tag{38}$$

From 38 the conditional distribution of β^{+++} is obtained from the standard non spatial Bayesian GLS approach as:

$$p(\beta^{+++}|\lambda^1, \sigma, \Omega, y) \sim \mathcal{N}[H(X^{+++}\tilde{N}\Omega^{-1}\tilde{N}y + \sigma^2 T^{-1}c, \sigma^2 H)], \tag{39}$$

with $H = (X^{+++\prime}\tilde{N}\Omega^{-1}\tilde{N}X^{+++} + T^{-1})^{-1}$, $\tilde{N} = (I^1_\lambda W^{+1})$, mean c and the corresponding variance covariance matrix T.

The conditional distribution of σ is

$$p(\sigma|\lambda^1, \Omega, \beta^{+++}, y) \propto \sigma^{-(N+1)} exp\left\{\frac{1}{2\sigma^2}\xi'\Omega^{-1}\xi\right\}. \tag{40}$$

Next the conditional distribution of every element v_i of Ω is considered. (Geweke, 1993) shows, that the conditional distribution for $v_i \in \Omega$ represents a χ^2 distribution with $s + 1$ degrees of freedom:

$$p\left(\left[\frac{(\sigma^{-2}e_i^2 + s)}{v_i}\right]|\beta^{+++}, \lambda^1, v_{-i}, \lambda^1\right) \sim \chi^2(s + 1), \tag{41}$$

with $v_{-i} = \{v_1, ...v_{i-1}, v_{i+1}, ..., v_N\}$.

Now consider the conditional distribution for the parameter σ assuming that we already know the parameters, given we know β^{+++}, λ^1 and Ω. This distribution would be:

$$p\left[\sum_{i=1}^{N} \frac{e_i^2}{v_i}/\sigma^2|\beta^{+++}, \lambda^1, \Omega\right] \sim \chi^2(N). \tag{42}$$

With 42 we adjust estimated residuals e_i with estimated weights or relative variance terms v_i. This approach corresponds to the simple weighted least square procedure (WLS) known from basic econometricians toolbox.

Finally, the conditional posterior of λ^1 is calculated as follows:

$$p(\lambda^1|\sigma,\Omega,\beta^{+++},y) \propto |\tilde{N}|exp\left\{\frac{1}{2\sigma^2}\xi'\Omega^{-1}\xi\right\}. \tag{43}$$

With exception of 43, all other posterior distributions are standard and therefore a Markov Chain Monte Carlo method (MCMC) can be applied to estimate parameters $\beta^{+++}, \lambda^1, \sigma^2, \Omega$. Usually, a Gibbs sampling approach, which is based on the conditional posterior densities is used.

We wish to make several draws to generate a large sample from which we can approximate the posterior distributions of our parameters. Unfortunately, we cannot approximate a posterior distribution for expression 43, because this type of distribution do not correspond to any so called standard class of probability densities. For this reason, Gibbs sampling cannot be readily used. Fortunately, a method called "Metropolis-Hasting" sampling which is an additional sequence in Gibbs sampling procedure[43], allows us to approximate the posterior distribution for λ^1.[44] The only problem one has to solve is to find a suitable proposal density. (LeSage, 2000) suggests to assume a normal or Student t-distribution. Because of the fact, that λ^1 has to be handled as a restricted parameter, which is situated between minus one and one, the sampler rejects values outside the interval $(-1, 1)$ from the sample. This is called "rejection sampling".[45]

The "Metropolis-Within-Gibbs" sampling algorithm can be expressed as follows:

1. *Set t=0.*

2. *Define a starting vector $S_{t=0}$ which contains the initial parameter of interest:* $S_0 = [\beta_0^{+++}, \sigma_0^2, v_{i0}, \lambda_i^1]$.

3. *Compute the mean and variance of β^{+++} using 39 conditional on all other initial values stacked in S_0.*

4. *Use the computed mean and variance of β^{+++} do draw from a multivariate normal distribution a normal random vector β_1^{+++}.*

[43] Because of this reason, the method is also called "Metropolis-Within-Gibbs".

[44] Refer to (Gelman et al., 1995).

[45] Refer to (Gelfand et al., 1990).

5. *Calculate 42 refering on β_1^{+++} from step 4 and use this expression in combination with $\chi^2(N)$ random draw to determine σ_1^2 for $i = \{1, 2, ..., N\}$.*

6. *Use β_1^{+++} and σ_1^2 to calculate 41 and use this value together with a N-dimensional vector of $\chi^2(s+1)$ random draws to determine $v_i \in \Omega$ for $i = \{1, 2, ..., N\}$.*

7. *Use metropolis within Gibbs sampling to calculate λ^1 using values $v_i \in \Omega$ for $i = \{1, 2, ..., N\}$, β_1^{+++} and σ_1^2.*

8. *Set t=t+1.*

The question which remains is, how to select the correct Bayesian model. It is sometimes the case that several competing models M_u with $u = \{1, 2, ..., U\}$ exist. Then usually posterior probabilities are computed which should give advice, which model is the correct model in terms of probability. The posterior probability p_u^{pos} for model u is given by[46]:

$$p_u^{pos} \equiv p(M_u|y) = \frac{p(y|M_u)}{\sum_{u=1}^{U} p(y|M_u)}. \tag{44}$$

Bayesian model averaging suggests to weight all possible Bayesian models M_u with $u = \{1, 2, ..., U\}$ with their corresponding posterior probabilities. In terms of probability this means:

$$p(y^*|y) = \sum_{u=1}^{U} p(y^*|y, M_u)p(M_u|y), \tag{45}$$

with $p(y^*|y)$ as the posterior, $p(M_u|y)$ as the posterior model probability and $p(y^*|y, M_u)$ as the likelihood function of model M_u. The reason why model averaging should be used is quite simple. The traditional approach is to choose the single best model based on calculating posterior model probabilities with 44 for every model of interest.[47] But one has to remember that this rather excluding approach could be

[46] Please refer to (Hepple, 2004), p. 105.

[47] A large bulk of literature on Bayesian model averaging (BMA) over alternative linear regression models containing differing explanatory variables exists. For instance refer to (Raferty et al., 1997), (Fernandez et al., 2001b) and (Fernandez et al., 2001a). The MC^3 approach, is set forth for in (Madigan and York, 1995) for the SAR and SEM models.

lead to wrong decisions, because a researcher has to decide on the basis of model probabilities what is the "good model" and what is the "not so good model" from a sometimes large set of models. Additionally, only referring to the "good model"' ignores model uncertainty. In this study, relying on model probabilities is not a good idea, because "posterior model probabilities cannot be meaningful calculated with improper non informative priors,"[48] which are not common for all models. Therefore we refer to the MCMC literature to compute a posteriori model probabilities. This so called MC^3 approach, introduced by (Madigan and York, 1995) is based on a stochastic Markov Chain process which moves through the model space and samples those regions which has a high superior model support. Thus this approach is very efficient because not the entire model space is of interest.[49]

Knowing these facts, we are now able to interpret our estimation results for both approaches, the Maximum-Likelihood and the Bayesian approach. The results for the first approach can be found in column (1) and (2) of table 5. The first regression is a mixture model of spatial lag and spatial error model, the so called spatial ARMA model, which is in this case labeled as SEC(r,r) to avoid confusion with respect to time series context.[50] This regression is done to corroborate our model selection on inductive statistics, done in the forgoing chapter. After estimation of all possible combinations of first order and second order spatial models[51], we have chosen the SEC(1,1) model as the appropriate model on basis of the value of the log-likelihood. Leaving out the insignificant parameter ρ^1, estimating a pure spatial error model (column (2)) and comparing this with column (1) we can see, that only minor changes of coefficient values result. This is an indicator, that the spatial lag does not provide any further information for our model. Thus, it is justified, to model a spatial SEM(1) model, which is printed in column (2), because the spatial error coefficient λ^1 is highly significant. Comparing the SEM(1) model with the fourth column of 3 we can find moreover, that the coefficient for $\ln(I^{+1})$ is not positive,

[48](Koop, 2003), p. 268.

[49]Refer to (LeSage and Parent, 2007) for an excellent contribution to this topic.

[50]See appendix 1 for a deviation of the log-likelihood of the spatial ARMA model.

[51]See appendix 2 for a summary.

Bayesin model (4)[SEM(1)]	Model (4,1)	Model(4,2)	Model (4,3)	Model(4,4)
Runs	10,000	10,000	100,000	100,000
Informative Priors	No	Yes	No	Yes
p_u^{pos}	0.2770	0.2509	0.2374	0.2374

Table 4: MC^3 a posteriori model probabilities p_u^{pos} for variants of model (4)[SEM(1)]

but again highly non significant. All other coefficient have, compared to (4) in table 3, roughly the same dimension, the same sign and the same level of significance.

Additionally, the results for the second approach, an estimation of the Bayesian counterpart of equation 23 can be found in column (3) of table 5. Before discussing the results, we first should get an intuition of what is behind the Bayesian estimation approach.

To obtain estimates from our Bayesian approach we have to simulate draws. To ensure stability of simulated results, one should do a simulation on non informative priors and on informative priors, for which starting values are obtained from a corresponding Maximum-Likelihood estimation with different draws. For this reason, two Bayesian estimations, one with 10,000 draws and one with 100,000 draws, each with informative and non informative priors have been conducted. At all we get 4 models, for each number of draws one should estimate a model with informative and non informative priors.[52] The model probabilities p_u^{pos} for the relevant models can be found in table 4.

Calculating this probabilities and comparing them with each other, we find, that the first model $(4, 1)$ has slightly a higher probability to be the correct model.

Furthermore, MCMC-convergence checks the four relevant models have been performed.[53] to ensure convergence of the sampler. If the means and variances for the posterior estimates are similar from all runs, convergence seems ensured at all.

[52]Because of the fact, that the initial model estimation results on which Bayesian model specification is based are drawn in column (4) of 5, we label variants of the Bayesian model as model (4,1), (4,2), (4,3) and (4,4).

[53]Please refer to appendix 3 for a short description of convergence criteria and appendix 4 for convergence diagnostic of all selected models.

The convergence tests for all regressions show, that convergence of the sampler is guaranteed for all simulations. Therefore, we rely on model $(4, 1)$ because it request fewer draws. The estimation results for this model can be found in 5 in column (3). If we now turn back to 5 and compare the heteroscedastic Bayesian counterpart in column (3) with the homoscedastic Maximum-Likelihood based estimation in column (2) then we can easily see, that estimation results do not differ dramatically. Picture 3 and picture 4 confirm this result. Again, the coefficient of $\ln(I^{+1})$ is positive but not significant. On contrary, the heteroscedastic Bayesian approach estimates a lower value for the spatial lag component λ^1, as the homoscedastic Maximum-Likelihood does. But again, the parameter range for λ^1 is comparable between the two approaches and both coefficient values are highly significant on a 1% significance level.

The last point we have to tackle is to ask, whether the spatial Bayesian estimation provides us with some evidence of spatial heterogeneity. Picture 5 shows a plot of the mean of the v_i draws which should serve as an estimate of these relative variance terms. We can see that one outlier is identified, irrespectively what model we choose. If spatial homogeneity is observed, all elements of Ω should realize the value one. Obviously, this not the case for all four Bayesian models, as we can see from figure 5. From this point of view, we should conclude, that spatial heterogeneity matters, although Maximum-Likelihood and Bayesian estimates correspond each other with respect to parameter estimates. Therefore, we should choose the Bayesian model represented column (3) in table 5 as the optimal one, which delivers efficient parameter estimates.

dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$			
independent variables $\mathbf{x} \in \mathbf{X}$	ML	ML	Bayes
Preferred Model	(4)[SAC(1,1)]	(4)[SEM(1)]	(4)[SEM(1)]
Column	(1)	(2)	(3)
Constant	8.531315	10.09034	10.07207
	(0.0034)$^{\diamond}$	(0.0000)	(0.0000)
ln(K)	0.306738	0.303988	0.294401
	(0.0000)	(0.0000)	(0.0019)
ln(L)	-0.232205	-0.235542	-0.231121
	(0.0004)	(0.0003)	(0.012526)
ln(I)	0.009437	0.011475	0.006971
	(0.6514)	(0.5828)	(0.3904)
ln(H)	0.196265	0.183675	0.187209
	(0.0006)	(0.0003)	(0.0024)
ln(P)	-0.050593	-0.043691	-0.043108
	(0.0847)	(0.1101)	(0.1145)
$\ln(H^{+1})$	-0.055016	-0.071008	-0.004407
	(0.5594)	(0.4149)	(0.4871)
$\ln(P^{+1})$	0.070208	0.083117	0.062044
	(0.0477)	(0.0008)	(0.0164)
$\ln(I^{+1})$	0.028737	0.035589	0.030584
	(0.6262)	(0.5416)	(0.3233)
d	0.261555	0.252277	0.255723
	(0.0000)	(0.0000)	(0.0005)
ρ^1	0.132883	—	—
	(0.5905)	(—)	(—)
λ^1	0.696998	0.710951	0.561134
	(0.0000)	(0.0000)	(0.0081)
Observations	39	39	39
$\ln(\mathcal{L})$	90.47	67.93	—
adjusted pseudo R^2	0.83	0.83	0.81

Table 5: Estimation results for German NUTS-2 regions

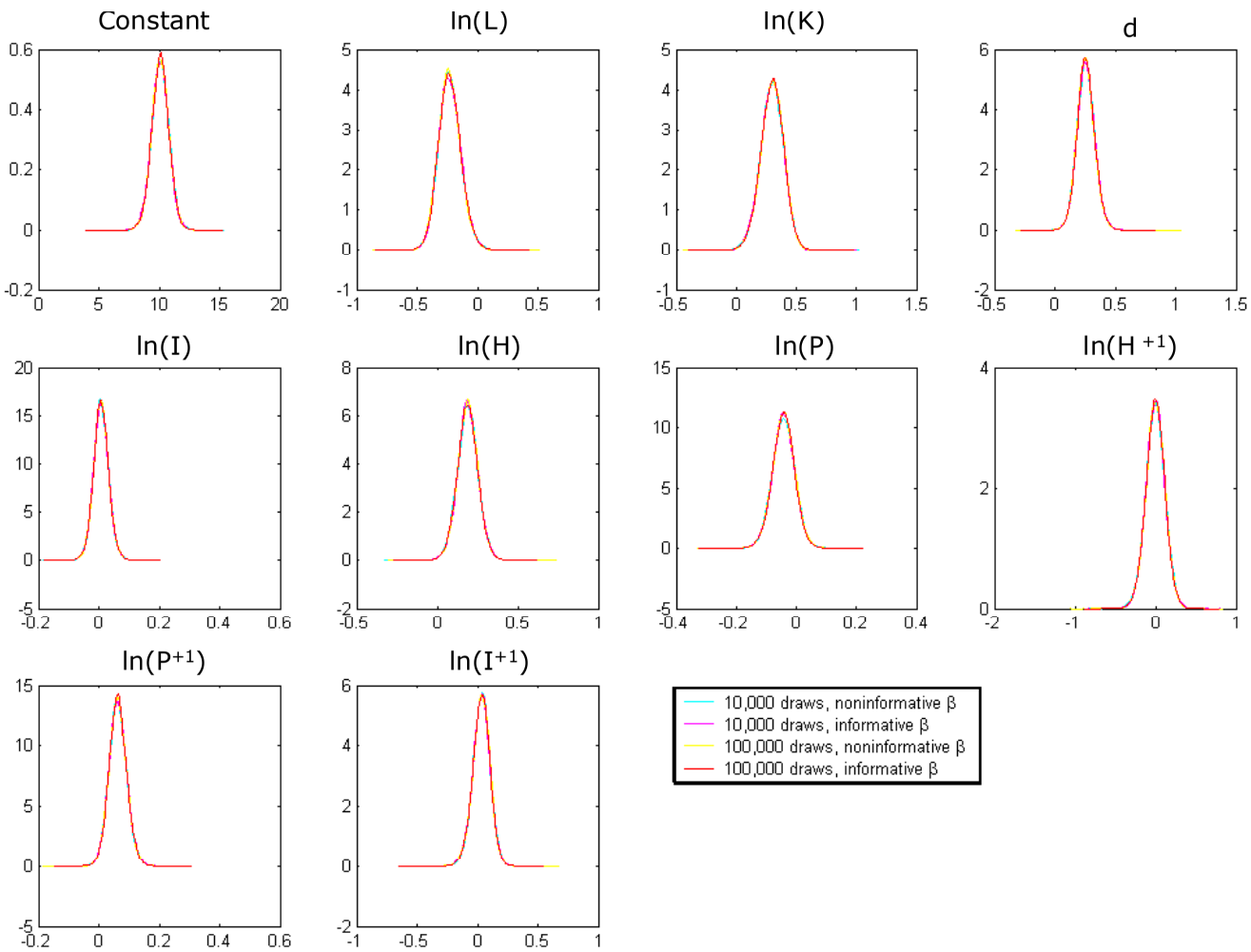

Figure 3: Density plots of estimated β^X and β^{++}

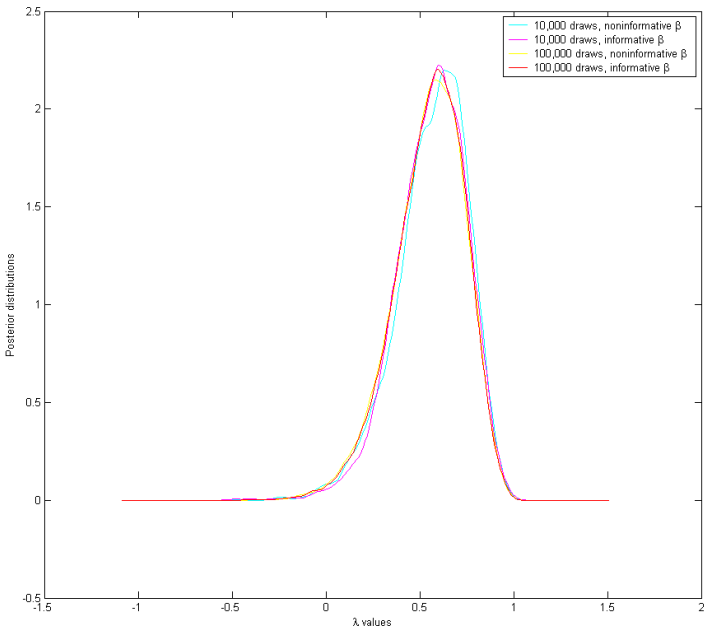

Figure 4: Density plots of estimated λ^1

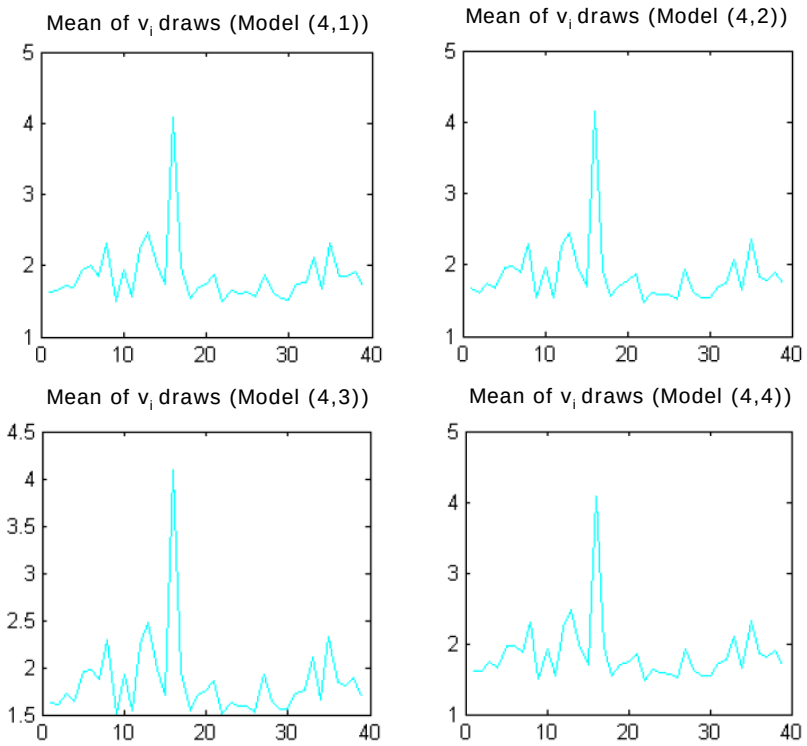

Figure 5: Computation of v_i draws of Ω

10 Interpretation of obtained results

In this section we have tried to find out how regional labour productivity is affected by spatial knowledge processes. We found, that first order neighbouring patent activity influences the regions own labour productivity, while own patent activity does not exhibit a significant influences on own labour productivity. Additionally, most of spatial activity cannot be explained fully by exogenous spatial lagged knowledge. This is the case, because the spatial error term is highly significant, even if one includes spatial lagged counterparts of exogenous variables. Additionally, it was shown with a spatial Bayesian analysis, that spatial heterogeneity is a reasonable assumption and neglecting this issue would lead to inefficient parameter estimates. The next step is to investigate further the impact of knowledge diffusion on German NUTS-2 regions more systematically. So far, we only have obtained some evidence, that the data generating process can be described also by spatial effects. The next

step is, to isolate the spatial neigbbouring influence from the data. In this way it is possible, to distinguish between region specific or home effects and neighbour effects. For instance, regions might have a high labour productivity compared to the average, but this level of labour productivity might be influenced negatively by neighbouring regions et vice versa. The goal is to identify strength and weakness of German NUTS-2 regions and derive implications for regional policy instruments.

11 Spatial filtering

In this section we try to isolate spatial spillover effects from region specific labour productivity. In this way it is possible to create a strength and weakness profile of German Nuts-2 regions. Particularly, one should be interested in answering the question which regions have positive effects on neighbouring regions and which regions provide negative effects on neighbouring regions. This has also implications for an appropriate regional policy. In this way we can say that labour productivity is a sum of own labour productivity and spillovers from neighbouring regions which can be either positive or negative. The question is, if the overall effect is positive or negative. We base the spatial filtering procedure on the so far obtained results. Thus we set $r = 1$ and include only patents p and human capital h as exogenous variables in our filter procedure.

11.1 Concept of the filtering approach

Spatial filtering is a well established analysis method in spatial econometrics applications. The idea is based on a two step estimation technique. In the first step we have to filter every exogenous variable and in the second step we have to regress the dependend endogenous variables on all spatial filtered exogenous variables.
The starting point of spatial filtering is the Morans's $\mathcal{I}$. From equation 12 we know that Moran's $\mathcal{I}$ for a standardized matrix W^{+1} can be computed as follows:

$$\mathcal{I} = \frac{e'W^{+1}e}{e'e}. \tag{46}$$

This equation can be reformulated [54] as

$$\mathcal{I} = \frac{y' C^{+1} y}{e'e}, \tag{47}$$

with

$$C = \left(I - \frac{\iota\iota'}{N} \right) W^{+1} \left(I - \frac{\iota\iota'}{N} \right), \tag{48}$$

with ι as a $(N \times 1)$ vector of ones and I as the $(N \times N)$ identity matrix. In this way, the Eigenvectors of C^r load every spatial effect. The first Eigenvalue contains the largest Morans's $\mathcal{I}$ coefficient with a given standardized matrix W^{+r}. The second Eigenvalue contain the value, which leads to the maximal Morans's $\mathcal{I}$ given the second Eigenvalue is not correlated with the first one, which is ensured, because W^{+r} is standardized.

Because of missing degrees of freedom, one cannot use every Eigenvector for spatial filtering [55]. Therefore a rule of thumb for Eigenvector selection is needed. (Griffith, 2003) has proposed to use only those Eigenvectors which fulfill the following condition:

$$\mathcal{I} > 0.25\, \mathcal{I}_{max}. \tag{49}$$

Equation 49 provides us with an indicator regarding the maximum number of Eigenvectors L which should be included into our regression framework. Based on a top down procedure, one can eliminate all Eigenvectors which do not provide a substantial potential of explanation. Given we have identified the relevant Eigenvectors, we can proceed with the filtering scheme. On the first step we filter the vector of independent variables X by running the following regression:

$$x_k = \gamma_0 + \sum_{l=1}^{L} \gamma_l \hat{v}_l + \epsilon_k, \tag{50}$$

with $\epsilon \sim (0, \sigma^2 \Omega)$, v_l the l^{th} Eigenvector and x_k the k^{th} exogenous variable.

[54] Refer to (Griffith, 2000), p. 145.

[55] Refer to (Griffith, 2003), p. 107.

It is clear that the estimated residual vector $\hat{\epsilon}_k$ contains the spatial filtered counterpart of the not filtered variable x_k. The second step is to regress y on spatial filtered variables and on Eigenvectors v_l. In this equation every variable is spatial filtered and therefore OLS estimation is unbiased. The corresponding regression on the second step can be written as:

$$y = \gamma_0 + \sum_{l=1}^{L} \gamma_l \hat{v}_l + \sum_{k=1}^{K} \gamma_l \hat{\epsilon}_k + \kappa, \tag{51}$$

with $\kappa \sim (0, \sigma^2\Omega)$, v_l the l^{th} Eigenvector and x_k the k^{th} exogenous variable. Of course equation 51 can be consistently estimated with OLS.

11.2 Eigenvector computation

From matrix C of equation 48 one can derive the Eigenvectors and compute Moran's-$\mathcal{I}$ with 12. This can be done using Matlab for instance. For every Eigenvector v_l the corresponding Moran $\mathcal{I}$ coefficient was computed. As one can see from picture 6 only 10 of 39 Eigenvalues meet 49. The second Eigenvector leads to $\mathcal{I}_{max}$ which takes the value $\mathcal{I}_{max} = 0.97437$.

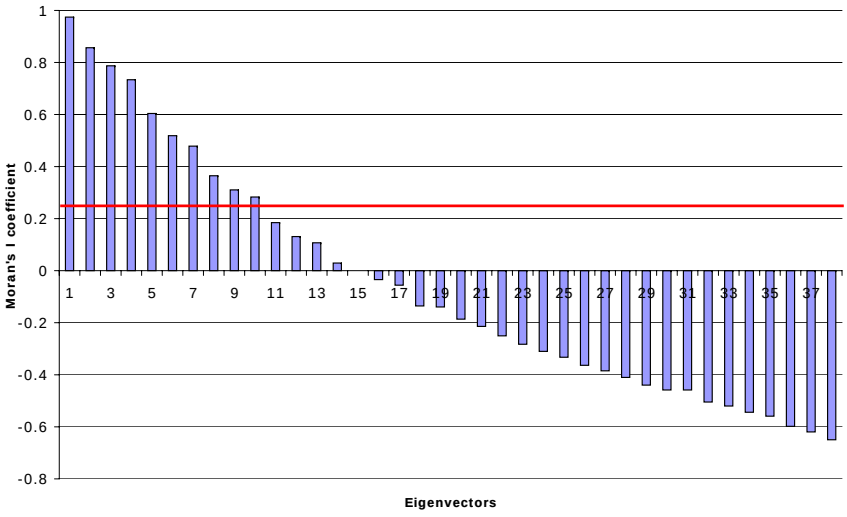

Figure 6: Eigenvector selection

dependend variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$				
estimation method	OLS	OLS	OLS	OLS
dependent variables $\mathbf{x} \in \mathbf{X}$	l	k	h	p
Constant	13.21070	12.30990	2.090592	5.293690
	$(0.0000)^{\diamond}$	(0.0000)	(0.0000)	(0.0000)
v_1	—	—	—	-3.049842
	(—)	(—)	(—)	(0.0019)
v_3	—	—	0.623305	—
	(—)	(—)	(0.0755)	(—)
v_6	-1.259430	-1.229270	—	-1.816913
	(0.0084)	(0.0424)	(—)	(0.0287)
v_7	0.950438	—	—	1.500619
	(0.0176)	(—)	(—)	(0.0961)
Observations	39	39	39	39
adjustedR^2	0.16	0.08	0.07	0.26

Table 6: Spatial filtering of exogenous variables X

11.3 Spatial filtering estimation

First we estimated separately 12 for k, l, h and p.[56] The results of these regressions, corresponding to equation 50 can be found in table 6.

As we can see from figure 7, we cannot observe a clear spatial pattern, represented by the Eigenvectors v_6 and v_7, labeled as (a) and (b). On contrary, in figure 8 the Eigenvector v_1, labeled as (c), the Eigenvector v_2, labeled as (d) and the Eigenvector v_3, labeled as (e) show a clear spatial pattern. The first Eigenvector v_1, labeled as (c) is declining from North to South, the second Eigenvector, the Eigenvector v_2, labeled as (d) exhibit a significant declining West-East pattern, whereas the third Eigenvector v_3, labeled as (e) is affected by low values in North-West and South-East

[56] Based on Wald-tests, we should not include i and p both. As argued above again, a Wald-test based on ML estimation ignoring spatial dependence in the data is not valid. Because of the fact that we want to include possible knowledge spillover variables, we only eliminated i. From table 5 we see, that neither i nor the spatial lagged counterpart of i are significant, whereas the spatial lagged counterpart of p is significant.

of Germany.

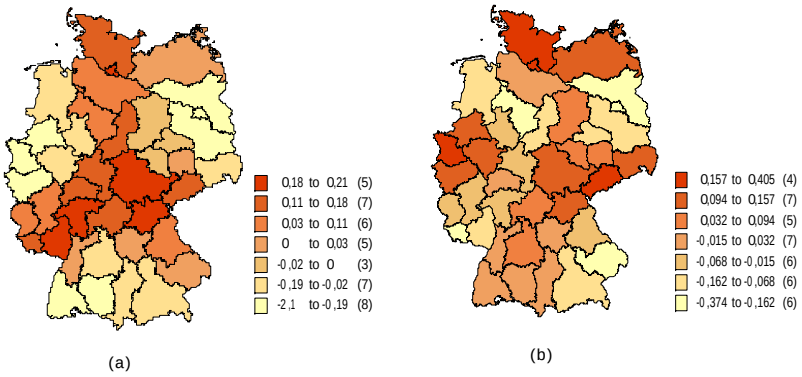

Figure 7: Graphical representation of Eigenvectors (I)

After filtering the exogenous variables, the next step is estimating 51. Therefore, a stepwise estimating procedure of labour productivity on Eigenvectors and spatial filtered variables is employed. In the regression context, no dummy variable for West-East differences is included, because the dummy would filter spatial information potential and could lead to a biased regression in this context. The results of this estimation can be found in table 7. We find, that the first three Eigenvectors, which cover spatial effects, determine a considerable amount of labour productivity. This leads us to conclude, that labour productivity of a given region is not only determined by its own economic potential, but also by neighbouring labour productivity. This implies, that network effects play an important role and should be considered within the embodiment of regional policy. We can therefore conclude, that patents p and human capital h are mainly affected by spatial effects. The latter is only partial true for capital k and labour l.

As we can see from 7, the results for the constant estimated labour elasticity l, the constant estimated capital elasticity k, the constant estimated human capital elasticity h and the constant estimated patent elasticity p have all positive signs and have been, with respect to their dimension correct estimated. With the exception of p, all estimated coefficients are significant on a 5% or 10% level.

Now we are prepared to decompose labour productivity in home effects and neigh-

dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$	
estimation method	OLS
Constant	11.21121
	$(0.0000)^{\diamond}$
$\hat{\epsilon}_l$	-0.358058
	(0.0031)
$\hat{\epsilon}_k$	0.261464
	(0.0183)
$\hat{\epsilon}_h$	0.136156
	(0.0747)
$\hat{\epsilon}_p$	0.075354
	(0.1114)
v_1	-0.421669
	(0.0002)
v_2	0.211905
	(0.0323)
v_3	-0.305651
	(0.0225)
Observations	39
adjustedR^2	0.66

Table 7: Spatial filtering of labour productivity y

bour effects. The residual of this simple decomposition cannot be returned neither to home effects, nor to neighbour effects and therefore they are treated as not systematic. Noting the fact, that both, Eigenvectors and spatial filtered variables exhibit a mean of zero, we can conclude that the constant term contains the mean of labour productivity. If we subtract the mean $\bar{y}$ from equation 51 we obtain:

$$\breve{y} \equiv y - \bar{y} = \gamma_0 + \sum_{l=1}^{L} \gamma_l v_l + \sum_{k=1}^{K} \gamma_l \hat{\epsilon}_k - \bar{y} + \kappa, \tag{52}$$

with $\kappa \sim (0, \sigma^2\Omega)$, v_l the l-th Eigenvector and x_k the $k-th$ exogenous variable. The term $\gamma_0 + \sum_{k=1}^{K} \gamma_l \hat{\epsilon}_k - \bar{y}$ can be defined as the own region effect, whereas the term $\sum_{l=1}^{L} \gamma_l v_l$ represents the neighbour effects. The term κ represents the unsystematic component. Because of the fact, that all effects are centered around zero, we can interpret equation 52 as a deviation from the mean specification. For instance, if $\breve{y} > 0$, which means that a region exhibits a superior labour productivity, this can be due to home effects or due to neighbouring effects. Even if a region has a superior home effect, a negative neighbour effect could lead to a negative overall

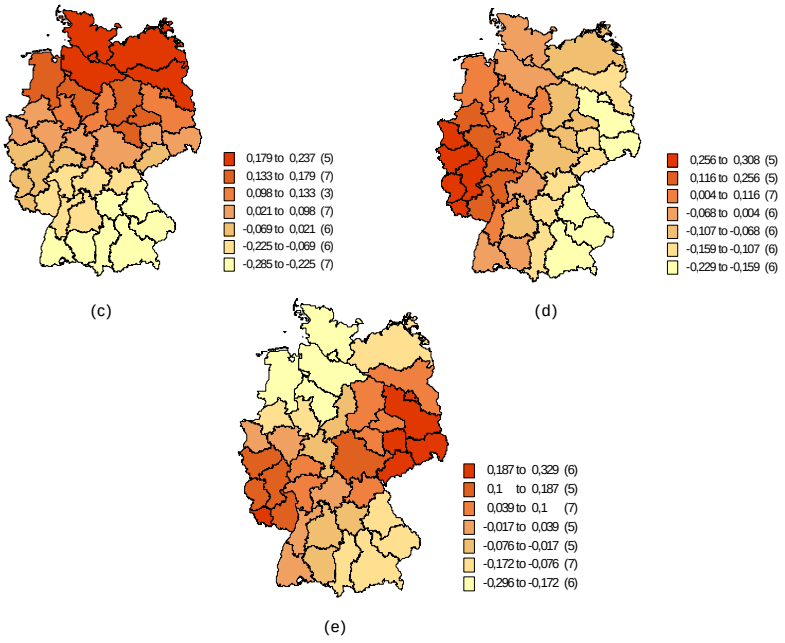

Figure 8: Graphical representation of Eigenvectors (II)

effect regarding labour productivity, et vice versa.[57]

11.4 Interpretation of simulation results

The next two figures in 9give an impression of the results of labour productivity simulation, based on equation 52. First, we should investigate own regions effects regarding labour productivity, which are separated from regions neighbour effects. The labour productivity effects are deviations from the mean which is, as mentioned above, centered around zero. As we can see with respect to the own region effect (RE) in the left hand map, especially some East German regions, such as "Süd Brandenburg", "Sachsen" and "Berlin" would exhibit a relative high labour productivity if we only apply for own region effects. But with respect to the overall effect (OE), which is plotted in the right hand map, some negative influence from neighbouring

[57]From equation 52 we see, that an inclusion of a dummy variable as done before, would bias within the regression context. Even more, a spatial filtering of a dummy is by definition not plausible. Besides that and give, we include the dummy we cannot rule out that these variable also contains spatial information.

regions leads to a reduction of labour productivity in those regions. On the other side, the region "Oberfranken" benefits for the most part from neighbouring effects.

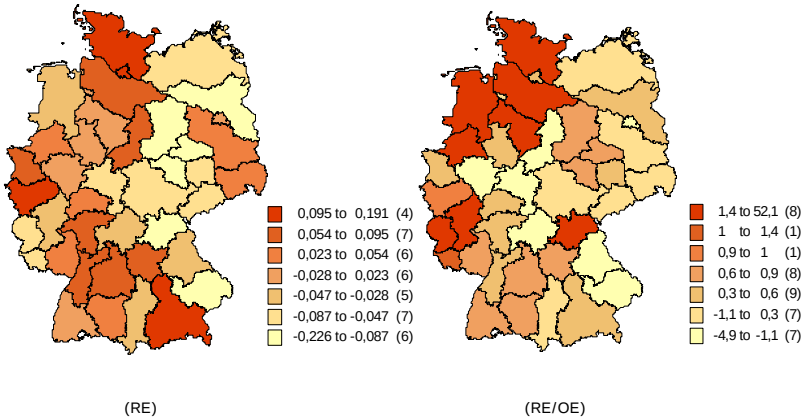

Figure 9: Absolute and relative regional effects

Now we turn our attention to the neighbouring effects. These are visualized in figure 10. First, we find in the map on the left hand side a rather impressive confirmation that especially South German regions and with some cut backs also West German regions, settled in the "Rhein-Main -Gebiet" and the "Ruhrgebiet", are the source of knowledge spillovers. On contrary, we find maximum negative neighbour influence throughout East German regions. With respect to the overall effect (OE), we find in the map on the right hand side some dramatic changes. The effects for South German regions, some regions of the "Rhein-Main-Area" and some regions of the "Ruhrgebiet" are rather low. Thus, only a little fraction of the superior labour productivity of these regions are due to neighbouring effects.

Finally, we can categorize the regions in a strength-weakness profile, both for the own region and neighbouring region effects.

Regions which are settled in the top right corner of figure 11 can be characterized by a superior labour productivity. For those regions, positive or negative neighbouring effects play only a minor role. In these regions, with the exception of "Hamburg", which is top leader with respect to own and overall effects and "Schleswig-Holstein", you can find mainly South German regions, such as "Oberbayern", "Stuttgart",

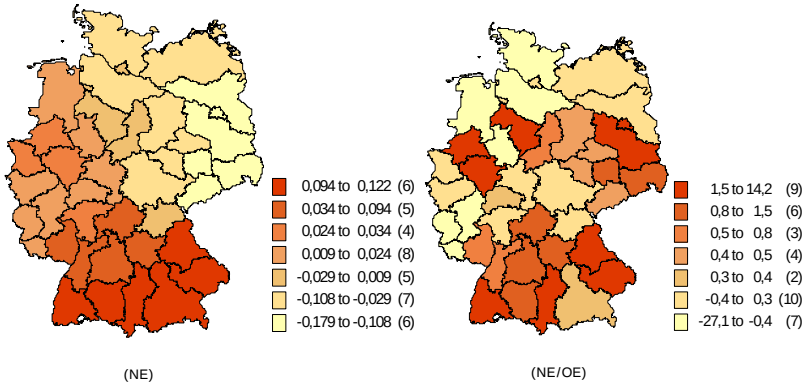

Figure 10: Absolute and relative neighbour effects

"Tübingen" etc. and West German regions, such as "Düsseldorf", "Köln" etc. situated in this area. These findings supports the findings of (Eckey et al., 2007) for German labour market regions.

Regions which can be found in the down right corner of figure 11 exhibit a positive over all effect, because of the positive neighbouring effects. But, without these effects, a negative overall effect would occur. In this area you can find mainly West German regions, which profit from spillover regions, situated in the upper right regions. This is especially true for some Bavarian regions, such as "Schwaben", "Oberpfalz", "Niederbayern" who profit mainly from "Oberbayern" spillover centers like "Greater Munich area".

Regions in top left exhibit a negative overall effect, despite the fact, that the home effect is positive. In other words, if negative neighbouring effects did not affect those regions, those regions could be associated with a superior labour productivity. In this region you find primarily East German regions, which are compared to other East German regions are relative prosperous with respect to their economic development. This is especially the case for "Dresden", "Süd-Brandenburg". But also "Berlin" and "Braunschweig" can be found in this area.

Regions in the down left regions can be characterized as regions which require economical and political support and should therefore be in the focus of political debate when talking about the allocation of supranational grants. Neither their own labour

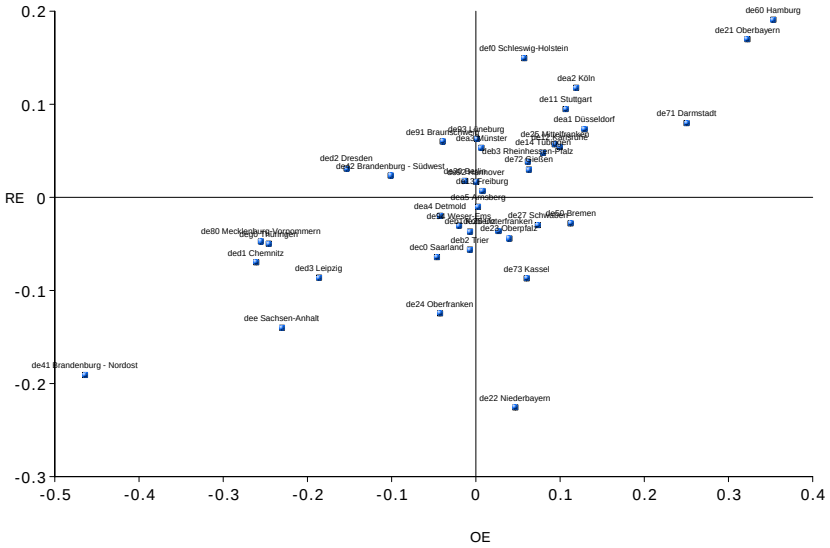

Figure 11: Regional and overall effect

productivity is superior, nor they can benefit from positive knowledge spillovers from neighbouring regions. Again, in this region we can find by majority East German regions. But, in contrast to East German regions which are settled in the upper left, those regions suffer from structural weaknesses. This is especially the case for "Nord-Brandenburg" but also for West German regions, such as "Saarland" and "Oberfranken". In this picture we can see that when talking about economic performance of German regions, the entity "Bundesländer" is to crude. For instance, it is not as easy as it seems ex ante to get a correct impression of the economic performance of "Sachsen". For NUTS-2 case Dresden performs rather well, whereas "Chemnitz" and "Leipzig" perform bad. This is also true for West German regions, especially "Bayern", which seems to be more heterogeneous than "Baden-Württenberg".

Figure 12, which has at the ordinate the neighbour effects and at the abscissa the overall effect has to be interpreted analogously as figure 11 and provides an alternative view on the same results obtained before. Again, some regions would exhibit a positive overall effect, unless negative neighbour effects are taken into consideration. Again, this is especially true for "Berlin", "Brandenburg-Südwest" and "Dresden" for instance.

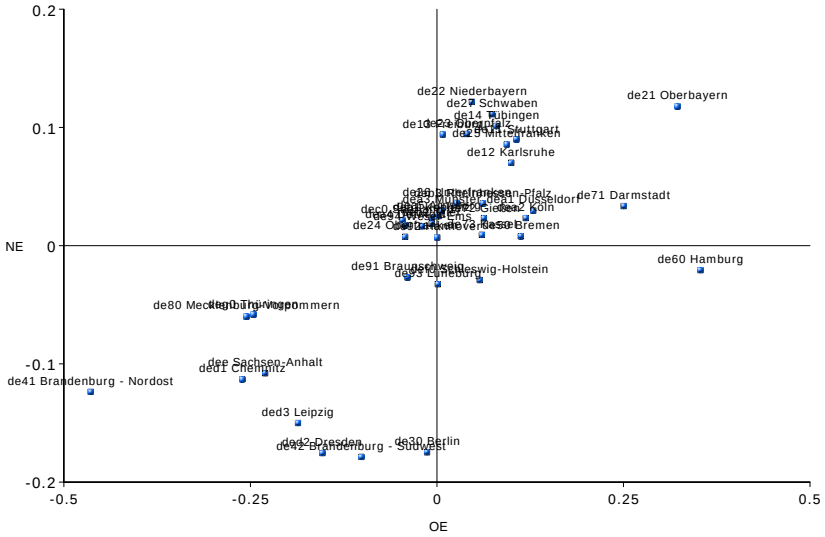

Figure 12: Neighbour and overall effect

12 Policy implications

The purpose of this section is to derive some implications for regional policy, based on simulation results which have been obtained in the last section. The core of regional policy since 1969, coordinated by German administration is the so called "Bund-Länder-Gemeinschaftsaufgabe "Verbesserung der regionalen Wirtschaftsstruktur"" (GA). Regional policy in the German sense is a cooperation of the German countries and the Federal Republic of Germany which is controlled by Art. 91.a in the German constitution. But (GA) is not only a traditional funding instrument. (GA) is the framework of strategy, regulation and coordination of regional policy, also for EU related founds.

If we have again a look at figure 11, and more precisely, if we again take the upper left region. Then from a policy maker view it should be clear that policy instruments are required which take into account that a comprehensive regional approach goes into the wrong direction. For those regions a mixture of traditional structural sponsorship should be supplemented by appropriate public-private-partnerships. Especially for East German regions regional policy has focused instruments which should hinder the Brain drain towards West German regions. Since 2005 the German economic

department programme "Kooperationsnetzwerke und Clustermanagement" supports regional and national cooperation between companies, scientific and economy as well as between local administration to strength the network abilities and competitiveness of regions. This seems an appropriate policy instrument for those mentioned regions. For regions in the down left area in figure 11, traditional regional policy arrangements seem to be appropriated. Additionally, as mentioned by (Moll, 2000) EU region wide cooperations between country such as Germany and France near the German-French boarder or with Czech Republic to promote former so called borderlands or with Poland to promote close to boarder regions of "Mecklenburg" have been aspired. Thus, EU as German funding instruments includes a prominent regional component which aims to support the creation of regional clusters.

13 Conclusion

The main intention of this paper is to identify knowledge spillovers in a spatial context. Particularly, it is assumed, that spatial effects per se are heterogeneous, an assumption which seems to be plausible. Hence, this paper combines spatial heterogeneity and spatial dependence which are the two main aspects of spatial econometrics.

These aspects have been integrated in a cross section production function approach, proposed by (Griliches, 1979), which should measure the effects of innovativeness, measured by knowledge capital, such as human capital, patents or $R\&D$ and spatial spillovers on output. This is done for German NUTS-2 regions. These administration level has been selected due to the fact, that referring on NUTS-3 regions could lead to spatial dependence by "construction" caused by streams of commuters for instance. Spatial econometricians methods have been employed to measure the before mentioned effects. Spatial heterogeneity is mostly neglected in hitherto empirical studies. Thus, employing a new model selection mechanism, which accounts for spatial heterogeneity and which is based both on Maximum-Likelihood and Bayesian methods, one can find that significant spatial knowledge spillovers exist in the data, even though they are small. Especially, patents spillovers have been detected as the driving forces of economic performance. Further, the selected model found that spatial heterogeneity matters. Controlling for spatial heterogeneity is important because neglecting it could lead to insufficient estimates. Until today, the majority of existing studies assume ex ante spatial homogeneity. This could be due to the fact, that Maximum-Likelihood methods are very clumsy for spatial model estimation. Coevally, Bayesian methods are still on the fringes, especially in spatial econometrics, although the conceptual idea of Bayesian methods are more eligible to cover spatial model design than Maximum- Likelihood methods so far. Hence, it can be expected that in the next years some improvements of Maximum-Likelihood methods will be made in terms of efficient spatial model estimation.

Another way to investigate spatial data, is to employ spatial filter methods. This method should be used, if spatial effects should be removed from data. In this

context it is obvious to ask the question if regions specific economic strength benefits from economic activity from their neighbours or not. The filtering method is easy to implement and can be conducted with a traditional two step OLS procedure. One of the key findings is, that economic performance differs not primarily between East and West German regions, but is more complex. Especially for East German regions we find, that some well performing regions suffers in great extent from negative neighbour influence. This is also true for some West German regions but plays a minor role. Against this background it is rather logical, that cluster phenomena are suitable for explaining the distribution of economic activity of German NUTS-2 regions over space.

This cluster phenomena can be graphically replicated with a weakness-strength profile. To obtain this, on the basis of the employed filter method a simulation of labour productivity has been conducted. Using the simulated data it is found that especially, South German regions, such as "Bayern" and "Baden-Württemberg" and regions in the "Ruhr-Gebiet" perform well, due to their inherent economic strength. These regions do not rely on positive neighbour effects to beef up their economic performance. Therefore, these regions can be labeled as knowledge generation areas. On contrary, some regions would perform significantly better, if negative spillover from neighbouring regions could be eliminated. This is particularly true for "Brandenburg-Süd" and "Dresden".

What are the political implications? As mentioned above, EU has launched several economic policy programmes to foster regional economic performance. Most of the EU related programmes have recognized the outstanding role of knowledge and economic clusters for regional development. Knowledge spillovers, generated by knowledge generation areas, such as "Munich Greater Area"' should contribute to boost neighbouring regions, which suffer from insufficient knowledge generating potential. Hence, regional politics is on the right track, but should provide further incentives for strengthen regional knowledge networks.

14 Appendix

14.1 Appendix 1

The log-likelihood function for the spatial variant of a ARMA(1,1), SAC(1,1) can be derived as follows.[58]

$$y = \rho^1 W^{+1} y + X^{+++}\beta^{+++} + \epsilon, \tag{53}$$

with

$$\epsilon = \lambda^1 W^{+1}\epsilon + \kappa, \tag{54}$$

with $\kappa \sim (0, \sigma^2 I)$ can be written as

$$\xi = \frac{1}{\sigma}(I - \lambda^1 W^{+1})[(I - \rho^1 W^{+1})y - X^{+++}\beta^{+++}] \tag{55}$$

with $\xi \sim N(0, I)$. The corresponding determinant of the Jacobian $\mathcal{J} \equiv det\frac{\partial \xi}{\partial y}$ can be rewritten as

$$\mathcal{J} \equiv det\frac{\partial \xi}{\partial y} = |\frac{1}{\sigma}[I - \lambda^1 W^{+1}]\,||\,[I - \rho^1 W^{+1}]|. \tag{56}$$

Employing the fact that $\xi \sim N(0, I)$ we can write the log-likelihood for the joint distribution as

$$\ln \mathcal{L} = -\frac{N}{2}\ln 2\pi - \frac{N}{2}(\sigma^2) + \ln|\tilde{N}| + \ln|[I - \rho^1 W^{+1}]| - \frac{1}{2}\xi'\xi, \tag{57}$$

If $\rho^1 = 0$ then the log-likelihood 25 results.

[58] The proof is based on (Anselin, 1988), p. 74 with some minor adjustments.

14.2 Appendix 2

Dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$				
Independent variables $\mathbf{x} \in \mathbf{X}$				
Preferred model	Number of parameters	$\ln(\mathcal{L})$	$\hat{\rho}$	$\hat{\lambda}$
Model 4 [SAC(1,1)]	12	90.47	0.133	0.697$^{\natural}$
Model 4 [SAC(1,2)]	12	88.55	0.284	-0.989
Model 4 [SAC(2,1)]	12	90.28	-0.000	0.722$^{\natural}$
Model 4 [SAC(2,2)]	12	87.74	-0.000	-0.987
Model 4 [SEM(1)]	11	67.93	—	0.711$^{\natural}$

$^{\diamond}$ Selected model. † indicates 10% significance. ‡ indicates 5% significance. $^{\natural}$ indicates 1% significance

Table 8: Comparison of selected models

14.3 Appendix 3

In the relevant literature, there are some convergence checks for convergence of MCMC based samplers for linear models. In this section there is given a short motivation of some convergence checks instruments. All below mentioned diagnostic tools are implemented in the Matlab function "coda".

14.3.1 Autocorrelation estimates

From time series it is known that if ρ is a stationary correlated process, then $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\rho_i$ is a consistent estimate of $E(\rho)$. Therefore it is allowed to simulate some correlated draws from our posterior distribution to get a hint how many draws we need for uncorrelated draws for our Gibbs sampler. A high degree of correlation should cause someone to carry out more draws which should result in a sample which allows to draw correct posterior estimates.

14.3.2 Raftery-Lewis diagnostics

(Raftery and Lewis, 1992b), (Raftery and Lewis, 1992a) and (Raftery and Lewis, 1995) have suggested a set of diagnostic tools which they have first implemented in `FORTRAN` named "Gibbsit". This function was converted in `Matlab` and called "raftery". (Raftery and Lewis, 1992b), (Raftery and Lewis, 1992a) and (Raftery and Lewis, 1995) have focused on the quantiles of the marginal posterior. The diagnostic itself is based on the properties of a two state Markov-Chain, because for a given quantile the chain is dichotomized using a binary time series that is unity, if $\rho_i \leq q_{quant}$ and zero otherwise, where q_{quant} denotes the quantile which has to be chosen from the researcher ex ante. For an independent chain, the zeros and ones should be appear randomly. The "coda" function prints the so called thinning-ratio, which is an indicator of autocorrelation in the draws. "Thinning" means, that only every third, fifth,... draw for instance are saved for inference, because the draws from a Markov Chain are not independent. Additionally, the number "burn-in-draws" are reported. The number of "burn-in-draws" are excluded from sampling based on inference. Finally, the I-statistic is reported which is the ratio of the number of total draws and the minimum number of draws to ensure an i.i.d. chain, represented by the draws. (Raftery and Lewis, 1992b), (Raftery and Lewis, 1992a) and (Raftery and Lewis, 1995) indicate that values larger than 5 exhibit convergence problems of the sampler and therefore, more draws should be carried out.

14.3.3 Geweke diagnostics

The `Matlab` function "coda" additionally estimates the numerical standard errors and relative numerical standard errors based on the work of (Geweke, 1992). The code can be found at http://www.biz.uiowa.edu/cbes/code.htm, which is based on BACC. The BACC code itself as `Matlab`, `R` and `S-Plus` routines can be found at http://www2.cirano.qc.ca/ bacc/bacc2003/index.html. This diagnostics are based on elements of spectral analysis. From time series analysis we know, that an estimate of variance of ρ is based on $Var[\hat{\rho}_i] = \frac{\Delta(0)}{k}$ with Δ_0 as the spectral density of ρ_i

evaluated at ω_0 of $\Delta(\omega)$. The question is, how to approximate $\Delta(\omega)$. For this reason, alternative tapering of the spectral window should be used. Using numerical standard errors and relative numerical i.i.d. standard errors and compare them with numerical standard errors and relative numerical standard errors from the tapered version. If the relative numerical standard error of the tapered version is close to one, then convergence seems to be ensured.

14.3.4 Geweke-χ^2 test

Geweke's-χ^2 test is based on the intuition that sufficiently large draws have been taken, estimation based on the draws should rather identical, provided the Markov chain has reached an equilibrium state. This test is a simple comparison of the means for each split of the draws. In this work, the χ^2 test, based on the null hypothesis of equality of the means of splits is carried out for each tapered case.
It should be mentioned that the diagnostic tools introduced here are not foolproof and sometimes MCMC diagnostic tools lead to misleading decisions.[59]

14.4 Appendix 4

For appendix 6, please refer to the following pages.

[59]Refer for this topic to (Koop, 2003), p. 66.

Dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$						
1. AWEPC						
Variable	Lag 1	Lag 5	Lag 10	Lag 50	—	—
λ^1	0.800	0.414	0.205	-0.035	—	—
σ	0.410	0.022	-0.029	0.009	—	—
2. RLDEPC						
Variable	Thin	Burn in	N	N_{min}	I-statistic	—
λ^1	1	18	4736	937	5.054	—
σ	1	18	4736	937	5.054	—
3. GDEPC						
Variable	Mean	Std. deviation	NSE (iid)	RNE (iid)	—	—
λ^1	0.561134	0.187292	0.001922	1.000000	—	—
σ	0.003283	0.001214	0.000012	1.000000	—	—
Variable	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%
λ^1	0.007401	0.067106	0.007782	0.060969	0.008809	0.047582
σ	0.000014	0.767126	0.000014	0.830173	0.000013	0.904161
4. GCSTEPC						
λ^1	NSE	Mean	p-Value	—	—	—
i.i.d.	0.569899	0.002264	0.000000	—	—	—
4% taper	0.565011	0.008253	0.057527	—	—	—
8% taper	0.562932	0.008323	0.055775	—	—	—
15%taper	0.565765	0.008427	0.065577	—	—	—
σ	NSE	Mean	p-Value	—	—	—
i.i.d.	0.003270	0.000015	0.378822	—	—	—
4% taper	0.003267	0.000021	0.592235	—	—	—
8% taper	0.003267	0.000018	0.536563	—	—	—
15%taper	0.003266	0.000017	0.530511	—	—	—

$^{\diamond}$ Note: "AWEPC" stands for "Autocorrelation within each parameter chain". "RLDEPC" stands for "Raftery-Lewis Diagnostics for each parameter chain". "GDEPC" stands for "Geweke Diagnostics for each parameter chain". "GCSTEPC" stands for "Geweke-χ^2-Test for each parameter chain". "RNE" stands for "Relative Numerical Efficiency", "NSE" stands for "Numerical Standard Error".

Table 9: MCMC-convergence summary for model (4,1)

Dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$						
1. AWEPC						
Variable	Lag 1	Lag 5	Lag 10	Lag 50	—	—
λ^1	0.734	0.245	0.076	0.039	—	—
σ	0.457	0.034	-0.002	0.035	—	—
2. RLDEPC						
Variable	Thin	Burn in	N	N_{min}	I-statistic	—
λ^1	1	19	5047	937	1.000	5.386
σ	1	19	5047	937	1.000	5.386
3. GDEPC						
Variable	Mean	Std. deviation	NSE (iid)	RNE (iid)	—	—
λ^1	0.555595	0.183009	0.001878	1.000000	—	—
σ	0.003266	0.001241	0.000013	1.000000	—	—
Variable	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%
λ^1	0.006591	0.81155	0.006416	0.085646	0.005519	0.115728
σ	0.000018	0.528629	0.000015	0.682827	0.000014	0.856455
4. GCSTEPC						
λ^1	NSE	Mean	p-Value	—	—	—
i.i.d.	0.559135	0.002221	0.000089	—	—	—
4% taper	0.560335	0.006356	0.152210	—	—	—
8% taper	0.561610	0.006683	0.162430	—	—	—
15%taper	0.560764	0.006756	0.173295	—	—	—
σ	NSE	Mean	p-Value	—	—	—
i.i.d.	0.003255	0.000015	0.380291	—	—	—
4% taper	0.003257	0.000022	0.582485	—	—	—
8% taper	0.003259	0.000019	0.572221	—	—	—
15%taper	0.003259	0.000018	0.547883	—	—	—

$^{\diamond}$ Note: "AWEPC" stands for "Autocorrelation within each parameter chain". "RLDEPC" stands for "Raftery-Lewis Diagnostics for each parameter chain". "GDEPC" stands for "Geweke Diagnostics for each parameter chain". "GCSTEPC" stands for "Geweke-χ^2-Test for each parameter chain". "RNE" stands for "Relative Numerical Efficiency", "NSE" stands for "Numerical Standard Error".

Table 10: MCMC-convergence summary for model (4,2)

Dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$						
1. AWEPC						
Variable	Lag 1	Lag 5	Lag 10	Lag 50	—	—
λ^1	0.000	-0.003	0.015	-0.005	—	—
σ	0.146	-0.006	0.000	-0.002	—	—
2. RLDEPC						
Variable	Thin	Burn in	N	N_{min}	I-statistic	—
λ^1	1	2	974	937	1.039	—
σ	1	2	974	937	1.039	—
3. GDEPC						
Variable	Mean	Std. deviation	NSE (iid)	RNE (iid)	—	—
λ^1	0.156684	0.193235	0.001983	1.000000	—	—
σ	0.008386	0.002139	0.000022	1.000000	—	—
Variable	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%
λ^1	0.002051	0.934374	0.001926	1.059081	0.001864	1.131157
σ	0.000024	0.864318	0.000023	0.924805	0.000023	0.941651
4. GCSTEPC						
λ^1	NSE	Mean	p-Value	—	—	—
i.i.d.	0.159254	0.002375	0.694409	—	—	—
4% taper	0.159402	0.002474	0.689738	—	—	—
8% taper	0.159500	0.002392	0.672743	—	—	—
15%taper	0.159553	0.002236	0.648753	—	—	—
σ	NSE	Mean	p-Value	—	—	—
i.i.d.	0.008387	0.000026	0.285231	—	—	—
4% taper	0.008380	0.000031	0.325237	—	—	—
8% taper	0.008381	0.000031	0.329946	—	—	—
15%taper	0.008380	0.000030	0.303602	—	—	—

$^{\diamond}$ Note: "AWEPC" stands for "Autocorrelation within each parameter chain". "RLDEPC" stands for "Raftery-Lewis Diagnostics for each parameter chain". "GDEPC" stands for "Geweke Diagnostics for each parameter chain". "GCSTEPC" stands for "Geweke-χ^2-Test for each parameter chain". "RNE" stands for "Relative Numerical Efficiency", "NSE" stands for "Numerical Standard Error".

Table 11: MCMC-convergence summary for model (4,3)

Dependent variable $\mathbf{y}$: $\ln\left(\frac{Y}{L}\right)$						
1. AWEPC						
Variable	Lag 1	Lag 5	Lag 10	Lag 50	—	—
λ^1	0.677	0.160	0.025	-0.003	—	—
σ	0.440	0.032	0.004	-0.006	—	—
2. RLDEPC						
Variable	Thin	Burn in	N	N_{min}	I-statistic	—
λ^1	1	15	4023	937	4.293	—
σ	2	15	4023	937	4.293	—
3. GDEPC						
Variable	Mean	Std. deviation	NSE (iid)	RNE (iid)	—	—
λ^1	0.541436	0.189626	0.000601	1.000000	—	—
σ	0.003268	0.001236	0.000004	1.000000	—	—
Variable	NSE 4%	RNE 4%	NSE 8%	RNE 8%	NSE 15%	RNE 15%
λ^1	0.001422	0.178616	0.001433	0.176005	0.001312	0.210008
σ	0.000007	0.296403	0.000007	0.305018	0.000007	0.285996
4. GCSTEPC						
λ^1	NSE	Mean	p-Value	—	—	—
i.i.d.	0.541633	0.000722	0.013795	—	—	—
4% taper	0.541560	0.001743	0.299201	—	—	—
8% taper	0.541598	0.001789	0.316022	—	—	—
15%taper	0.541719	0.001697	0.306570	—	—	—
σ	NSE	Mean	p-Value	—	—	—
i.i.d.	0.003267	0.000005	0.018101	—	—	—
4% taper	0.003269	0.000009	0.173843	—	—	—
8% taper	0.003273	0.000008	0.135064	—	—	—
15%taper	0.003275	0.000007	0.077024	—	—	—

$^{\diamond}$ Note: "AWEPC" stands for "Autocorrelation within each parameter chain". "RLDEPC" stands for "Raftery-Lewis Diagnostics for each parameter chain". "GDEPC" stands for "Geweke Diagnostics for each parameter chain". "GCSTEPC" stands for "Geweke-χ^2-Test for each parameter chain". "RNE" stands for "Relative Numerical Efficiency", "NSE" stands for "Numerical Standard Error".

Table 12: MCMC-convergence summary for model (4,4)

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Dordrecht.

Anselin, L. (1992). *SpaceStat tutorial: A workbook for using SpaceStat in the analysis of spatial data.* Morganton, West Virginia: Regional Research Institute, West Virginia University.

Anselin, L. (2005). *Exploring Spatial Data with GeoDa: A Workbook.* Can be downloaded from www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf.

Anselin, L., Bera, A., Florax, R. J. G. M., and Yoon, M. (1996). *Simple diagnostic tests for spatial dependence.* In: Regional Science and Urban Economics, 26, pp. 77-104.

Anselin, L. and Florax, R. J. G. M. (1995a). *New Directions in Spatial Econometrics: Introduction.* In: Anselin, L., Florax, R. J. G. M. (eds.). Advances in Spatial Science, pp. 3-16, Berlin.

Anselin, L. and Florax, R. J. G. M. (1995b). *Small sample properties of tests for spatial dependence in regression models: Some further results.* In: Anselin, L., Florax, R. J. G. M. (eds.). New Directions in Spatial Econometrics, pp. 21-74, Berlin.

Anselin, L. and Rey, S. (1991). *The performance of tests for spatial dependence in linear regression models.* In: Geographical Analysis, 23, pp. 110-131.

Anselin, L., Varga, A., and Acs, Z. (1997). *Local Geographic Spillovers between University Research and High-Technology Innovations.* In: Journal of Urban Economics, 42, pp. 422-448.

Audretsch, D. and Mahmood, T. (1994). *The Knowledge Production Function and R&D Spillovers.* Wissenschaftszentrum Berlin für Sozialforschung. Discussion Paper FS IV 94-6.

Baldwin, R. E. and Martin, P. (2003). *Agglomeration and Regional Growth.* CEPR Discussion Paper No. 3960.

Barro, R. J. and Sala-I-Martin, X. (1997). *Technological diffusion, convergence, and growth.* In: Journal of Economic Growth, 2, pp. 126.

Barry, R. P. and Kelley, R. P. (1999). *Monte Carlo Estimates of the Log Determinant of Large Sparse Matrices.* In: Linear Algebra and its Applications, 289, pp. 41-54.

Bayes, T. (1763). *An Essay towards solving a Problem in the Doctrine of Chances.* In: Philosophical Transactions of the Royal Society of London, 53, pp. 370-418.

Bottazzi, L. and Peri, G. (2003). *Innovation and spillovers in regions: evidence from European patent data.* In: European Economic Review, 47, pp. 687-710.

Coe, D. T. and Helpman, E. (1995). *International R&D Spillovers.* In: European Economic Review, 39, pp. 859-887.

Dollar, D. (1986). *Technological innovation, capital mobility, and the product cycle in North-South trade.* In: The American Economic Review, 76, pp. 177-190.

Eckey, H. F., Kosfeld, R., and Türck, M. (2007). *Regionale Entwicklung mit und ohne räumliche Spillover-Effekte.* In: Jahrbuch für Regionalwissenschaft, 27, pp. 23-42.

Fernandez, C., Ley, E., and Steel, M. (2001a). *Benchmark priors for Bayesian model averaging.* In: Journal of Econometrics, 100, pp. 381-427.

Fernandez, C., Ley, E., and Steel, M. (2001b). *Model uncertainty in cross-country growth regressions.* In: Journal of Applied Econometrics, 16, pp. 563-576.

Florax, R. J. G. M., Folmer, H., and Rey, S. J. (2003). *Specification searches in spatial econometrics: the relevance of Hendry's methodology.* In: Regional Science and Urban Economics, 33, pp. 557-579.

Frauenhofer (2000). *Regionale Verteilung von Innovations- und Technologiepotentialen in Deutschland und Europa.* Fraunhofer-Institut für Systemtechnik und Innovationsforschung, Karlsruhe.

Fujita, M. and Thisse, J. (2002). *Economics of Agglomeration.* Cambridge.

Gelfand, A. E., Racine-Poon, A., and Smith, A. F. M. (1990). *Illustration of Bayesian inference in normal data models using Gibbs sampling*. In: Journal of the American Statistical Association, 85, pp. 972-985.

Gelman, A., Carlin, B., J., Stern, H. S., and Rubin, D. (1995). *Bayesian Data Analaysis*. London.

Geweke, J. F. (1992). *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments*. In: Berger, J. O., Bernardo, J. M. , Dawid, A. P., Smith, A. F .M. (eds.). Bayesian Statistics, 4, pp. 169-194.

Geweke, J. F. (1993). *Bayesian treatment of the independent Student-t linear model*. In: Journal of Applied Econometrics, 8, pp. 19-40.

Glass, A. (1997). *Product cycles and market penetration*. In: International Economic Review, 38, pp. 865-891.

Greif, S. (1998). *Patentatlas Deutschland*. München.

Griffith, D. A. (1996). *Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data*. In: Canadian Geographer, 40, pp. 351-367.

Griffith, D. A. (2000). *A Linear Regression Solution to the Spatial Autocorrelation Problem*. In: Journal of Geographical Systems, 2, pp. 141-156.

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

Griliches, Z. (1979). *Issues in Assessing the Contribution of Research and Development to Productivity Growth*. In: The Bell Journal of Economics, 10, pp. 92-116.

Grossman, G. M. and Helpman, E. (1991a). *Innovation and Growth in the Global Economy*. Cambridge.

Grossman, G. M. and Helpman, E. (1991b). *Quality ladders and product cycles*. In: The Quarterly Journal of Economics, 106, pp. 557-586.

Hendry, D. F. (1979). *Predictive failure and econometric modelling in macroe-*

conomics: The transactions demand for money. In: P. Ormerod (ed.). Economic Modelling. London.

Hendry, D. F. (2006). *A comment on specification searches in spatial econometrics: The relevance of Hendrys methodology*. In: Regional Science and Urban Economics, 36, pp. 309-312.

Hepple, L. W. (2004). *Bayesian Model Choice in spatial econometrics*. In: Lesage, J. P., Pace, R. K. (eds.). Advances of Econometrics, 18, pp. 101-126.

Kahnert, R. (1998). *Wirtschaftsentwicklung, Sub- und Desuburbanisierung*. In: Informationen zur Raumentwicklung, 24, pp. 509-520.

Kaltenbrunner, R. (2003). *Scholle und Rand. Wohnen und Suburbanisierung-ein kaum steuerbarer Zusammenhang*. In: Raumordnung und Raumforschung, 61, pp. 319-333.

Keilbach, M. (2000). *Spatial Knowledge Spillovers and the Dynamic of Agglomeration and Regional Growth*. Heidelberg.

Kelejian, H. H. and Prucha, I. R. (1998). *A generalized spatial two stage least square procedure for estimating a spatial autoregressive model with autoregressive disturbances*. In: Journal of Real Estate Finance and Economics, 17, pp. 99-121.

Kelejian, H. H. and Robinson, d. P. (2004). *The influence of spatially correlated heteroskedasticity on tests for spatial correlation*. In: Anselin, L., Florax, R. J. G. M., Rey, S. J. (eds.). Advances in Spatial Econometrics, pp. 79-98. Berlin.

Keller, W. (2001). *Geographic Localization and International Technology Diffusion*. In: American Economic Review, 92, pp. 120-142.

Kühn, M. (2001). *Regionalisierung der Städte. Eine Analyse von Stadt-Umland-Diskursen räumlicher Forschung und Planung*. In: Raumordnung und Raumforschung, 59, pp. 402-411.

Kim, C. W., Phipps, T. T., and Anselin, L. (2003). *Measure the benefits of air quality improvement: A spatial hedonic approach*. In: Journal of Enviornmental Economics and Management, 45, pp. 24-39.

Klotz, S. (1996). *Ökonometrische Modelle mit raumstruktureller Autokorrelation: eine kurze Einführung*. Mimeo: Univeristät Konstanz. Can be downloaded from www.uni-konstanz.de/FuF/wiwi/pohlmei/www.html.

Koop, G. (2003). *Bayesian Econometrics*. Wiley.

Krugman, P. (1979). *A model of innovation, technology transfer, and the world distribution of income*. In: Journal of Political Economy, 87, pp. 253-266.

Krugman, P. (1991). *Increasing Returns and Economic Geography*. In: Journal of Political Economy, 99, pp. 483-499.

Krugman, P. (1998a). *Space: The Final Frontier*. In: Journal of Economic Perspectives, 12, pp. 161-174.

Krugman, P. (1998b). *Whats New About New Economic Geography?* In: Oxford Review of Economic Policy, 14, pp. 7-17.

Lechevalier, S., Ikeda, Y., and Nishimura, J. (2007). *Investigating Collaborat ive R&D Using Patent Data: The Case Study of Robot Technology in Japan*. Discussion Paper Series No. 498, The Institute of Economic Research, Hitot subashi University.

LeSage, J. P. (2000). *Bayesian estimation of limited dependent vaiable spatial autoregressive models*. In: Geographical Analysis 32, pp. 19-35.

LeSage, J. P. and Parent, O. (2007). *Bayesian Model Averaging for Spatial Econometric Models*. In: Geographical Analysis, 39, pp. 241-267.

Lichtenberg, F. and van Pottelsberghe de la Potterie, B. (1996). *International R&D Spillovers: A Re-Examination*. NBER Working Paper No. 5668.

Lucas, R. E. (1988). *On the Mechanics of Economic Development*. In: Journal of Monetary Economics, 22, pp. 3-39.

Madigan, D. and York, J. (1995). *Bayesian Graphical Models for Discrete Data*. In: International Statistical Reveiws, 63, pp. 215-232.

Marshall, A. (1920). *Principles of Economics*. 8th edition. London.

Maurseth, P. B. and Verspagen, B. (2002). *Knowledge spillovers in Europe. A patent citations analysis.* In: Scandinavian Journal of Economics, 104, pp. 531-545.

Moll, P. (2000). *Probleme und Ansätze zur Raumentwicklung in der europäischen Grenzregion Saarland-Lothringen-Luxemburg-Rheinland-Pfalz-Wallonien.* In: Raumordnung und Raumforschung, 58, pp. 265-275.

Moran, P. A. P. (1948). *The interpretation of Statistical Maps.* In: Journal of the Royal Statistical Society, 10, pp. 243-251.

Moran, P. A. P. (1950). *Notes on Continuous Stochastic Phenomena.* In: Biometrika, 37, pp. 17-33.

Morgan, K. (1997). *The learning region: institutions, innovation and regional renewal.* In: Regional Studies, 31, pp. 491-503.

Mur, J. (1999). *Testing for spatial autocorrelation: Moving average versus autoregressive processes.* In: Enviornment and Planning, 31, pp. 1371-1382.

Olejnik, A. (2008). *Using the spatial autoregressively distributed lag model in assessing the regional convergence of per-capita income in the EU25.* In: Regional Science 87, pp. 371-384.

Ord, K. (1975). *Estimation Methods for Models of Spatial Interaction.* In: Journal of American Statistical Association, 70, pp. 120-126.

Raferty, A., Madigan, D., and Hoeting, J. (1997). *Bayesian Model Averaging for Linear Regression Models.* In: Journal of of the American Statistical Association, 92, pp. 179-191.

Raftery, A. E. and Lewis, S. (1992a). *Comment: One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo.* In: Statistical Science, 7, pp. 493-497.

Raftery, A. E. and Lewis, S. (1992b). *How many iterations in the Gibbs sampler?* In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. (eds.). Bayesian Statistics, 4, pp. 763-773.

Raftery, A. E. and Lewis, S. (1995). *The number of iterations, convergence diagnostics and generic Metropolis algorithms.* In: Gilks, W. R., Spiegelhalter, D. J., Richardson, S. (eds.). Practical Markov Chain Monte Carlo. London.

Rivera-Batiz, L. A. and Romer, P. M. (1991). *International trade with endogenous technological change.* In: Economic Review, 35, pp. 971-1004.

Romer, P. (1986). *Increasing Returns and Long-Run Growth.* In: Journal of Political Economy, 94, pp. 1002-1037.

Romer, P. (1990). *Endogenous Technical Change.* In: Journal of Political Economy, 98, pp. 71-102.

Santolini, R. (2008). *A spatial cross-sectional analysis of political trends in Italian municipalities.* In: Regional Science, 83, pp. 431-451.

Scotchmer, S. and Thisse, J.-F. (1992). *Space and Competition: A Puzzle.* In: Annals of Regional Science, 26, pp. 269-286.

Tappeiner, G., Hauser, C., and Walde, J. (2008). *Regional knowledge spillover: Fact or artifact?* In: Research Policy, 37, pp. 861-874.

Theil, H. and Goldberger, A. S. (1961). *On pure and mixed statistical estimation in economics.* In: International Economic Review, 2, pp. 65-78.

Trivezg, F. J., M. J. (2004). *Some proposals for discriminating between spatial process.* In: Getis A., Mur J. and Zoller H. G. (eds). Spatial Econometrics and Spatial Statistics, pp. 150-175.