Discussion Paper No. 10-005

Spatial Model Selection and Spatial Knowledge Spillovers: A Regional View of Germany

Torben Klarl

ZEW
Zentrum für Europäische Wirtschaftsforschung GmbH
Centre for European Economic Research

Discussion Paper No. 10-005

Spatial Model Selection and Spatial Knowledge Spillovers: A Regional View of Germany

Torben Klarl

Non-technical summary

Knowledge and technological change are often assumed to be the driving forces for long run economic growth. Regions with a higher level of knowledge compared to other regions exhibit a higher per-capita income on average. Agglomeration effects can lead to a steady increase of income and widen the productivity gap between rich and poor regions.

It follows that regions are spatially related. From this point of view, it is not surprising that neighbouring region's spillover affects the own economic performance in a positive or negative manner. Recent contributions show that particularly knowledge spillover leads to an enhancing of agglomeration effects: superior regions with respect to per capita income are attractive for company establishment, due to the fact of superior human capital endowment compared to other regions. These core regions play the role of so called knowledge generators because of the fact that they may benefit from well established research density. Densely located underperforming neighbouring regions may benefit from spillover created by core regions to boost their economic performance. Of course, absolute distance towards the core region affects underperforming regions probability of participating in knowledge spillover potential. The implication is, that regions are not only spatially related to each other but they are also spatially heterogeneous with respect to their knowledge potential, which is endogenously influenced by knowledge spillovers.

One essential purpose of spatial econometrics is to uncover knowledge spillovers from regional data. Although several empirical contributions have devoted to the identification of spatial dependence in several contexts, the majority of them do not control for spatial heterogeneity in the data.

The aim of this paper is to introduce a spatial model selection mechanism for cross sectional data, which controls for both spatial heterogeneity and spatial dependence. Furthermore this mechanism also considers the fact of spatial limited spatial effects. So far existing model selection criteria only tackle the problem of spatial dependence. Using regional data on German NUTS-2 regions, this paper investigates, whether own regional economic performance - measured by per capita productivity - is in-

fluenced by neighbouring regions.

The selection mechanism which is established in this paper provides evidence of spatially bounded spillover effects, particularly caused by patenting activity of neighbouring regions, if we control for spatial heterogeneity of regions. If one neglects this heterogeneity, results and, eventually conclusions regarding the significance of determining factors of regional economic performance can be severely biased.

Nicht-technische Zusammenfassung

Wissen und technologischer Wandel sind die treibenden Kräfte für wirtschaftliches Wachstum. Regionen, die ein im Vergleich zu anderen Regionen höheres Wissenspotential aufweisen, können in der Regel durch ein im Durchschnitt zu anderen Regionen höheres Pro-Kopf-Einkommen charakterisiert werden. Agglomerationseffekte sorgen dafür, dass das Einkommens- und Produktionsdifferential dieser im Durchschnitt prosperierenden Regionen mit den verbleibenden Regionen im Zeitverlauf immer größer wird.

Regionen interagieren und stehen in einer räumlichen Abhängigkeit zueinander. Es ist folglich nicht verwunderlich, wenn man annimmt, dass die von Nachbarregionen generierten Spillover die eigene ökonomische Leistungsfähigkeit beeinflussen. In der Literatur werden insbesondere sog. Wissensspillover eine große Bedeutung zugemessen, da sie die Agglomerationstendenzen verstärken können: Regionen mit einem im Vergleich hohen Pro-Kopf-Einkommen sind attraktiv für Unternehmensansiedelungen, da in diesen Regionen relativ viel Humankapital akkumuliert und genutzt wird. Diese Kernregionen produzieren vermehrt Wissen, da in diesen eine hohe Dichte an Hochschulen, Universitäten und Forschungseinrichtungen vorzufinden ist. Schwächere Regionen, die in unmittelbarer Nachbarschaft liegen, können nun ihre eigene Leistungsfähigkeit durch Ausnutzung von Wissenspillover der Kernregion verbessern. Je weiter eine Region jedoch von dem Zentrum der Wissensbildung entfernt liegt, desto geringer ist die Chance, an diesen Wissensspillovern teilzuhaben. Der Wissensverbreitung sind demnach räumliche Grenzen gesetzt, selbst wenn man annimmt, dass jede Region das generierte Wissen gleich gut absorbieren kann. Demnach sind Regionen nicht nur räumlich voneinander abhängig, sie sind auch heterogen bzgl. des Wissenspotentials, welches endogen durch Wissenspillover beeinflusst wird.

Ein Anwendungsgebiet der räumlichen Ökonometrie ist nun gerade darin zu sehen, diese Wissensspillover in Regionaldatensätzen sichtbar zu machen. Obwohl sich einige Beiträge der Identifikation von räumlicher Abhängigkeit von Regionen gewidmet haben, so ist doch zu konstatieren, dass diese häufig nicht für räumliche

Heterogenität in den Daten kontrollieren.

Im Rahmen dieses Aufsatzes wird ein Modellselektionsmechanismus für räumliche Querschnittsdaten vorgestellt, der nicht nur wie bereits bestehende Modellauswahlverfahren, ausschließlich für räumliche Abhängigkeit kontrolliert, sondern auch die Heterogenität von Regionen und für die Tatsache, dass räumliche Effekte wie Wissenspillover in ihrer Wirkung räumlich beschränkt sind. kontrolliert.

Für deutsche NUTS-2 Regionen wird daraufhin untersucht, ob die wirtschaftliche Leistungsfähigkeit von heterogenen Regionen durch Nachbarregionen beeinflusst wird. Als Ergebnis lässt sich festhalten, dass die Anwendung des Selektionsmechanismus Hinweise auf räumlich beschränkte Spillovereffekte (etwa bei den Patenten) und auf räumliche Heterogenität liefert. Würde man die Heterogenität im Rahmen der Schätzung vernachlässigen, führt dies zu falschen Schlussfolgerungen bzgl. der wirtschaftlichen Leistungsfähigkeit von Regionen.

Spatial model selection and spatial knowledge spillovers: A regional view of Germany*

Torben Klarl[†] [‡]

January 8, 2010

Abstract

The aim of this paper is to introduce a new model selection mechanism for cross sectional spatial models. This method is more flexible than the approach proposed by Florax et al. (2003) since it controls for spatial dependence as well as for spatial heterogeneity. In particular, Bayesian and Maximum-Likelihood *(ML)* estimation methods are employed for model selection. Furthermore, higher order spatial influence is considered. The proposed method is then used to identify knowledge spillovers from German *NUTS-2* regional data. One key result of the study is that spatial heterogeneity matters. Thus, robust estimation can be achieved by controlling for both phenomena.

Keywords: Spatial econometrics, Bayesian spatial econometrics, Spatial heterogeneity
JEL Classification Number: C11, C31, C52

---

*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.

[†]Address: Centre for European Economic Research (ZEW), P.O. Box 103443, D-68034 Mannheim, Germany. Email: klarl@zew.de

[‡]Address: Department of Economics, University of Augsburg, Universitätsstraße 16, D-86159 Augsburg, Germany. Email: torben.alexander.klarl@wiwi.uni-augsburg.de

1 Introduction

It is an undisputable fact that knowledge and technological change are driving forces for long run economic growth. It is rather intuitive that spatial barriers to knowledge diffusion can be used as an argument for income and productivity differentials between regions. That should be considered as one reason why we observe clusters of regions or countries with similar productivity and income values and thus spatial patterns in economic long run growth. Regions (take cities for example), which are more productive are more attractive for innovative companies. Consequently, these regions become more attractive for investments and this process leads to a more and more increasing productivity differential between regions and. From this point of view, it is not surprising that economic growth and agglomeration are positively correlated[1].

Furthermore, it is known that knowledge concentration enforces region-specific growth differentials. As mentioned by Fujita and Thisse (2002), knowledge spillovers can be interpreted as a source for sustainable regional growth, given that decreasing returns of learning are excluded. The empirical literature has not paid much attention to the grasp of knowledge spillovers. Anselin et al. (1997) and Anselin et al. (2000) are two of a few studies that mentioned concrete numbers for the scope of knowledge spillovers. Anselin et al. (1997) found that a significant positive effect can be detected within a 50 mile radius of metropolitan statistical areas *(MSAs)*, but only for university research. Concerning private $R\&D$, such a significant effect of knowledge transfer could not be detected. Anselin et al. (2000), using a similar setup as Anselin et al. (1997) have shown, that not only spillovers within $MSAs$ but also between $MSAs$ can be found. The key cognition of the study performed by Anselin et al. (2000) is that without exact geographical distance measures it can be shown that spatial influence is bounded locally. Audretsch and Mahmood (1994) for instance, point to the limited effects of knowledge spillovers for metropolises borders. Furthermore, applicants of spatial econometrics sometimes neglect that spatial effects can appear as two types: the first is spatial dependence, the second spatial

[1] See Baldwin and Martin (2003).

heterogeneity. Spatial dependence, which is consistently assumed in the above-mentioned studies, covers the not directly observable dependence between regions, caused by spillovers for instance. In contrast to spatial dependence, spatial heterogeneity takes into account that spatial effects are not uniformly distributed across space and that outliers could exist. From an econometrician's view, this should be interpreted as a spatial kind of heteroscedasticity. Although several arguments militate in favour of the significance of spatial heterogeneity[2], this aspect is not commonly "seen as a serious problem in spatial regression"[3].

One reason could be that spatial econometrics is still a developing discipline, dominated by *(ML)* based estimation methods, which are primarily not designed for dealing with spatial heterogeneity. In contrast to *(ML)* methods, Bayesian methods are well suited to incorporate spatial heterogeneity within the model setup and thus provide a fruitful way of analysing spatial phenomena and could lead to a robustification of coefficient estimates.

Another crucial issue within the spatial econometric research agenda is the topic of adequate model selection from a pool of possible models. A widely used selection strategy has been introduced by Florax et al. (2003). This strategy is based on Lagrange-Multiplier tests ($\mathcal{LM}$) tests but has some drawbacks: First, it suffers from the test strategy itself, second, only spatial error or spatial lag models are possible, but no combination of both; and third, the procedure does not control for spatial heterogeneity. This forms the scope for future research.

Thus, the aim of this paper is first, to introduce a new model selection mechanism which is more flexible than the method proposed by Florax et al. (2003) through controlling for spatial dependence as well as for spatial heterogeneity. In particular, Bayesian and *(ML)* methods are employed to select a suitable model. The paper is organised as follows: After motivating spatial models and their corresponding

[2] Anselin (1988) for instance comment on page 13 with respect to the importance of spatial heterogeneity in econometrics 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)."

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

spatial weights in chapter two and three and introducing the concept of higher order spatial influence in chapter four, chapter five deals with the model selection mechanism. The first part of chapter five is dedicated to give a brief sketch of the method established by Florax et al. (2003), while the focus of the second part is laid on a new spatial model selection method. Chapter six deals with an application of the proposed method and tries to identify spatial knowledge spillovers from German *NUTS-2* data. Chapter seven provides a short summary of the obtained results and highlights some avenues for further research.

2 Motivation

I start with the basic cross-section regression model which can be written as follows:

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

where, to make it easy, y is a non stochastic $(N \times 1)$ vector of observations, 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 control for spatial dependence. But if spatial dependence, particularly spatial autocorrelation, exists in the data, and if it is neglected when estimating equation (1), results may not be consistent[4]. This argumentation is familiar when talking about estimation problems within a pure time series context, for instance performing an *OLS* regression with non-stationary time series.
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, a spatial lag model or a spatial $AR(r)$ model has to be estimated, with r as the degree of spatial influence. Second, spatial dependence is only observed in the error vector ϵ, which means that a spatial error or equivalently, a spatial $MA(r)$ model has to be set up. Finally a combination of both, i.e. a spatial lag and a spatial error model can be taken under consideration. Similar

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

to the time series context, the third type can be labeled as an spatial $ARMA(r, r)$ model. The vast majority of empirical studies dealing with spatial events sets $r = 1$. Given that is true, equation (1) can be expanded to 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 $X = [x_1, x_2, ..., x_K]$, $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_M]$, the $(K\times1)$ vector $\beta^X = [\beta_1^X, \beta_2^X, ..., \beta_K^X]'$, the $(M \times 1)$ vector $\beta^{\tilde{X}} = [\beta_1, \beta_2, ..., \beta_M]'$ and $\kappa \sim N(0, \sigma^2 I)$.

The parameter ρ denotes 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 $M = K$. We do so here.

The $(N \times K)$ matrix X comprise 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 defining $\tilde{\beta} = [\rho, \beta^X, \beta^{\tilde{X}}, \lambda]'$ leads to

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

Expression (2) can be considered also as a spatial $ARIMA(r, I, r)$-model, with I as the degree of integration if $|\rho| = 1$. If we observe a significant coefficient of ρ close to one[5], a spatial $ARIMA(r, I, r)$-model should be estimated to avoid results based on a spurious regression.

It can be concluded that time series and spatial econometrics are closely connected. But it has to be pointed out that, in contrast to time series phenomena, which go in one direction by definition, spatial spillovers are often characterized by feedback-processes, as mentioned before.

Although it is common to set $\kappa \sim N(0, \sigma^2 I)$, it is more realistic to assume $\kappa \sim N(0, \sigma^2\Omega)$ with $\sigma_i = \tilde{h}(f_i'\alpha)$ and $h(\cdot) > 0$ as an unknown, continuous function.

[5]Of course it could be difficult to decide ex ante, whether one is confronted with a highly persistent or an unit root process with respect to space.

These elements are defined as the diagonal elements of the error covariance matrix $\sigma^2\Omega$. This is one way to introduce spatial heterogeneity in the model setup.

Interestingly, Keilbach (2000) and Klotz (1996) argue that spatial heterogeneity does not have to be seen a serious in spatial econometrics context this should not be taken for granted. Remember for instance that some regions do not follow the same spatial regime 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.

The challenge of estimating a heterogeneous spatial model is that when allowing for heteroscedasticity we have to estimate N additional parameters for each σ_i. Of course, this leads to the so called "degree of freedom" problem, because we do not have enough observations to compute an estimate for every point located in space. One way to deal with this problem is to refer to Bayesian econometrics. Bayesian methods avoid the degree of freedom problem by using informative priors. We will see later, that the prior distribution for our N diagonal elements of Ω is 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 parameters of the diagonal elements of Ω by adding a single parameter r to our standard regression procedure.

Hence, the estimation strategy is defined as follows: start with an estimation of a spatial $ARMA(r, r)$-model with homogeneous errors based on equation (2). Equation (2) can be consistently estimated via Maximum-Likelihood (ML) as mentioned by Anselin and Rey (1991). Since (ML) based models are not suitable to model spatial heterogeneity, Bayesian models with the additional assumption of heterogeneous errors are introduced. After performing the model selection mechanism, a direct model comparison of the (ML) based and the Bayesian models should be used to select the model which best fits to the data-generating process.

3 Spatial weights

To the best of my knowledge, there is no concise theory about how to specify the "correct" spatial weight matrix W. Of course, answering this question adequately is the big deal in applied spatial econometrics. Therefore, the choice of the spatial weights should be done according to the specific research topic. The first question is how to find a proxy for spatial proximity. One approach is to say that spatial proximity is best approximated by geographical distances. Another way is to argue that geographical borders are less important for spatial proximity and for this reason one should better rely on non geographical data, such as trade shares[6] or data on FDI[7].

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, on the first sight, intuitive to refer to trade shares or FDI data, for instance, to find a proxy for spatial proximity which is related to knowledge diffusion. But as noted by Krugman (1991) "[k]nowledge flows are invisible; they leave no paper trail by which they may be measured and tracked[...]". Thus, solely relying on this data, tell us not the whole story. Second, there is a methodological problem: it is very likely that these weights are endogenous and therefore lead to biased estimators, unless using an IV or GMM approach[8].

Hence, the majority of the literature refers to geographical weights. It is common to use geographical distances Keller (2001), or more precisely, to use great circle distances between regions' centroids Anselin (1988). But this contains the inherent assumption that knowledge spillover sources are located in the regions' centroids. Another way, which is also considered in this study, is simply to refer to a binary weighting scheme[9]. 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. Of course, the definition of the weighting scheme is irrelevant for the proposed selection method.

[6] Refer to Coe and Helpman (1995).

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

[8] Finding valid and non weak instruments in this context seems to be rather difficult.

[9] Refer to Tappeiner et al. (2008).

Thus, for the symmetric $(N \times N)$ matrix W with weights w_{ij} the following rule is defined:

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

Often, this matrix is weighted or standardized because this facilitates the interpretation of the estimated coefficients[10] and guarantees that the Moran's $\mathcal{I}$, which stands for an indicator for spatial dependency, is situated in the interval $[-1; 1]$[11]. Using the weighting scheme proposed by Anselin (1988), for the standardized elements w_{ij}^{+} of W^{+} it can be written:

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

In this way we have created row-standardized spatial weighting elements $w_{ij} \in W^{+}$ which are used in the following estimation exercise.

4 Higher order spatial influence specification

One major drawback of model (2) is, that it is not designed to account for higher order spatial dependencies. To obtain a higher order weighting matrix W^{+r} for $r = \{1, ..., R\}$ we should increase the power of the simple contiguity matrix. Labelling the order of the spatial dependency with $r = \{1, 2, 3, ..., R\}$ then $\tilde{X}$ can be expanded as a $(N \times (R \times M))$ matrix as follows:

$$\tilde{X}^{++} = \left[\begin{pmatrix} \tilde{x}_{11}^{1} & \tilde{x}_{12}^{1} & \cdots & \tilde{x}_{1M}^{1} \\ \tilde{x}_{21}^{1} & \tilde{x}_{22}^{1} & \cdots & \tilde{x}_{2M}^{1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{N1}^{1} & \tilde{x}_{N2}^{1} & \cdots & \tilde{x}_{NM}^{1} \end{pmatrix}, ..., \begin{pmatrix} \tilde{x}_{11}^{R} & \tilde{x}_{12}^{R} & \cdots & \tilde{x}_{1M}^{R} \\ \tilde{x}_{21}^{R} & \tilde{x}_{22}^{R} & \cdots & \tilde{x}_{2M}^{R} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{N1}^{R} & \tilde{x}_{N2}^{R} & \cdots & \tilde{x}_{NM}^{R} \end{pmatrix} \right] \tag{6}$$

or in short hand notation:

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

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

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

Defining $P = [\rho^1, \rho^2, ..., \rho^R]'$, $\tilde{y} = [W^{+1}y, ..., W^{+R}y]$, $\Lambda = [\lambda^1, \lambda^2, ..., \lambda^R]'$, $\tilde{\epsilon} = [W^{+1}\epsilon, ..., W^{+R}\epsilon]$, and the $((R\times M)\times 1)$ dimensional vector $\beta^{++} = [\beta^{\tilde{X}^1}, \beta^{\tilde{X}^2}, ..., \beta^{\tilde{X}^R}]'$ with the the $(M \times 1)$ vector $\beta^{\tilde{X}^r} = [\beta_1^{\tilde{X}^r}, ..., \beta_M^{\tilde{X}^r}]$ 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 $R = 1$ the model collapses to model (2). From the general $SAC(r, r)$ model (8) we can derive three major submodels for a given degree of higher spatial influence $r = \{1, ..., R\}$: 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:

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

with $\kappa \sim N(0, \sigma^2\Omega)$. For the $SEM(r)$ the following equation holds:

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

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 model (8) and its submodels (9), the estimation of model (10) and model (11) could lead to biased and inconsistent OLS estimates $\forall r \in R$. Take submodel (9) for instance: $\tilde{y}P$ is correlated not only with κ but also with neighbourings of κ. If all elements of $\tilde{y}P$ are zero OLS estimates are unbiased but inefficient $\forall r \in R$. 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 $\kappa \sim N(0, \sigma^2 I)$, $\forall r \in R$.

OLS is even more unbiased if estimating a spatial error model, thus referring on submodel (10) $\forall r \in R$[12].

5 Spatial model selection

In this section a spatial model selection strategy is introduced, which can be interpreted as an expansion of the proposed and well established strategy by Florax et al. (2003). Before introducing the new model selection method, the classic method of Florax et al. (2003) for cross section analysis is briefly sketched. As we will see, this selection methods ignores the appearance of spatial heterogeneity.

5.1 Established method

Start by estimating an initial model $y = X\beta + \epsilon$, which goes in line with the stated assumption of model (1). Second, on the basis of the estimated model, Lagrange Multiplier ($\mathcal{LM}$) tests on the basis of of model (1) are conducted to test in favour of a spatial lag or spatial error model. If the null hypothesis of no spatial correlation is rejected, then 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 adopted in the following manner:

$$\mathcal{LM}_{\rho^r} = \frac{\left(\frac{e'W^{+r}e}{s^2}\right)^2}{T}, \tag{12}$$

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 the contrary, the test statistic for $\mathcal{LM}_{\lambda^r}$ under $H := \lambda^r = 0$, $\forall r$ can be written as:

$$\mathcal{LM}_{\lambda^r} = \frac{\left(\frac{e'W^{+r}y}{s^2}\right)^2}{NJ}, \tag{13}$$

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 (1).

[12]Refer to appendix 1 for corresponding proofs of the last mentioned propositions.

Third, if for both, the $\mathcal{LM}_{\rho^r}$ and $\mathcal{LM}_{\lambda^r}$ the null hypothesis cannot be rejected, then the initial model (1) should be used. If both tests indicate significance, it is advisable to compute the robust versions of $\mathcal{LM}_{\rho^r}$ and $\mathcal{LM}_{\lambda^r}$ statistics to come to a final decision.

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

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

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

$$\tilde{\mathcal{LM}}_{\lambda^r} = \frac{\left(\frac{e'W^{+r}e}{s^2} - T(NJ)^{-1}\frac{e'W^{+r}y}{s^2}\right)^2}{T[1 - T(NJ)]^{-1}}. \tag{15}$$

Again, if both robust test statistics are significant, then the following rule of thumb should be employed: If $\tilde{\mathcal{LM}}_{\rho^r} > \tilde{\mathcal{LM}}_{\lambda^r}$, then one should decide to estimate a spatial lag model, otherwise if $\tilde{\mathcal{LM}}_{\rho^r} < \tilde{\mathcal{LM}}_{\lambda^r}$, then one should refer to a spatial error model.

Given, only $\mathcal{LM}_{\rho^r}$ is significant but $\mathcal{LM}_{\lambda^r}$ 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. Experiment-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. Florax et al. (2003) instead have shown that the classical top down approach, i.e. relying on the non robust $\mathcal{LM}$ tests, outperforms the robust strategy in terms of performance and accuracy. Thus, the same authors emphasize, that one should use the classic approach when testing for spatial effects. From the discussion above, it can be concluded that this classic strategy is not theoretically justified yet, it is the only systematic approach of model selection provided by the literature and used in empirical studies[13].

The estimation strategy proposed by authors such as Anselin (2005) has three limitations: first, the strategy lacks regarding its underlying tests strategy. More precisely,

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

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. To be inferentially correct, one has to construct a test with the null hypothesis of a spatial lag and the alternative hypothesis of a spatial error model, for instance. Mur (1999) and Trivezg (2004) tackle this problem. 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).

Second, the strategy is exclusive in the sense, that this strategy does not allow for a spatial $ARMA(r, r)$ model specification, which is as mentioned above a combination of spatial lag and spatial error models[14]. There is no reason, why a spatial $ARMA(r, r)$ model should be excluded ex ante. But if doing so and referring to a "from simple to complex" estimation strategy, there is an inherent potential of misspecification[15].

Third, both tests do not sufficiently control for heterogeneity of the error term; nor do they cover the aspect of outliers. In other words, this method neglects spatial heterogeneity entirely. Fortunately, spatial heterogeneity can be elegantly considered in a Bayesian approach. From this way it seems reasonable to incorporate the Bayesian approach within the model selection mechanism.

[14]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 $\mathcal{L}\tilde{\mathcal{M}}_{\lambda^r} > \mathcal{L}\tilde{\mathcal{M}}_{\rho^r}$.

[15]This could lead to serious problems, because even if λ^r differs significantly 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, in line with Florax et al. (2003), a spatial lag model.

5.2 Proposed method

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 or only to a small extent included in standard econometricians tool boxes, such as *EViews* or *Stata*. Second, these methods normally require extended programming techniques. In addition, their use for large sample applications is problematic, because one is often confronted with numerical problems, particularly when calculating the determinant of spatial weight matrix for a large N[16], which is a corner-stone of spatial model estimation[17]. Third, Bayesian methods are often rejected or disregarded by the class of frequentist or "main-stream" econometricians, mainly because of the inherent Bayesian assumption that the vector of coefficients is treated as random, whereas the frequentists treat the vector of coefficients estimate as random[18].

Not ignoring but for this purpose neglecting the ongoing dispute between frequentists and Bayesian supporters, the model selection plan of the proposed strategy can be formulated as follow:

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

2. *Use Moran's $\mathcal{I}$[19] 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*

[16] This is also relevant for spatial panel setups.

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

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

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

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(r,r)) and separate spatial lag (SAR(r)) and spatial error models (SEM(r)) with MLE.*

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. This model assumes spatial homogeneity at the end.*

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 exhibit 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.*

6 Application

6.1 Negligence of spatial dependence

As an application for the proposed selection strategy, a cross section analysis for German *NUTS-2* regions has performed. The aim of this study is to give an answer

to the question, whether spatial knowledge spillover affects regional labour productivity and further tries to answer the question, whether spatial heterogeneity is a justified assumption for the data generating process.

For this purpose I construct a cross-section regression of the regional output, measured as gross value added on regional $R\&D$-effort, human capital H, regional number of patent applications P, regional capital stock K, regional number of low qualified labour force L, regional infrastructure I, spatial weighted dependent variables $\tilde{X}^{++}$ and an East-West dummy d, which covers the fact that East German regions are less productive than West German regions[20].

The initial model, based on a per capita Cobb Douglas production technique, with $ln\left(\frac{Y}{L}\right)$ as the dependent variable, can be written in log-log form as follows[21]:

$$\ln\left(\frac{Y}{L}\right) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + d\gamma + \kappa, \quad (16)$$

or in a more compact manner with $y \equiv \frac{Y}{L}$ as

$$y = X\beta^X + d\gamma + \kappa, \quad (17)$$

with $\beta^X = [\beta^c, \beta^k, \beta^l, \beta^h, \beta^p, \beta^i]'$, $X = [1, k, l, h, p, i]$, and with $\kappa \sim (0, \sigma^2\Omega)$, $\sigma^2\Omega \neq \sigma^2 I$, $\Omega = diag(v_1, ..., v_N)$, and d as East-West dummy. Two remarks regarding the specification of equation (16) or equation (17) should be made: First, as usual, the coefficient vector β^X contains constant production elasticities of the respective values stacked in X. The elasticity of production for labour l in this context is defined as $(1 + \beta^l)$. Therefore, we expect a negative sign of β^l. Second, the inclusion of both $R\&D$ expenditures and patent applications 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 outperform $R\&D$ expenditures regarding their interpretation as a quality measure of innovativeness[22]. From this point of view patent data have been used for this application.

[20] Refer to appendix 8.4 for a detailed description of employed data and variable construction.

[21] For all variables X, we define $x \equiv \ln(X)$.

[22] See Lechevalier et al. (2007) for instance.

Let us start with the first step of the above strategy. An OLS estimation of the initial model has been performed first[23].

Column 1 of table 6 reports a simple estimation of y on k and l and a East-West dummy d. The values of the elasticity of production for capital and labour indicate the expected positive sign and have the expected dimension[24]. 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.

For this specification, both Moran's $\mathcal{I}$ tests for first and second order spatial influence cannot reject the null hypothesis of no spatial correlation.[25] The findings indicate that, the $\mathcal{LM}$ lags for $r = 1$ and $r = 2$ are not significant. This is also 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.

Thus, a certain 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 can be detected for the specification in first column of table 6. The second column of table 6 reports the estimation results of a knowledge capital augmented regression. Knowledge capital is approximated by human capital ($\ln(H)$) and patent applications ($\ln(P)$). Further, in this regression it is controlled for public infrastructure ($\ln(I)$) expenditures.

For all three additionally included variables we should expect a positive sign with respect to their specific regressors. This is true for the estimated coefficients of human capital and infrastructure, but not for patents, which is of course counter intuitive at first glance. Referring to the significance level, we find that patents are not significant at the 10% significance level. This is also true for infrastructure which is not significant at the 10% significance level. In contrast, human capital is significant at the 5 %. Referring to the test statistics depicted in the lower part of the second column of table 6, it should be noted, that the $\mathcal{LM}$ test for spatial lag is

[23] 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`.

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

[25] The choice of the maximum dimension of R has been carried out on the basis of the variable individual Moran's $\mathcal{I}$. A graphical representation can be found in figure 1 and 2.

significant at the 5% significance level. Moran's $\mathcal{I}$ for $r = 1$ suggests, that a spatial error model should be estimated which is underpinned by the significant $\mathcal{LM}$ test for the spatial error component for $r = 2$.

6.2 Spatial dependence

Given our selection strategy, we should expand model (16) by exogenous spatial lagged variables. Keeping in mind our results obtained from figures 1 and 2, first order spatial lags of human capital $\ln(H^{+1})$, of patents $\ln(P^{+1})$ and of infrastructure $\ln(I^{+1})$. Additionally, the second order spatial lag of patents $\ln(P^{+2})$ have been included. Stacking these 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 $\beta^{\tilde{X}^2} = [\beta^{p^{+1}}]$, lead to the following expansion of equation (16):

$$\ln(y) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + \tag{18}$$
$$+ \ln(H^{+1})\beta^{H^{+1}} + \ln(P^{+1})\beta^{P^{+1}} + \ln(I^{+1})\beta^{I^{+1}} + \ln(P^{+2})\beta^{P^{+2}} + d\gamma + \kappa,$$

or again in compact notation as an expansion of equation (17):

$$y = X\beta^X + \tilde{X}^{++}\beta^{++} + d\gamma + \kappa, \tag{19}$$

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

The estimation results for model (19) can be found in column 3 of table 6. Once again, we expect positive effects from neighbouring regions. But with the exception of patents, we find negative signs of coefficients for neighbouring human capital and infrastructure. Moreover, the latter two coefficients are highly non-significant. The negative second order spillover coefficient of patents is highly insignificant, too. In contrast, the first order neighbouring patent activity has a significant positive effect on the domestic labour productivity. If we refer to our spatial test statistics depicted in column 3 we find that the $\mathcal{LM}$ test for first order spatial lag is, on the contrary

to column 2, not significant at the 5% level 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 the 10% significance level, whereas the first order $\mathcal{LM}$ error test is now significant at the 5% significance level. Also the first order Moran's $\mathcal{I}$ test is significant at the 5% significance level.

Given that our argument is a based on a 10% significance level, this leads to the conclusion that a first order spatial error model should be modelled because of the fact that $\mathcal{LM}_{\lambda^1} > \mathcal{LM}_{\lambda^2}$. Based on a 5% significance level, a $SEM(1)$ model seems appropriate. The last column of table 6 shows the same regression as column 3 except of excluding the highly non significant spatial second order patent activity. Comparing column 3 and column 4 shows that the exclusion of spatial second order patent activity does not change the sign and significance of the estimated coefficient. In line with the proposed model selection mechanism, we should base the further analysis on the specifications in column 4 of table 6.

6.3 Spatial heterogeneity

From the discussion before we have concluded that we need to expand our regression model (18) by a first order spatial lagged error component. The spatial error model $SEM(r)$ of order $r = 1$ is specified through equation (20):

$$\begin{aligned} \ln(y) = \beta^c + \beta^k \ln(K) + \beta^l \ln(L) + \beta^h \ln(H) + \beta^p \ln(P) + \beta^i \ln(I) + \\ + \ln(H^{+1})\beta^{H^{+1}} + \ln(P^{+1})\beta^{P^{+1}} + \ln(I^{+1})\beta^{I^{+1}} + d\gamma + \epsilon, \end{aligned} \tag{20}$$

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

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

with $\beta^X = [\beta^c, \beta^k, \beta^l, \beta^h, \beta^p, \beta^i]$, $X = [1, k, l, h, p, i]$, $\Lambda = [\lambda^1]$, $W^{++} = [W^{+1}]$, $X^{+++} = [X, \tilde{X}]$, $\beta^{+++} = [\beta^X, \beta^{++}]$ and d as West-East dummy.

With respect to the proposed estimation strategy, model (21) should be estimated via two different ways:

- *The first approach is to estimate this model under the assumption of $\sigma^2\Omega = \sigma^2 I$, implying spatial homogeneity, which is a common assumption in the literature*[26]*. As mentioned above, model (20) should be estimated via Maximum Likelihood.*

- *The second approach is to estimate this model under the assumption that $\sigma^2\Omega \neq \sigma^2 I$. Hence, estimation applying the Bayesian approach is required.*

6.3.1 Maximum-Likelihood approach

First approach, first we have to set up our Likelihood function $\mathcal{L}$. This is for model (20):

$$\mathcal{L} = \frac{\left|\tilde{N}\right|}{(2\pi\sigma^2)^{\frac{N}{2}}} exp\left\{\frac{1}{2\sigma^2}(y - X^{+++}\beta^{+++})'\Theta^{-1}(y - X^{+++}\beta^{+++})\right\}, \tag{22}$$

with $\Theta^{-1} = \tilde{N}'\tilde{N}$ and $|\Theta|^{\frac{1}{2}} = |\tilde{N}|$ and N the numbers of observations.
The corresponding log-likelihood $\ln(\mathcal{L})$ for equation (22) is:

$$\ln\mathcal{L} = -\frac{N}{2}\ln 2\pi - \frac{N}{2}(\sigma^2) + \ln|\tilde{N}| - \frac{1}{2}\xi'\xi, \tag{23}$$

with $\tilde{N} = (I - \lambda^1 W^{+1})$ and $\xi = \tilde{N}(y - X^{+++}\beta^{+++})$. Expression (23) can be written in concentrated form as

$$\ln\mathcal{L}_c \propto \ln|\tilde{N}| - \frac{N}{2}\tilde{\xi}'\tilde{\xi}, \tag{24}$$

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

$$\tilde{\beta}^{+++}_{ML} = (X^{+++\prime}\tilde{N}'\tilde{N}X^{+++})^{-1}X^{+++\prime}\tilde{N}'\tilde{N}y, \tag{25}$$

and

$$\hat{\sigma}^2_{ML} = \frac{1}{N}(\tilde{\xi}'\tilde{\xi}), \tag{26}$$

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

obtained by maximizing equation (23). Equation (24) is highly non linear in the parameter space λ^1. Because both β^{+++} and κ are a function of λ^1 we should use an iterative method to estimate λ^1: First, estimate β^{+++} via OLS, then compute λ^1 on the basis of the concentrated maximum likelihood function (24) and update β^{+++}. Given the updated values of β^{+++} estimate a new λ^1, based on the updated estimated residuals. Convergence is achieved, if values for the residuals and for the estimated values of β^{+++} do not change from one to the next iteration step, which means the difference between the estimated values of $\beta_t^{+++} - \beta_{t-1}^{+++} < \vartheta$ for a small value of ϑ close to zero[27].

From a computational view, it is worth noting, that when referring to Maximum-Likelihood based methods in spatial econometrics, we have to impose a restriction on the parameter space λ^1. Referring to Anselin and Florax (1995a), p. 34, reasonable parameter values are in the range of:

$$\frac{1}{\tilde{\lambda}^1_{min}} < \lambda^1 < \frac{1}{\tilde{\lambda}^1_{max}}. \tag{27}$$

$\tilde{\lambda}^1_{min}$ corresponds to the minimum Eigenvalue of the matrix W^{+1}, whereas $\tilde{\lambda}^1_{max}$ represents the maximum Eigenvalue of W^{+1}. This requires a constrained Maximum-Likelihood maximization. If W^{+r} is row standardized, then $\lambda^1_{max} = 1$. That procedure gets extremely cumbersome with respect to computational issues. More precisely, the computational costs increase with the dimension of the weighting scheme matrix W^{+1}. Alternatively, one can set ex ante values for λ^1, such that $\lambda^1 \in (0, 1)$ which implies only positive spatial error dependence[28].

6.3.2 The Bayesian approach

The second approach dealing with the estimation of model (21) is to rely on a Bayesian approach, with the additional assumption of $\sigma^2\Omega \neq \sigma^2 I$. If spatial heterogeneity matters, the Bayesian estimates should lead to a robustifaction of the ML results.

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

[28] In general, $\lambda^1 \in (-1, 1)$. Because of the fact that knowledge spillovers are generally assumed to be positive, it is assumed that $\lambda^1 \in (0, 1)$.

Based on the Likelihood function expressed by equation (22), a corresponding spatial Bayesian heteroscedastic model is set up. The core of Bayesian econometrics is the Theorem of Bayes (1763). Assume that θ is a vector of unknown parameters to be estimated. Before any data are observed, the true parameter values of θ are uncertain, but we can express our beliefs concerning our expectations about θ. These beliefs are called "a priori" probabilities which are fully represented by the probability function $p(\theta)$. The probability model itself is entirely defined by the likelihood $p(y\,|\theta)$. This likelihood can be described as the key element of Bayesian econometrics, because it contains the entire set of information stemming 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, labeled as $p(\theta|y)$, which is represented by

$$p(\theta|y) = \frac{p(y\,|\theta)p(\theta)}{p(y)}, \tag{28}$$

with $p(y) = (y|\theta)p(\theta)$, defined by the law of total probability. Because the marginal probability of y, $p(y)$ does not contain any information about θ and, moreover, we are only interested in θ itself, we can ignore $p(y)$. Thus the "a posteriori" probability is proportional to the likelihood times the "a priori" probability:

$$p(\theta|y) \propto p(y\,|\theta)p(\theta). \tag{29}$$

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[29]. For instance we can write:

$$p(\theta_1|y) = \int p(\theta|y)d\theta_2 d\theta_3 ... \,. \tag{30}$$

The entire information needed for inference about θ_1 is contained in the marginal distribution of θ_1. What is left to do is to specify the exogenously given priors and the likelihood function.

[29]Refer to Geweke (1993).

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. The v_i have to be treated 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 K exceeds the number of observations N. 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 traces 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 prior 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.

In this way we come back to the assumptions made by the *(ML)* approach. Here, small values of s lead to a skewed parameter distribution. Therefore, the role of v_i is, as with traditional *(WLS)*, to down-weight observations with large variances. For this reason, the degrees of freedom s plays a crucial role when robustifying against outliers. What kind of priors should we attach to s? One way is to assign an improper value to s. The other possibility is to use a proper prior for s which is Gamma distributed:

$$s \sim \Gamma(a, b), \tag{31}$$

with hyperparameter a and b. It has to be pointed 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 should be achieved 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

$$\pi\left(\frac{s}{v_i}\right) \sim \text{i.i.d. } \chi^2(s), \; \forall\, i, \tag{32}$$

with $\pi(\cdot)$ denoting the prior. This implies that the normal mixture model with prior

(32) 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 a 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:

$$\pi(\beta^{+++}) \sim \mathcal{N}(c, T), \tag{33}$$

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

$$\pi(\lambda^1) \sim \mathcal{U}[-1, 1]. \tag{35}$$

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

$$\begin{aligned} & p(\beta^{+++}, \sigma, \lambda^1) = p(\beta^{+++}) p(\sigma) p(\lambda^1) \propto \\ & \propto \left|I - \lambda^1 W^{+1}\right| \sigma^{-N} exp\left\{-\frac{1}{2\sigma^2}(\xi' \Omega^{-1} \xi)\right\} \sigma^{-1} \times \\ & \times \exp\left\{-\frac{1}{2\sigma^2}(\beta^{+++} - c)' T^{-1} (\beta^{+++} - c)\right\}. \end{aligned} \tag{36}$$

From equation (36) the conditional distribution of β^{+++} is obtained from the standard non spatial Bayesian *(WLS)* approach as:

$$p(\beta^{+++} | \lambda^1, \sigma, \Omega, y) \sim \mathcal{N}[H(X^{+++} \tilde{N} \Omega^{-1} \tilde{N} y + \sigma^2 T^{-1} c, \sigma^2 H)], \tag{37}$$

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

$$p(\sigma | \lambda^1, \Omega, \beta^{+++}, y) \propto \sigma^{-(N+1)} exp\left\{\frac{1}{2\sigma^2} \xi' \Omega^{-1} \xi\right\}. \tag{38}$$

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:

$$p\left(\left[\frac{(\sigma^{-2}e_i^2+s)}{v_i}\right]|\beta^{+++},\lambda^1,v_{-i},\lambda^1\right) \sim \chi^2(s+1), \tag{39}$$

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 β^{+++}, λ^1 and Ω. This distribution reads as:

$$p\left[\sum_{i=1}^{N}\frac{e_i^2}{v_i}/\sigma^2|\beta^{+++},\lambda^1,\Omega\right] \sim \chi^2(N). \tag{40}$$

With equation (40) 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 toolboxes.

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

$$p(\lambda^1|\sigma,\Omega,\beta^{+++},y) \propto |\tilde{N}|exp\left\{\frac{1}{2\sigma^2}\xi'\Omega^{-1}\xi\right\}. \tag{41}$$

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 (41), because this type of distribution does 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[30], allows us to approximate the posterior distribution for λ^1[31]. The only problem one has to solve is to find a suitable proposal density. LeSage (2000) suggests assuming a normal or Student t-distribution. Because of the fact that λ^1 has to be handled as a restricted parameter, the sampler rejects values outside the interval $(-1,1)$ from the sample[32]. This is called "rejection sampling"[33].

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

[31] Refer to Gelman et al. (1995).

[32] With this assumption spatial unit roots are excluded.

[33] Refer to Gelfand et al. (1990).

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. To solve this problem, so called posterior probabilities are computed which should give advice on which model is the correct model in terms of probability. The posterior probability p_u^{pos} for model u is given by[34]:

$$p_u^{pos} \equiv p(M_u|y) = \frac{p(y|M_u)}{\sum_{u=1}^{U} p(y|M_u)}. \tag{42}$$

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

$$p(y^*|y) = \sum_{u=1}^{U} p(y^*|y, M_u)p(M_u|y), \tag{43}$$

with $p(y^*|y)$ as the posterior, $p(M_u|y)$ as the posterior model probability and the expression $p(y^*|y, M_u)$ as the likelihood function of model M_u. The reason why model averaging should be used is quite simple. The traditional *(ML)* approach is to choose the single best model based on calculating posterior model probabilities with equation (42) for every model of interest[35]. One has to remember that this rather excluding approach could 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 "worse model" from a 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 meaningfully calculated with improper non-informative priors"[36]. Therefore, I refer to the *MCMC* literature to compute a posteriori model probabilities. This so called MC^3 approach, introduced

[34] Refer to Hepple (2004), p. 105.

[35] A large bulk of literature on Bayesian model averaging *(BMA)* about 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(r)* and *SEM(r)* models.

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

by Madigan and York (1995), is based on a stochastic Markov Chain process which moves through the model space and samples those regions which have a superior model support. Thus, this approach is very efficient because not the entire model space is of interest[37].

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 columns 1 and 2 of table 7. The first regression is a mixture model of spatial lag and spatial error model, the $SEC(1,1)$ model[38]. This regression carried out in order to underpin our model selection on inductive statistics, done in the preceding chapter.

Performing estimation runs for all possible combinations of first order and second order spatial models[39], the $SEC(1,1)$ model has been chosen as the appropriate model on basis of the value of the log-likelihood as can be seen from table 5.

Leaving out the insignificant parameter ρ^1, estimating a pure spatial error model (column 2) and comparing this with column 1, it can be concluded that only minor changes of coefficient estimates are observed in contrast to the estimate for λ^1, which is highly significant. Thus, choosing a $SEM(1)$ seems to be the adequate choice. The estimation results of the $SEM(1)$ model can be found in column 2 of table 7. Comparing the $SEM(1)$ in column 2 of table 7 with the specification depicted in the fourth column of table 6 we can find that with the exception of $\ln(I^{+1})$ all other coefficients have roughly the same dimension, the same sign and the same level of significance.

The results for the second approach, an estimation of the Bayesian counterpart of equation (21) can be found in column 3 of table 7. Before discussing the results, we first should get an intuition of how to interpret the obtained results.

To get estimates from our Bayesian approach, first simulation draws have been made. To ensure stability of simulated results, it seems to be a good strategy to conduct simulation studies based on informative and on non-informative priors, for which

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

[38] See appendix 1 for a deviation of the log-likelihood of the $SEC(1,1)$ model.

[39] See appendix 2 for a summary.

starting values are obtained from a corresponding Maximum-Likelihood estimation procedure. 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 performed.

In total we obtain four models. For each number of draws, one should estimate a model with informative and non informative priors[40]. If we refer to the probabilities p_u^{pos} for the relevant models which can be found in table 4, and if we further draw attention to figure 4, which contains the simulated densities of the λ^1, we can conclude that every chosen model contains nearly the same amount of information with respect to λ^1.

Furthermore, *MCMC*-convergence checks of the four relevant models have been performed[41]. If the means and variances for the posterior estimates are similar for all runs, convergence seems to be ensured. Furthermore, convergence tests for all regressions show that convergence of the sampler is guaranteed for all simulations. Because of the fact that model $(4, 1)$ has a slightly higher probability to be the correct model and, at the same time, requires fewer runs, it is the preferred model in this application. The estimation results for this model can be found in the third column of table 7.

If we now turn back to table 7 and compare the heteroscedastic Bayesian counterpart in column 3 with the homoscedastic *(ML)* based estimation in column 2 we can conclude that estimation results do not differ dramatically with respect to λ^1. The Bayesian model estimates a lower value for the spatial lag component λ^1 compared to the homoscedastic Maximum-Likelihood procedure, but both coefficient values are highly significant on the 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 in the data. Figure 5 shows

[40]Because of the fact, that the initial model estimation results on which Bayesian model specification is based are drawn in column (4) of table 7, we label variants of the Bayesian model as model (4,1), (4,2), (4,3) and (4,4).

[41]Please refer to appendix 8.6 for a description of convergence criteria and refer to tables 8 to 11 for convergence diagnostics of all selected models.

a plot of the mean of the v_i draws, which should serve as an estimate of these relative variance terms. If the assumption of spatial homogeneity is appropriate, all elements of Ω should be equally weighted with value one. Obviously, this is 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, and thus, a robustification of *(ML)*-parameter estimates can be achieved by conducting an estimation based on Bayesian principles. With respect to the conducted study, this is achieved for labour, human capital, physical capital, the first order spatial lagged patents, as well as for the first order spatial error component. Thus, first order neighbouring patent activity positively influences the region specific labour productivity. Neglecting this fact would yield an underestimation of region specific labour productivity.

7 Conclusion

The aim of this paper was to introduce a new spatial model selection mechanism for cross section data. In particular, Bayesian and *(ML)* methods have been employed to select a suitable model. The proposed model is more flexible than the approach introduced by Florax et al. (2003), as it incorporates the two major issues of spatial econometrics: spatial dependence and spatial heterogeneity. Furthermore, the new model selection mechanism framework has been used to identify spatial knowledge spillover from German *NUTS-2* regions. It has been found that first order neighbouring patent activity positively influences the region-specific labour productivity, while own patent activity does not exhibit a significant influence 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. On the basis of a spatial Bayesian analysis it was further shown that spatial heterogeneity is a reasonable assumption, and controlling for it within a spatial Bayesian framework leads to a robustification of the parameter estimates.

Of course there are avenues for further research: first of all, the question which is a suitable weighting matrix is a still not answered question in spatial econometric ap-

plication. Until today, there is no concise theory about how to choose the "correct" weighting scheme as it is for the estimation results itself. This is from fundamental importance due to the fact that the model selection mechanism is a dependent function of the weighting scheme. Second, Bayesian model convergence checks sometimes lead to misleading conclusions regarding the sampler convergence. Third, it is still unclear how model selection procedures for spatial models an sufficiently control for spatial unit roots.

References

Almus, M. and Czarnitzki, D. (2003). *The Effects of Public R&D Subsidies on Firms' Innovation Activities: The Case of Eastern Germany*. In: Journal of Business and Economic Statistics, 21, pp. 226-236.

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

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.

Anselin, L., Varga, A., and Acs, Z. J. (2000). *Geographical Spillovers and University Research: A Spatial Econometric Perspective.* In: Growth and Change, 31, pp. 501-515.

Aschauer, D. (1989). *Aschauer, D., 1989. Is public expenditure productive?* In: Journal of Monetary Economics, 23, pp. 177-200.

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.

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.

Bellmann, L. and Brussig, M. (1998). *Ausmaß und Ursache der Produktivitätslücke von ost- und westdeutschen Betrieben. Eine Analyse auf der Grundlage des Betriebspanels.* In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 31, pp. 648-660.

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

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.

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.

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

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.

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.

Klodt, H. (2000). *Industrial Policy and the East German Productivity Puzzle.* In: German Economic Review, 1, pp. 315-333.

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

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.

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

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.

Sachverständigenrat (2005). *Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2005) Jahresgutachten 2005/06. Die Chance nutzen Reformen mutig voranbringen.* Wiesbaden.

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

Schmalwasser, O. and Schidlowski, M. (2006). *Kapitalstockrechnung in Deutschland*. In: Wirtschaft und Statistik, 11, pp. 1107-1123.

Sinn, H.-W. (2000). *Germanys Economic Unification-An Assessment After Ten Years*. CESifo, Working Paper Series, 247, Munich.

Smolny, U. (2003). *Smolny, U. (2003), Produktivitätsanpassung in Ostdeutschland. Bestandsaufnahme und Ansatzpunkte einer Erklärung*. In: Jahrbücher für Nationalökonomie und Statistik, 223, pp. 239-254.

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.

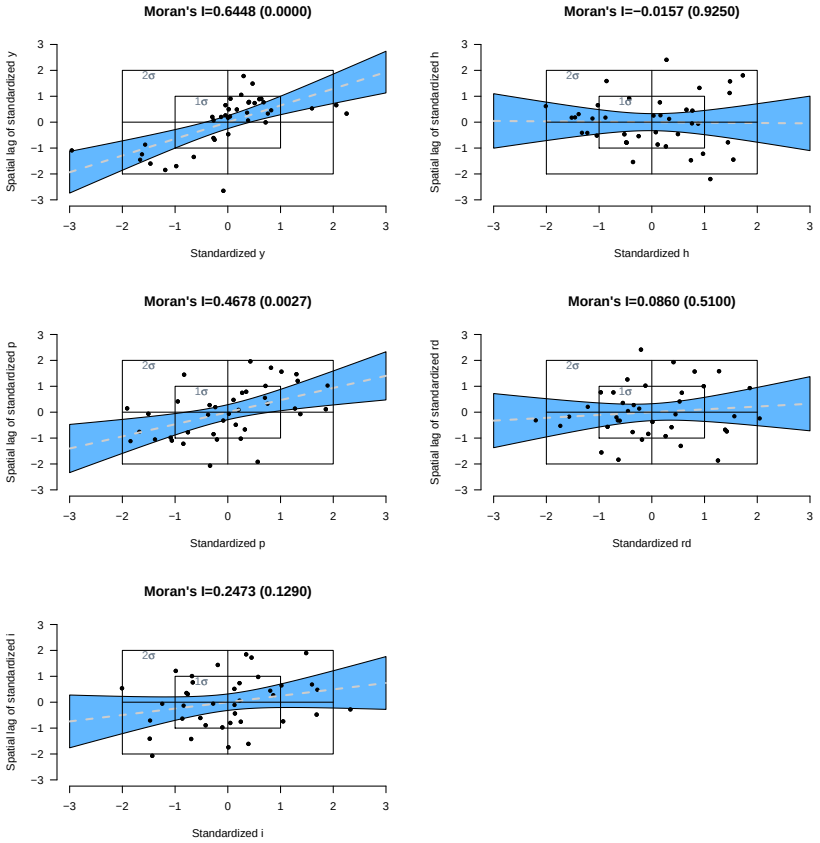

Figure 1: Computation of Moran's $\mathcal{I}$ with corresponding p-values for dependent and independent variable for $r = 1$

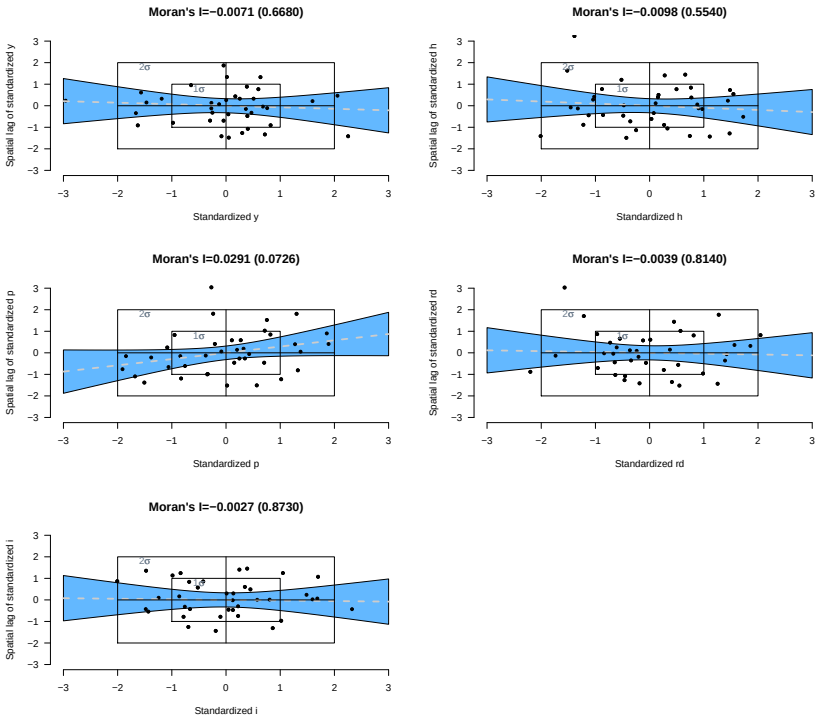

Figure 2: Computation of Moran's $\mathcal{I}$ with corresponding p-values for dependent and independent variable for $r = 2$

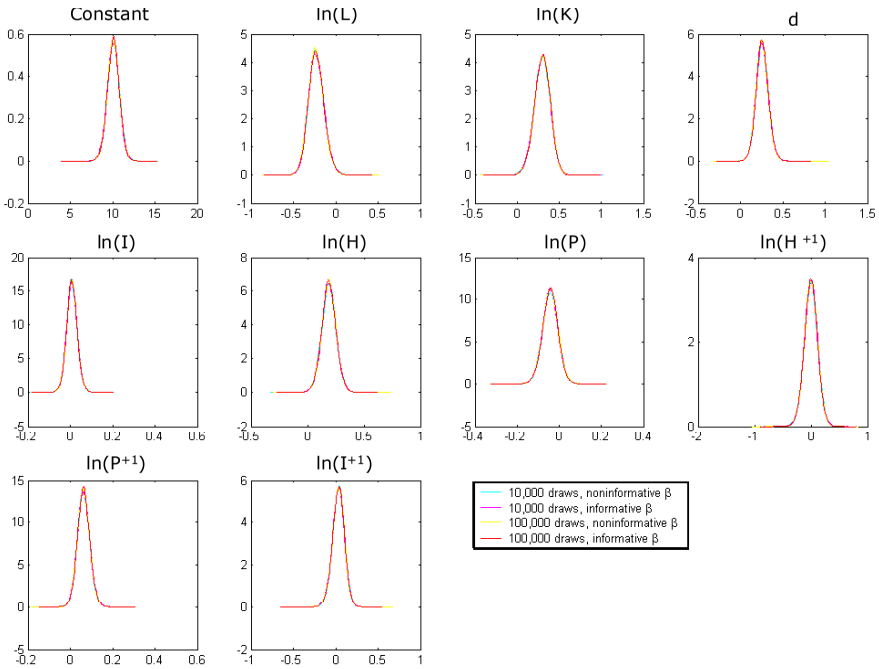

Figure 3: Density plots of estimated β^X and β^{++}

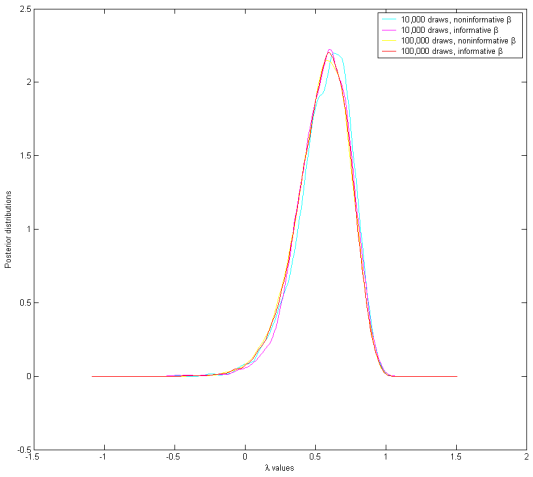

Figure 4: Density plots of estimated λ^1

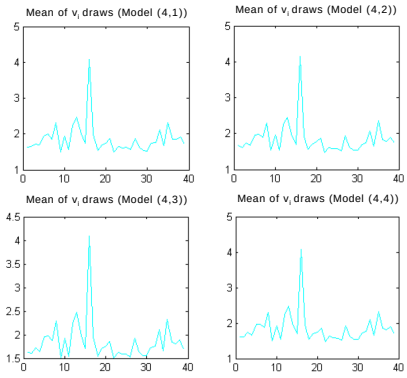

Figure 5: Computation of v_i draws of Ω

Code	German NUTS-2 region	Location
de11	Stuttgart	West
de12	Karlsruhe	West
de13	Freiburg	West
de14	Tübingen	West
de21	Oberbayern	West
de22	Niederbayern	West
de23	Oberpfalz	West
de24	Oberfranken	West
de25	Mittelfranken	West
de26	Unterfranken	West
de27	Schwaben	West
de30	Berlin	West
de41	Brandenburg-Nordost	East
de42	Brandenburg-Südwest	East
de50	Bremen	West
de60	Hamburg	West
de71	Darmstadt	West
de72	Gießen	West
de73	Kassel	West
de80	Mecklenburg-Vorpommern	East
de91	Braunschweig	West
de92	Hannover	West
de93	Lüneburg	West
de94	Weser-Ems	West
dea1	Düsseldorf	West
dea2	Köln	West
dea3	Münster	West
dea4	Detmold	West
dea5	Arnsberg	West
deb1	Koblenz	West
deb2	Trier	West
deb3	Rheinhessen-Pfalz	West
dec0	Saarland	West
ded1	Chemnitz	East
ded2	Dresden	East
ded3	Leipzig	East
dee	Sachsen-Anhalt	East
def0	Schleswig-Holstein	West
deg0	Thüringen	East

Table 1: List of German NUTS-2 regions

	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 2: 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 3: Table of descriptive statistics (II) of variables used for the analysis

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)]

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 5: Comparison of selected models

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)	(0.3879)	(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 6: Results of OLS estimation for German NUTS-2 regions

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)	(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 7: Estimation results for German NUTS-2 regions

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 (i.i.d.)	RNE (i.i.d.)	—	—
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 8: 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 (i.i.d.)	RNE (i.i.d.)	—	—
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 9: 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 (i.i.d.)	RNE (i.i.d.)	—	—
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 10: 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 (i.i.d.)	RNE (i.i.d.)	—	—
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 11: MCMC-convergence summary for model (4,4)

8 Appendix only for review purpose

8.1 Proofs of *OLS* estimations for spatial cross section models

Proposition: An OLS estimation of a spatial lag model would yield inconsistent and thus biased estimators. An OLS estimation of a spatial error model would yield inefficient but unbiased OLS estimators. An OLS regression of a spatial model with exogenous spatial lagged variables is unbiased but only $BLUE$ if spatial homogeneity is assumed. □

Let us start with

$$y = \tilde{y}P + X^{+++}\beta^{+++} + \tilde{\epsilon}\Lambda + \kappa, \tag{44}$$

or with the familiar notation from expression (8):

$$y = \tilde{y}P + X\beta^{X} + \tilde{X}^{++}\beta^{++} + \tilde{\epsilon}\Lambda + \kappa \tag{45}$$

with
$X^{+++} = [X, \tilde{X}^{++}]$ and $\beta^{+++} = [\beta^{X}, \beta^{++}]$. Labeling parts of (8) with I, II, III, IV, this yields:

$$y = \underbrace{\tilde{y}P}_{(I)} + \underbrace{X\beta^{X}}_{(II)} + \underbrace{\tilde{X}^{++}\beta^{++}}_{(III)} + \underbrace{\tilde{\epsilon}\Lambda}_{(IV)} + \kappa. \tag{46}$$

1. Assume, that I=II=IV=0.

   This yields

   $$y = \tilde{X}^{++}\beta^{++} + \kappa, \tag{47}$$

   with $\kappa \sim (0, \sigma^2\Omega)$ with $\sigma^2\Omega \neq \sigma^2 I$. From equation (47) we can obtain an OLS estimator $b^{++} = (\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}\tilde{X}^{++\prime}y$. This estimator is unbiased if $E[\tilde{X}^{++\prime}\kappa] = 0$ because:

   $$E[b^{++}] = \beta^{++} + E[(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}\tilde{X}^{++\prime}\kappa] = \beta^{++}. \tag{48}$$

The estimated variance covariance matrix of $V[b^{++}]$ is

$$V[b^{++}] = E[(b^{++} - \beta^{++})(b^{++} - \beta^{++})'] = \tag{49}$$

$$= E[(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}\tilde{X}^{++\prime}\kappa\kappa'\tilde{X}^{++}(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}], \tag{50}$$

or

$$V[b^{++}] = [(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}\tilde{X}^{++\prime}\Sigma\tilde{X}^{++}(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}] \neq \sigma^2(\tilde{X}^{++\prime}\tilde{X}^{++})^{-1}, \tag{51}$$

with $\Sigma = \sigma^2\Omega$. Thus, *OLS* is unbiased but only *BLUE* if $\sigma^2\Omega = \sigma^2 I$, thus if spatial homogeneity is assumed.

2. Assume, that III=IV=0.

   This yields

$$y = \tilde{y}P + X\beta^X + \kappa, \tag{52}$$

   $\kappa \sim (0, \sigma^2\Omega)$ with $\sigma^2\Omega \neq \sigma^2 I$. An *OLS* estimation of equation (52) would yield, under neglecting an element $r \in R$ of $\tilde{y}P$:

$$b^X = (X'X)^{-1}X'y. \tag{53}$$

   After inserting y this yields the estimator b^X of β^X:

$$b^X = (X'X)^{-1}X'(\rho^r W^{+r} y + X\beta^X + \kappa). \tag{54}$$

   Taking the expectation of equation (54) this yields

$$E[b^X] = E[(X'X)^{-1}X'(\rho^r W^{+r} y) + \beta^X] \neq \beta^X. \tag{55}$$

   Thus, the bias can be expressed as

$$E[b^X - \beta^X] = E[(X'X)^{-1}X'(\rho^r W^{+r} y)] = \rho^r \beta^l. \tag{56}$$

   The expression can be interpreted as ρ^r times the regression of X against $(\rho^r W^{+r} y)$ with the corresponding β^l which is equal to the expected value of the regression $(W^{+r} y)$ on X. Hence, *OLS* is biased if only one component of $\tilde{y}P$ is neglected.

3. Assume, that II=III=IV=0

   In this case we obtain

$$y = \tilde{y}P + \kappa, \tag{57}$$

   $\kappa \sim (0, \sigma^2\Omega)$ with $\sigma^2\Omega \neq \sigma^2 I$. An *OLS*-estimator of one element ρ^r of P would yield:

$$\hat{\rho^r} = [(W^{+r}y)'(W^{+r}y)]^{-1}(W^{+r}y)'y, \tag{58}$$

   or inserting expression (57) in equation (58)

$$\hat{\rho^r} = \rho^r + [(W^{+r}y)'(W^{+r}y)]^{-1}(W^{+r}y)'\kappa. \tag{59}$$

   The estimator $\hat{\rho}$ is not consistent because

$$\hat{\rho} \xrightarrow{p} \rho + \left(\frac{1}{N}(W^{+r}y)'W^{+r}y\right)^{-1}\left(\frac{1}{N}(W^{+r})'\kappa\right),$$

$$\hat{\rho} \xrightarrow{p} \rho + S^{-1}_{(W^{+r}y)'W^{+r}y}\left(\frac{1}{N}\kappa'W^{+r}(I-\rho^r W^{+r})^{-1}\kappa\right). \tag{60}$$

   The expression $\left(\frac{1}{N}(W^{+r}y)'W^{+r}y\right)$ converges to a regular and finite matrix $S_{(W^{+r}y)'W^{+r}y}$. The second term $\left(\frac{1}{N}(W^{+r})'\kappa\right)$ however converges to an expression which is quadratic in the errors, unless $\rho^r = 0$. Hence, estimating a spatial lag parameter ρ^r via *OLS* is biased and inconsistent.

4. Assume, that I=III=0

   Now it results

$$y = X\beta^X + \epsilon, \tag{61}$$

   with $\epsilon = \epsilon\Lambda + \kappa$ and $\kappa \sim (0, \sigma^2\Omega)$ with $\sigma^2\Omega = \sigma^2 I$. An *OLS* estimation of β^X would be unbiased, because

$$E[b^X - \beta^X] = E[(X'X)^{-1}X'\kappa] = 0, \tag{62}$$

   but β^X is not efficient because for a given λ^r from Λ for the estimated variance covariance matrix it is obtained:

$$V[b^{++}] = [(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}X], \tag{63}$$

which is

$$V[b^{++}] = \sigma^2[(X'X)^{-1}X'[(I - \lambda^r W^{+r})'(I - \lambda^r W^{+r})]^{-1}X(X'X)^{-1}X]. \quad (64)$$

In consequence, the *OLS* estimator of b^X is unbiased but inefficient for a given λ^r from Λ. ■

8.2 Derivation of the log-likelihood function for the $SAC(1,1)$ model

The log-likelihood function for the $SAC(1,1)$ can be derived as follows:[42]

$$y = \rho^1 W^{+1} y + X^{+++}\beta^{+++} + \epsilon, \quad (65)$$

with

$$\epsilon = \lambda^1 W^{+1}\epsilon + \kappa, \quad (66)$$

with $\kappa \sim (0, \sigma^2 I)$. Next define

$$\kappa = \frac{1}{\sigma}(I - \lambda^1 W^{+1})^{-1}[(I - \rho^1 W^{+1})y - X^{+++}\beta^{+++}] \quad (67)$$

with $\kappa \sim N(0, \sigma^2 I)$. The corresponding determinant of the Jacobian $\mathcal{J} \equiv det\frac{\partial \kappa}{\partial y}$ can be rewritten as

$$\mathcal{J} \equiv det\frac{\partial \kappa}{\partial y} = |\frac{1}{\sigma}[I - \lambda^1 W^{+1}]\,||\,[I - \rho^1 W^{+1}]|. \quad (68)$$

Employing the fact that $\kappa \sim N(0, \sigma^2 I)$ we can write the log-likelihood for the joint distribution as

$$\ln \mathcal{L} = -\frac{N}{2}\ln 2\pi - \frac{N}{2}(\sigma^2) + \ln|\tilde{N}| + \ln|[I - \rho^1 W^{+1}]| - \frac{1}{2}\kappa'\kappa, \quad (69)$$

If $\rho^1 = 0$ equation (23) results.

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

8.3 Derivation of formula (78)

Using PIM, the capital stock K_t can be computed as

$$K_t = \iota_0 I_t + \iota_1 I_{t-1} + ... + \iota_T I_{t-T}, \tag{70}$$

with I_t as investment in new capital K_t. It is common to set $\iota_0 = 1$ and $\iota_t = (1-\delta)^t$ for $t > 0$. Using a Koyck transformation, from equation (70) we obtain:

$$K_t = I_t + (1-\delta)K_{t-1}, \tag{71}$$

with $\delta = \frac{\iota_{T-1} - \iota_t}{\iota_{T-1}}$. To obtain expression (78) we assume that investment I_t in stock of capital K_t is growing from $t = 0$ with constant rate ζ. Therefore we can write:

$$I_t = (1+\zeta)I_{t-1} = (1+\zeta)(1+\zeta)I_{t-2} = ... = (1+\zeta)^{\infty^+} I_{t-\infty^+}. \tag{72}$$

Further it is assumed that devaluation of capital K_t follows a geometric series:

$$K_t = I_t + (1-\delta)I_{t-1} + (1-\delta)^2 I_{t-2} + ... + (1-\delta)^{\infty^+} I_{t-\infty^+}. \tag{73}$$

Using (72) and (73) leads to

$$K_t = I_t + \left[\frac{1-\delta}{1+\zeta}\right] I_t + \left(\frac{1-\delta}{1+\zeta}\right)^2 I_t + ... + \left(\frac{1-\delta}{1+\zeta}\right)^{\infty^+} I_t = I_t \sum_{\kappa=0}^{\infty^+} \left(\frac{1-\delta}{1+\zeta}\right)^{\kappa}. \tag{74}$$

Rearranging equation 74 by writing:

$$\left[1 - \left(\frac{1-\delta}{1+\zeta}\right)\right] K_t = I_t \left[1 - \left(\frac{1-\delta}{1+\zeta}\right)^{\kappa+1}\right] \tag{75}$$

and letting $\kappa \to \infty$ leads to

$$K_t = I_t \frac{1}{\left[1 - \left(\frac{1-\delta}{1+\zeta}\right)\right]}, \tag{76}$$

because of $\left(\frac{1-\delta}{1+\zeta}\right) < 1$. Noting, that $I_{t+1} = I_t(1+\zeta)$ yields

$$K_t = \frac{I_{t+1}}{1+\zeta} \frac{1}{\left[1 - \left(\frac{1-\delta}{1+\zeta}\right)\right]}. \tag{77}$$

8.4 Data and variables

This section gives a description of the data which have been used in this study. 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[43]. This problem is boosted by the empirical fact of suburbanization, which has increasingly appeared in the last years[44]. 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 of less importance or even cancelled 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 included in the 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 Länder"" as well as from the INKAR-database CD-Rom published by the "Bundesamt für Bauwesen und Raumordnung".

In detail, the following variables are specified:

1. ***Output (Y)*** is approximated with Gross Value Added. The data are published

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

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

annually on the CD-Rom "Statistik regional" by the "Statistische Ämter des Bundes und der Länder" and have been stated in Mio. Euros.

2. ***Human capital (H)*** is measured as the percentage of the employees on NUTS-2 level, subject to social insurance contribution, who obtained a high level degree, such as an university, a polytechnical or a technical college degree. With the exception of Sachsen-Anhalt, the data stem from the CD-Rom "Statistik regional" edited by the "Statistische Ämter des Bundes und der Länder"[45]. To exclude the above mentioned commuter problem, the data correspond to the activity area, not to the place of residence of the employees. Naturally, this assumption implies that added value is created at the activity area. Unfortunately, the data do not exhibit the desirable attribute that they are restricted to the employed human capital in production sector. Hence, as mentioned by Keilbach (2000) we have to bear in mind implicit spillovers of employed human capital from the non-producing sectors.

3. ***Labour (L)*** is measured as number of employees in thousands on NUTS-2 level subject to social insurance contribution less human capital, defined above. Data have been obtained from the CD-Rom "Statistik regional" edited by the "Statistische Ämter des Bundes und der Länder".

4. ***Capital (K)*** stock construction for the regional NUTS-2 manufacturing sector is a serious problem. By mischance, it is not possible to hark back to regionally disaggregated stock of capital data for NUTS-2 regions from the official statistic suppliers. Only for the German "Bundesländer" the "Arbeitskreis "Volkswirtschaftliche Gesamtrechnungen der Länder"" offers capital stock data. Naturally, on this rather aggregated level, capital stock estimation via the perpetual inventory method (PIM) is rather easy to implement. The fundamental idea of PIM is that different vintages of the stock of capital exhibit different efficiencies being used in the production process. This idea has

[45]The data of Sachsen-Anhalt have been obtained directly from the "Statistisches Landesamt Sachsen-Anhalt".

to be considered when calculating the stock of capital. Therefore, one has first to determine the average durability of an asset. Next, calculation of long term data regarding the annual investments is needed to initialize PIM. Basically, it is common to refer to gross fixed capital formation as a proxy, because PIM is nothing else than computing an average weighted sum of past investments. Long term data are necessary, especially in the case of Germany, because the cumulative investment data have to be corrected using a survival function and depreciation function to obtain an estimation for the capital stock. In the case of Germany the Gamma distribution has to be consulted to obtain a measure for the mortality function from which the depreciation function can be obtained directly[46]. Based on the gamma function, it cannot be ruled out ex ante that the service life of an asset oscillates more than twice of an average service life of an asset. That is exactly the reason why it is strongly recommended to use long investment data[47]. In this way it is possible to calculate the stock of capital K in period t using data of gross fixed capital formation I from the period $t+1$, a depreciation rate on stock of capital δ, obtained from the depreciation function and an average growth rate ζ of gross fixed capital formation. In a more formal manner the following relationship results[48]:

$$K_t = I_t \sum_{\kappa=0}^{\infty} \left(\frac{1-\delta}{1+\zeta}\right)^{\kappa} = \frac{I_{t+1}}{1+\zeta} \frac{1}{\left(1-\left(\frac{1-\delta}{1+\zeta}\right)\right)} = \frac{I_{t+1}}{\zeta+\delta}. \tag{78}$$

As mentioned above, long term data for the gross fixed capital formation are needed to initialize PIM. Unfortunately, long term series of desired data are not available for Germany on NUTS-2 level. EUROSTAT offers data for gross fixed capital formation on NUTS-2 level for German regions only for the years 2002 and 2003[49]. Concerning the above mentioned, it is not reasonable to rely

[46] It is suitable to set the dilation parameter of the mortality function to the value $p = 9$.

[47] The starting date for series of gross fixed capital formation is 1799 for buildings, 1899 for machinery and equipment and 1945 or later for intangible assets. For an deeper introduction of PIM, particularly for Germany, consult Schmalwasser and Schidlowski (2006).

[48] Refer to appendix 2 for a deviation of expression 78.

[49] The Statistische Landesamt Baden-Wurttemberg offers data for the gross fixed capital formation from 1998 onwards online.

on PIM estimating the stock of capital for NUTS-2 regions.

Because of that reason, the estimation of the stock of capital is performed with a method similar to the shift analysis. The basic idea of the shift analysis is to compute a so called structural factor and a location factor. The structural factor should provide information about the capital intensity of branches and furthermore, should give a hint, whether capital intensive branches are over- or underrepresented in a specific region. Assume for the moment[50] that we have $i = \{1, 2, ..., I\}$ branches and $h = \{1, 2, ..., N\}$ NUTS-2 regions in each of the $j = \{1, 2, ..., M\}$ states. It is worth mentioning that a single NUTS-2 region can represent an own state[51]. Due to the fact that we do not analyse specific branches, we can set $i = 1$.

Hence, for the structural factor SF for region h in a formal manner we can write:

$$SF_t^h = \frac{\sum_i g_{t-1,i} I_{t,i}^M}{\sum_i g_{t-1,i} I_{t-1,i}^M} / \frac{\sum_i I_{t,i}^M}{\sum_i I_{t-1,i}^M}, \tag{79}$$

where $I_{t,i}^M$ stands for the gross fixed capital formation in the state M in year t, $I_{t,i}^N$ represents the gross fixed capital formation in the NUTS-2 region N in year t and $g_{t,i} \equiv \frac{I_{t,i}^N}{I_{t,i}^M}$ is the weight for region h. Of course, $\sum_h g_{t,i} = 1$ and $\sum_j \sum_h g_{t,i}$ must equal the number of the states M. In the case of Germany $M = 16$.

Instead of the structural factor, the location factor assumes implicitly, that a specific region, which can be characterized by high investments in the past, must exhibit a high stock of capital relative to other regions in the present. For the location factor LOF one can write:

$$LOF_t^h = \frac{\sum_i g_{t,i} I_{t,i}^M}{\sum_i g_{t-1,i} I_{t,i}^M}. \tag{80}$$

[50] Please bear in mind that the national form is only valid for presentation of shift analysis.

[51] This is true for Hamburg, Bremen, Berlin, Mecklenburg-Vorpommern, Schleswig-Holstein, Saarland, Thüringen and Sachsen-Anhalt in the case of German "Bundesländer".

Multiplying the structural factor with the location factor one obtains the regional factor. More formally spoken, we obtain the regional factor RF:

$$RF_t^h = \frac{\sum_i I_{t,i}^N}{\sum_i I_{t-1,i}^N} / \frac{\sum_i I_{t,i}^M}{\sum_i I_{t-1,i}^M} = \frac{\sum_i g_{t,i} I_{t,i}^M}{\sum_i g_{t-1,i} I_{t-1,i}^M} / \frac{\sum_i I_{t,i}^M}{\sum_i I_{t-1,i}^M}. \tag{81}$$

A value greater than one for the regional factor implies that a specific region has grown faster than the average, a value less than one means that a specific region has grown less than the average.

To calculate the weights for RF_t^h and LOF_t^h for every region h we have to consult data of gross fixed capital formation for 2003. After calculating the region specific weights, the regional capital stocks for the German "Bundesländer" are weighted with these. In this way, we have estimated NUTS-2 specific stocks of capital in Mio. Euros have been computed.

5. ***R&D (R&D)*** effort is expressed as the total *R&D* expenditure (GERD). The expenditures include the business enterprise sector, the government sector, the higher education sector as well as the private non-profit sector. Data have been expressed in Mio. Euros and have been provided by EUROSTAT. Obviously, relying on this data, we cannot exclude spillovers from the non-producing sector to the producing sector. As mentioned by Keilbach (2000) this effect should be neglectable. Although it would be reasonable on the first sight, we should not use *R&D* employees as a proxy for *R&D*, because it is justified to assume that within the *R&D* sector, more than in the manufacturing sector, the majority of offered jobs requires a high skilled labour force, a subset of human capital, defined above.

6. ***Patent (P)*** applications to the European Patent Office (EPO) by priority year at the regional level have been gathered from EUROSTAT. The priority starts after the year filing the patent application. Data are expressed as total number of patent applications in a specific *NUTS-2* region.

7. ***Infrastructure (I)***: Since Aschauer (1989) there has been a intensively leading debate about how to measure infrastructure and what effects public in-

frastructure has on output growth using a production function approach. In general, the studies can be grouped in national level studies and regional or state level studies. One traditional approach is to use information about undeveloped areas serving for streets, railways or airways and traffic on waterways Keilbach (2000). Additionally, other factors, such as political interest, friendship ties, basic trust and quality of life etc. should flow into the regression context. Regrettably, these data are not available on NUTS-2 regions. Therefore, for this study on has to refer to data on highway density per squared kilometre published by EUROSTAT.

8. ***Density (DEN)*** is measured as inhabitants per square kilometre. Data for the average population for 2003 per NUTS-2 area as well as details for the NUTS-2-areas in square kilometre have been obtained from the CD-Rom "Statistik regional".

9. ***Dummy***: The dummy covers East-Western productivity differences. It is defined as follows:

$$d = \begin{cases} 1 & \text{if region } i \text{ belongs to the group of West German NUTS-2 regions} \\ 0 & \text{if region } i \text{ belongs to the group of East German NUTS-2 regions} \end{cases}.$$

It is reasonable to include the dummy, because a bulk of papers have found empirical evidence that a significant difference regarding the capital intensity still exists between East and West German region. After initial continuous progress concerning the productivity of East German regions right after the German reunification and an observed stagnation in the years 1996 and 1997 this gap seems to widen again in recent years[52]. For instance Smolny (2003) has found that East German capital intensity is 80% of corresponding West German capital intensity.

[52]For the convergence debate of East German regions refer to the empirical based analysis of Bellmann and Brussig (1998), Almus and Czarnitzki (2003), Klodt (2000), Smolny (2003), Sinn (2000) and Sachverständigenrat (2005).

8.5 "Metropolis-Within-Gibbs"

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 expression (37) 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^{+++}.
5. *Calculate equation (40) 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 equation (39) 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.*

8.6 MCMC convergence checks

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".

8.6.1 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.

8.6.2 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.

8.6.3 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.

8.6.4 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[53].

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