Model uncertainty in matrix exponential spatial growth regression models

Fischer, Manfred M.; Piribauer, Philipp

DOI:
10.57938/6c9fb19e-e49f-4039-8f3e-ea33bdb4b330

Published: 01/10/2013

Document Version:
Publisher's PDF, also known as Version of record

Department of Economics
Working Paper No. 158

Model uncertainty in matrix exponential spatial growth regression models

Philipp Piribauer
Manfred M. Fischer

October 2013

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Model uncertainty in matrix exponential spatial growth regression models

Philipp Piribauer*† Manfred M. Fischer†

Abstract

This paper considers the problem of model uncertainty associated with variable selection and specification of the spatial weight matrix in spatial growth regression models in general and growth regression models based on the matrix exponential spatial specification in particular. A natural solution, supported by formal probabilistic reasoning, is the use of Bayesian model averaging which assigns probabilities on the model space and deals with model uncertainty by mixing over models, using the posterior model probabilities as weights. This paper proposes to adopt Bayesian information criterion model weights since they have computational advantages over fully Bayesian model weights. The approach is illustrated for both identifying model covariates and unveiling spatial structures present in pan-European growth data.

Keywords: model comparison, model uncertainty, spatial Durbin matrix exponential growth models, spatial weight structures, European regions

JEL: C11, C21, C52, O47, O52, R11

*Department of Economics, Vienna University of Economics and Business (WU). E-mail: philipp.piribauer@wu.ac.at

†Department of Socioeconomics, Vienna University of Economics and Business (WU). E-mail: manfred.fischer@wu.ac.at

Introduction

Recent years have seen a virtual explosion in the application of cross-sectional spatial growth regression models. While undoubtedly considerable progress has been made, most applications ignore model uncertainty about the process describing regional economic growth. Model uncertainty arises from two sources: (i) the spatial weight or connectivity structure assigned to regions that form the observational basis of spatial data samples, and (ii) specific explanatory variables included. The first source of model uncertainty is unique to spatial regression modeling since conventional regression models assume independence between sample observations (regions). The hallmark of spatial growth regression models is the spatial weight matrix that distinguishes these from non-spatial growth regressions. The specification of this matrix is typically constructed by means of geographic criteria, such as contiguity (sharing a common border) or distance, including nearest neighbor distance. Uncertainty regarding the spatial weight matrix has long been recognized by practitioners who typically check whether estimates and inference are similar when alternative spatial weight structures are used.

The second source of uncertainty arises in conventional as well as spatial growth regression models since growth theories are not sufficiently explicit about which specific factors underlie the data-generating process for growth regressions. Hence, researchers are faced with a dilemma regarding the large number of potential regressors. There is a trade-off between arbitrary selection of a small subset of variables which may give rise to omitted variables bias, and the introduction of a large set of variables that will tend to increase the dispersion of the estimated coefficients, making it difficult to identify important factors.

Spatial growth regression models produce estimates and inference that are conditional on both the particular spatial weight matrix used to specify which observational units (regions) are linked and the set of explanatory variables employed. Selection of an appropriate spatial weight matrix and explanatory variables are central to the analysis of growth empirics. Competing specifications are usually non-nested alternatives so that conventional statistical procedures such as likelihood ratio tests are inappropriate (LeSage and Fischer 2008).

Model averaging provides a formal approach that can be used to incorporate model uncertainty in spatial regression models which arises from selecting both the spatial weight matrix and the explanatory variables when making inferences about model parameters. Instead of selecting a single model, this approach proposes to average estimates across different models. Bayesian model averaging represents one powerful approach to making parameter inference unconditional on model specification issues. There is a great deal of literature on Bayesian model averaging for non-spatial regression models. The approach involves averaging over models based on different sets of explanatory variables.

Work by Fernández et al. (2001a) considers cases where the number of possible models is sufficiently large so that calculation of posterior probabilities for all models is difficult or infeasible. A Markov chain Monte Carlo model comparison methodology proposed by Madigan and York (1995) has gained popularity in the mathematical statistics and econometrics literature, which eliminates the need to consider all potential models. See, for example, Fernández et al. (2001a) or Koop (2003). An extension to spatial autoregressive regression models is provided by LeSage and Parent (2007). LeSage and Fischer (2008) include simultaneous comparison of models based on both alternative explanatory variables and spatial weight matrices, albeit concentrating on the class of k-nearest neighbor spatial weight matrices. From a technical point of view, the authors suggest to use numerical integration techniques to obtain posterior model probabilities for specifications with different k-nearest spatial weight matrices which are then used to obtain Bayesian model averaged estimates. The computational costs of this procedure, however, makes it an impractical choice for a large set of alternative spatial weight matrices. Our methodology improves on LeSage and Fischer (2008) by adopting Bayesian information criterion (BIC) posterior model weights to overcome such computational costs, and thus allowing for the consideration of a wide range of weight matrices as potential spatial structures underlying the spillovers in the data.

We focus our discussion of model averaging on a matrix exponential spatial specification (MESS) of the spatial Durbin growth regression model that replaces the spatial autoregressive process with a matrix exponential spatial transformation. This specification has a computational advantage over the conventional spatial autoregressive based spatial Durbin model, since it eliminates the need to calculate the log-determinant when producing maximum likelihood estimates (LeSage and Pace 2007). We work with three types of spatial weight matrices that are widely used in applied practice: the binary contiguity-based matrix that considers regions to be neighbors if they share a common border, distance-band spatial weight matrices with critical distances and k-nearest neighbor spatial weight matrices. Distance between regions is measured in terms of geodesic distances and travel times.

The rest of the paper is organized as follows. The next section outlines the matrix exponential spatial Durbin model, and the third section the associated likelihood, and a closed-form for maximum likelihood estimates. The following section presents the model averaging approach that uses a BIC approximation to the marginal likelihood for the models of interest along with a Markov chain Monte Carlo model composition (MC^3) method to reduce the computational burden. The fifth section applies the methodology to both identify model covariates and unveil spatial structures present in a dataset of 273 European regions. The final section concludes.

Matrix exponential spatial growth models

We consider cross-sectional spatial growth regression models where the dependent variable undergoes a linear transformation as in Eq. (1):

$$\boldsymbol{S}\boldsymbol{y} = \beta_0 \boldsymbol{\iota}_N + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \tag{1}$$

$\boldsymbol{y}$ is an $N \times 1$ vector of spatial observations on the dependent variable representing growth rates. $\boldsymbol{X}$ is an $N \times Q$ matrix of observations on the Q independent variables representing growth determinants. Each of these observations on the dependent and explanatory variables comes from regions in space. The N-element vector $\boldsymbol{\varepsilon}$ is distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_N\sigma^2)$. $\boldsymbol{\beta}$ is a $Q \times 1$ vector of parameters associated with the explanatory variables, $\boldsymbol{\iota}_N$ is a vector of ones and β_0 is the associated intercept parameter. The matrix $\boldsymbol{W}$ is an $N \times N$ spatial weight matrix that contains non-zero elements $[\boldsymbol{W}]_{ij}$ if observations j and i are neighboring regions $(i, j = 1, \ldots, N)$ and zero otherwise. Characteristically, the matrix $\boldsymbol{W}$ is row-stochastic, so that the $N \times Q$ spatial lag matrix $\boldsymbol{WX}$ contains values constructed from an average of neighboring regions. The $Q \times 1$ parameter vector $\boldsymbol{\gamma}$ measures the marginal impact of the explanatory variables from neighboring observations (regions) on the dependent variable $\boldsymbol{y}$.

$\boldsymbol{S}$ is an N-by-N non-singular matrix of constants that may depend on an unknown real scalar parameter α. We focus on the transformation $\boldsymbol{S}$ used to model spatial dependence among the (regional) elements of the vector $\boldsymbol{y}$. A prominent member of the family of spatial growth regression models given by Eq. (1) is the conventional spatial Durbin model that arises when setting $\boldsymbol{S} = (\boldsymbol{I}_N - \rho\boldsymbol{W})$, where ρ is a scalar parameter reflecting the magnitude of spatial dependence.

The focus on this paper is on the matrix exponential as a specification for $\boldsymbol{S}$ defined in Eq. (2):

$$\begin{aligned}\boldsymbol{S}(\alpha) &= \exp(\alpha\boldsymbol{W}) = \sum_{t=0}^{\infty} \frac{\alpha^t}{t!}\boldsymbol{W}^t \\ &= \boldsymbol{I}_N + \alpha\boldsymbol{W} + \frac{\alpha^2}{2!}\boldsymbol{W}^2 + \frac{\alpha^3}{3!}\boldsymbol{W}^3 + \ldots \end{aligned} \tag{2}$$

where α is a real scalar parameter with $-\infty < \alpha < \infty$. $\boldsymbol{S}$ is a linear combination of row-stochastic matrices and hence is proportional to a row-stochastic matrix, since products of row-stochastic matrices are row-stochastic.

The matrix exponential spatial specification (MESS), introduced by LeSage and Pace (2007), replaces the conventional geometric decay of influence from higher-order neighboring relations by the spatial autroregressive process in conventional spatial Durbin model relationships with an exponential pattern of decay in influence from higher-order neighboring relationships. Spatial growth regression models of type (1) with a matrix exponential specification of $\boldsymbol{S}$ may be termed spatial Durbin matrix exponential models. For implementation purposes, we truncate the infinite expansion shown in Eq. (2) to R terms creating an approximation to $\boldsymbol{S} = \exp(\alpha\boldsymbol{W})$.

LeSage and Pace (2007, 2009) discussed several of the salient properties of MESS models, some of which may be enumerated as follows: First, $\boldsymbol{S}(\alpha)$ is non-singular (*Property* 1). Second, $\boldsymbol{S}(\alpha)^{-1} = \left[\exp(\alpha\boldsymbol{W})\right]^{-1} = \exp(-\alpha\boldsymbol{W})$ (*Property* 2) and third, $|\exp(\alpha\boldsymbol{W})| = \exp\left[\mathrm{tr}(\alpha\boldsymbol{W})\right]$ (*Property* 3). *Property* 1 guarantees a positive definite covariance matrix and thus avoids the need to restrict the parameter space, or to carry out tests for positive definiteness during parameter estimation. *Property* 2 leads to a simple mathematical inversion of the matrix exponential. *Property* 3 implies that the log-likelihood of the MESS model does not contain a troublesome log-determinant of an $N \times N$ matrix unlike for the case of the conventional spatial Durbin model. Finally, it is worth noting that there are approximate relations

$\alpha \approx \log(1-\rho)$ or $\rho \approx 1-\exp(\alpha)$, and this relation between α and ρ facilitates interpretations of parameter estimates for α allowing them to be viewed in terms of the parameter ρ from the conventional spatial Durbin model.

Maximum likelihood estimation

The log-likelihood for the spatial Durbin matrix exponential spatial model specification given by Eqs. (1)-(2) takes the form as in Eq. (3):

$$\begin{aligned}
\log \mathcal{L}(\alpha, \boldsymbol{\delta}, \sigma^2|\boldsymbol{y}) &= -\frac{N}{2}\left[\log(\sigma^2)+\log(2\pi)\right]+\log|\boldsymbol{S}(\alpha)| \\
&\quad -\frac{1}{2\sigma^2}\log\left[\boldsymbol{y}(\boldsymbol{\delta})'\boldsymbol{S}(\alpha)'\boldsymbol{M}\boldsymbol{S}(\alpha)\boldsymbol{y}(\boldsymbol{\delta})\right] && (3)\\
\boldsymbol{M} &= \boldsymbol{I}_N - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}' && (4)\\
\boldsymbol{Z} &= \begin{bmatrix}\boldsymbol{\iota}_N & \boldsymbol{X} & \boldsymbol{W}\boldsymbol{X}\end{bmatrix} && (5)\\
\boldsymbol{\delta} &= \begin{bmatrix}\beta_0 & \boldsymbol{\beta} & \boldsymbol{\gamma}\end{bmatrix}'. && (6)
\end{aligned}$$

The term $|\boldsymbol{S}(\alpha)|$ in Eq. (3) is the Jacobian of the transformation for $\boldsymbol{y}$ to $\boldsymbol{S}(\alpha)\boldsymbol{y}$. *Property* 3 allows to greatly simplify the log-likelihood, using $\text{tr}(\boldsymbol{W})=0$, $|\exp(\alpha\boldsymbol{W})| = \exp[\text{tr}(\alpha\boldsymbol{W})]=1$.

Concentrating out the noise variance parameter σ^2 of the log-likelihood yields a concentrated log-likelihood taking the form:

$$\log \mathcal{L}(\alpha|\boldsymbol{y}) = \kappa - \frac{N}{2}\log\left[\boldsymbol{y}'\boldsymbol{S}(\alpha)'\boldsymbol{M}\boldsymbol{S}(\alpha)\boldsymbol{y}\right] \quad (7)$$

where κ is a scalar constant that does not depend on the parameter α. Hence, maximizing the log-likelihood is equivalent to minimizing $\left[\boldsymbol{y}'\boldsymbol{S}(\alpha)'\boldsymbol{M}\boldsymbol{S}(\alpha)\boldsymbol{y}\right]$, the sum of squared errors. Given an estimator for α, the remaining parameters $\boldsymbol{\delta}$ and σ^2 can be found using simple expressions involving only sample data and an estimator for α.

We follow LeSage and Pace (2007) to show how to find a closed-form solution for the model and rely on the truncated approximation to implement the infinite series definition of $\boldsymbol{S}(\alpha)$ using a power series expansion containing R terms. For sufficiently large R, the truncated terms of the series may be made as small as desired. We do not need to compute $\boldsymbol{S}$ separately since $\boldsymbol{S}$ always appears in conjunction with $\boldsymbol{y}$. Let $\boldsymbol{Y}$ be the $N \times R$ matrix comprised of powers of $\boldsymbol{W}$ times $\boldsymbol{y}$:

$$\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{W}^0\boldsymbol{y} & \boldsymbol{W}^1\boldsymbol{y} & \boldsymbol{W}^2\boldsymbol{y} & \cdots & \boldsymbol{W}^{R-1}\boldsymbol{y}\end{bmatrix}. \quad (8)$$

Note that contiguity-based or k-nearest neighbor spatial weight matrices are typically sparse. Such sparsity leads to a dramatic decline in the number of operations needed to compute $\boldsymbol{S}(\alpha)\boldsymbol{y}$.

To solve for parameter estimates of the model, we define the $R \times R$ diagonal matrix $\boldsymbol{A}$ that contains some of the coefficients from the power series as shown in Eq. (9):

$$\boldsymbol{A} = \begin{pmatrix} \frac{1}{0!} & 0 & \cdots & 0 \\ 0 & \frac{1}{1!} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{(R-1)!} \end{pmatrix}. \tag{9}$$

Moreover, we define the R-element column vector $\boldsymbol{b}(\alpha)$ shown in Eq. (10) that contains powers of the scalar real parameter α, $|\alpha| < \infty$:

$$\boldsymbol{b}(\alpha) = \begin{bmatrix}\alpha^0 & \alpha^1 & \alpha^2 & \cdots & \alpha^{R-1}\end{bmatrix}'. \tag{10}$$

Using Eqs. (8)-(10) we can rewrite $\boldsymbol{S}(\alpha)\boldsymbol{y}$ as shown in Eq. (11), where the approximate equality arises from the truncation of the series expansion to R terms:

$$\boldsymbol{S}(\alpha)\boldsymbol{y} \approx \boldsymbol{Y}\boldsymbol{A}\boldsymbol{b}(\alpha). \tag{11}$$

Premultiplying $\boldsymbol{S}(\alpha)\boldsymbol{y}$ by $\boldsymbol{M}$ yields the $N \times 1$ vector of least squares residuals $\boldsymbol{e}(\alpha)$ for a given value α, allowing us to express the overall sum-of-squared errors as in Eq. (12):

$$\begin{aligned} \boldsymbol{e}(\alpha)'\boldsymbol{e}(\alpha) &= \boldsymbol{b}(\alpha)'\boldsymbol{A}(\boldsymbol{Y}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{Y})\boldsymbol{A}\boldsymbol{b}(\alpha) \\ &= \boldsymbol{b}(\alpha)'\boldsymbol{H}\boldsymbol{b}(\alpha) \end{aligned} \tag{12}$$

where $\boldsymbol{H} = \boldsymbol{A}(\boldsymbol{Y}'\boldsymbol{M}\boldsymbol{Y})\boldsymbol{A}$. This allows us to rewrite $\boldsymbol{b}(\alpha)'\boldsymbol{H}\boldsymbol{b}(\alpha)$ as the $(2R-2)$ degree polynomial $P(\alpha)$:

$$P(\alpha) = \sum_{r=1}^{2R-1} b_r(\alpha)\alpha^{r-1} = \boldsymbol{b}(\alpha)'\boldsymbol{H}\boldsymbol{b}(\alpha). \tag{13}$$

$P(\alpha)$ is a polynomial in α and has a closed-form solution that provides an optimal value of α and also the second derivative at this optimal value. With regard to second-order conditions, LeSage and Pace (2007) show that the optimal α is unique for MESS.

The model averaging approach

We are interested in averaging over matrix exponential spatial growth regression models that differ in two respects, namely the spatial weight matrix specification and the set of explanatory variables. With Q potential growth determinants and J potential spatial weight matrices, the cardinality of the model space $\mathcal{M}$ is (IJ) with $I = 2^{2Q}$. A particular model $M_{ij} \in \mathcal{M}$ $(i = 1, \ldots, I; j = 1, \ldots, J)$ is characterized by its parameter vector $\boldsymbol{\theta} = [\boldsymbol{\delta}, \alpha, \sigma^2]$. In the Bayesian model averaging framework[1], the posterior for the parameters $\boldsymbol{\theta}$, calculated using M_{ij}, is written as:

[1] For an introduction to Bayesian model averaging, see, for example, Koop (2003).

$$p(\boldsymbol{\theta}|M_{ij}, \mathcal{D}) = \frac{f(\mathcal{D}|\boldsymbol{\theta}, M_{ij})\pi(\boldsymbol{\theta}|M_{ij})}{p(\mathcal{D}|M_{ij})} \tag{14}$$

with $\mathcal{D} = (\boldsymbol{y}, \boldsymbol{Z})$ denoting the data. The notation makes clear that we now have a posterior $p(\boldsymbol{\theta}|M_{ij}, \mathcal{D})$, a likelihood $f(\mathcal{D}|\boldsymbol{\theta}, M_{ij})$, and a prior $\pi(\boldsymbol{\theta}|M_{ij})$ of the parameter vector for each candidate model.

The posterior model probability $p(M_{ij}|\mathcal{D})$ propagates model uncertainty into the posterior distribution of model parameters. By Bayes' rule, $p(M_{ij}|\mathcal{D})$ can be expressed as:

$$p(M_{ij}|\mathcal{D}) = \frac{f(\mathcal{D}|M_{ij})\pi(M_{ij})}{p(\mathcal{D})} \propto f(\mathcal{D}|M_{ij})\pi(M_{ij}) \tag{15}$$

such that the posterior model probability (weight) of model M_{ij} is proportional to the product of the model-specific marginal likelihood $f(\mathcal{D}|M_{ij})$ and the prior model probability $\pi(M_{ij})$. The model weights are converted into probabilities by normalizing relative to the set of all (IJ) models:

$$p(M_{ij}|\mathcal{D}) = \frac{f(\mathcal{D}|M_{ij})\pi(M_{ij})}{\sum_{r=1}^{I}\sum_{s=1}^{J} f(\mathcal{D}|M_{rs})\pi(M_{rs})}. \tag{16}$$

Model weights can hence be obtained using the marginal (or integrated) likelihood $f(\mathcal{D}|M_{ij})$ for each individual model M_{ij} after eliciting a prior $\pi(M_{ij})$ over the model space. The marginal likelihood of model M_{ij} is in turn given by:

$$f(\mathcal{D}|M_{ij}) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty f(\mathcal{D}|\boldsymbol{\theta}, M_{ij})\pi(\boldsymbol{\theta}|M_{ij}) \, \mathrm{d}\boldsymbol{\delta} \, \mathrm{d}\alpha \, \mathrm{d}\sigma^2. \tag{17}$$

The ratio of marginal likelihoods of two different models is the Bayes factor and it is closely related to the likelihood ratio statistic. Using the law of total probability the posterior density of the parameters for all the candidate models under consideration is given by:

$$p(\boldsymbol{\theta}|\mathcal{D}) = \sum_{i=1}^{I}\sum_{j=1}^{J} p(M_{ij}|\mathcal{D})p(\boldsymbol{\theta}|M_{ij}, \mathcal{D}). \tag{18}$$

Hence, the full posterior distribution of $\boldsymbol{\theta}$ is an average of the posterior distributions under each of the models, weighted by their posterior model probabilities $p(M_{ij}|\mathcal{D})$. When applying Bayesian model averaging according to Eq. (18), both estimation and inference come naturally together from the posterior distribution. This posterior distribution provides inference about $\boldsymbol{\theta}$ that takes full account of model uncertainty.

The posterior mean and variance of $\boldsymbol{\theta}$ are as follows:

$$\mathrm{E}(\boldsymbol{\theta}|\mathcal{D}) = \sum_{i=1}^{I}\sum_{j=1}^{J} p(M_{ij}|\mathcal{D})\mathrm{E}(\boldsymbol{\theta}|M_{ij}, \mathcal{D}) \tag{19}$$

$$\begin{aligned}\mathrm{Var}(\boldsymbol{\theta}|\mathcal{D}) &= \sum_{i=1}^{I}\sum_{j=1}^{J} p(M_{ij}|\mathcal{D})\mathrm{Var}(\boldsymbol{\theta}|M_{ij},\mathcal{D}) \\ &+ \sum_{i=1}^{I}\sum_{j=1}^{J} p(M_{ij}|\mathcal{D})\Big[\mathrm{E}(\boldsymbol{\theta}|M_{ij},\mathcal{D}) - \mathrm{E}(\boldsymbol{\theta}|\mathcal{D})\Big]^2. \end{aligned} \tag{20}$$

The posterior variance in Eq. (20) incorporates not only the weighted average of the estimated variances of the individual models, but also the weighted variance in estimates of the $\boldsymbol{\theta}$s across different models. This illustrates that model uncertainty is taken into account when making inference based on Bayesian model averaging. Within the Bayesian model averaging approach, we can also compute the posterior inclusion probability (PIP) for a given explanatory variable (spatial weight matrix). The posterior inclusion probability is calculated as the sum of the posterior model probabilities for all models including that variable (spatial weight matrix).

While Bayesian model averaging is an intuitively attractive solution to the problem of accounting for model uncertainty, its implementation in the context of our study is difficult because of two reasons. First, the number of terms in Eq. (18) is enormous, rendering exhaustive summation infeasible. Second, the integrals implicit in Eq. (18) are hard to compute. To tackle the first problem, Markov chain Monte Carlo model composition (MC^3) methods (Madigan and York 1995) can be used to directly approximate Eq. (18). This approach eliminates the need to consider all possible models by constructing a sampler that explores relevant parts of the very large model space.

Starting with an arbitrarily chosen model M we use a birth-death sampler which uniformly draws a candidate regressor out of the set of $2Q$ explanatory variables. Given this draw, a neighbor model M' is proposed. If the chosen regressor is already included in M, the sampler performs a *death* step by dropping the drawn regressor in the neighbor model. Conversely, the sampler performs a *birth* step by adding the chosen regressor, if the the drawn covariate is not in M. Models M and M' are therefore considered as being neighbors if their set of covariates differs only by a single regressor.

To account for uncertainty across different specifications of the spatial weight matrices, the sampler also has to address the transition between the spatial weighting schemes. We therefore extend the notion of the model neighborhood to include models containing the most similar spatial weight matrix, where similarity between spatial weight matrices is defined in terms of a concordance measure. Conditional on $\boldsymbol{W}_M$ (i. e. the spatial weight matrix in current state M), our MC^3 algorithm simultaneously draws a spatial weight matrix for each recursion. M' is then additionally modified by the chosen weight matrix $\boldsymbol{W}_{M'}$. The draw of $\boldsymbol{W}_{M'}$ depends on its similarities to $\boldsymbol{W}_M$, where similarity between matrices $\boldsymbol{W}_M$ and $\boldsymbol{W}_{M'}$ is rendered by matrix $\boldsymbol{G}$. Since we face J different spatial weight matrices, $\boldsymbol{G}$ is of size $J \times J$. The typical element of $\boldsymbol{G}$ is given by the sum of the cross-products of $\boldsymbol{W}_M$ and $\boldsymbol{W}_{M'}$:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}[\boldsymbol{W}_M]_{ij}[\boldsymbol{W}_{M'}]_{ij} \tag{21}$$

Indices given in Eq. (21) are described in more detail in Anselin (1995) and Hubert et al. (1985) and represent a cornerstone of various measures of spatial autocorrelation such as Moran's I. Row-normalization (i. e. dividing the elements by the respective row-sums) of $\boldsymbol{G}$ eventually yields a matrix where each row describes a transition probability vector from a particular spatial weight matrix to another.

After drawing both a regressor and a spatial weight matrix the algorithm moves from state M to M' with probability

$$\min\left[1, \frac{p(M'|\mathcal{D})}{p(M|\mathcal{D})}\right].$$

Otherwise the chain remains in state M. In our application we use $30,000,000$ iterations and discard the first $5,000,000$ draws (burn-ins) to dilute distortions emanating from the choice of a starting model.

The second difficulty in implementing Bayesian model averaging is that the integrals implicit in Eq. (18) are hard to compute. For matrix exponential spatial growth regression models, closed form integrals for the marginal likelihood (17) are not available. We therefore use Bayesian information criterion (BIC) weights as an approximation to $f(\mathcal{D}|M_{ij})$:

$$f(\mathcal{D}|M_{ij}) \simeq f(\mathcal{D}|\hat{\boldsymbol{\theta}}, M_{ij})N^{-d_{ij}/2} \tag{22}$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate of the parameter vector $\boldsymbol{\theta}$ in model M_{ij}, and d_{ij} denotes the number of parameters in model M_{ij}. This is the BIC approximation. The posterior model probability of model M_{ij} is obtained by premultiplying Eq. (22) by the prior model probability $p(M_{ij})$ and dividing by the sum of all (IJ) possible models:

$$p(M_{ij}|\mathcal{D}) = \frac{\pi(M_{ij})f(\mathcal{D}|\hat{\boldsymbol{\theta}}, M_{ij})N^{-d_{ij}/2}}{\sum_{r=1}^{I}\sum_{s=1}^{J}\pi(M_{rs})f(\mathcal{D}|\hat{\boldsymbol{\theta}}, M_{rs})N^{-d_{rs}/2}}. \tag{23}$$

The posterior model weights (23) equal the prior model weights times the (exponentiated) Bayesian information criterion, developed by Schwarz (1978). The BIC weights depend on the likelihood evaluated at the maximum likelihood estimate, but penalize relatively large models through the penalty term $N^{-d_{ij}/2}$. Bayesian information criterion model weights have been widely discussed in the literature (see Schwarz 1978, Kass and Wasserman 1995, Raftery 1995). There exist several approaches which make use of BIC as a means to approximate posterior model probabilities in the empirical growth literature. For example, Sala-i-Martin et al. (2004) advocate frequentist ordinary least squares estimates along with BIC model weights. This approach became known as Bayesian model averaging of classical estimates (BACE). A generalization to maximum likelihood estimates is provided by Moral-Benito (2012). The use of information-theoretic quantities to calculate model weights is moreover extensively used in the frequentist model averaging literature. Frequentist model averging techniques are thoroughly described in Burnham and Anderson (2004) and Claeskens and Hjort (2008).

Model priors $\pi(M_{ij})$ are the only quantities which remain to be chosen. Many studies use a uniform model prior structure by assigning equal probability to the models prior seeing the data $\mathcal{D}$. However, a uniform prior structure would implicitly impose a mean prior model size of Q, since the majority of models in $\mathcal{M}$ are of such intermediate size. Therefore, several alternatives to a uniform model prior have been proposed in the literature. For example, Ley and Steel (2009) propose the use of binomial-beta priors which uses a hyperprior on the inclusion of each regressor:

$$\pi(M_{ij}) \propto \Gamma(1+\varphi_{ij})\Gamma(1+2Q-\varphi_{ij}) \tag{24}$$

where $\Gamma(\cdot)$ and φ_{ij} denote the gamma function and the number of non-constant covariates in model M_{ij}, respectively. Unlike a uniform model prior, which attaches uniform prior mass to the respective models, the prior given in Eq. (24) is uniform over prior expected model size.[2]

An applied illustration

This section serves to illustrate the model averaging approach on a matrix exponential spatial specification of the spatial Durbin growth regression for both identifying model covariates and unveiling spatial structures in pan-European growth data.

The sample data

Our sample is a cross-section of 273 European regions representing 28 European countries[3] over the 2000-2010 period. The units of observation are the NUTS-2 regions (NUTS revision 2010).[4] These regions, though varying in size, are generally considered to be appropriate spatial units for modeling and analysis purposes. In most cases, they are sufficiently small to capture subnational variations. But we are aware that NUTS-2 regions are formal rather than functional regions, and their delineation does not represent the boundaries of regional growth processes very well. The sample regions include regions located in Austria (nine regions), Belgium (11 regions), Bulgaria (six regions), Czech Republic (eight regions), Denmark (five regions), Estonia (one region), Finland (five regions), France (22 regions), Germany (38 regions), Greece (13 regions), Hungary (seven regions), Italy (21 regions), Latvia (one region), Lithuania (one region), Luxembourg (one region), Netherlands (12 regions), Norway (seven regions), Poland (16 regions), Portugal (five regions), Republic of Ireland (two regions), Romania (eight regions), Slovakia (four regions), Slove-

[2]The appendix provides a simulation study to test the ability of our approach.

[3] EU-27 member states plus Norway and Switzerland minus Cyprus

[4] We exlude the Spanish North African territories of Ceuta and Melilla, the Canary Islands, the Portuguese non-continental territories Açores and Madeira, the French Départements d'Outre Mer Guadeloupe, Martinique, Guyane and Réunion. Moreover, because of data availability problems we use the NUTS revision of 2006 rather than 2010 for the Finnish regions Aland, Etelä-Suomi, Itä-Suomi, Länsi-Suomi and Pohjois-Suomi.

nia (two regions), Spain (16 regions), Sweden (eight regions), Switzerland (seven regions) and the United Kingdom (37 regions).

We use gross-value added (gva), rather than gross regional product at market prices as a proxy for regional income. The proxy is measured in accordance with the European System of Accounts (ESA) 1995. The data for the European regions come from the Cambridge Econometrics database. The dependent variable represents average growth rates over the period 2001-2010.

[Table 1 about here]

We consider a set of $Q = 32$ candidate explanatory variables as well as their spatially lagged forms. To avoid potential endogeneity problems all the variables are measured at the beginning of the sample period (that is, 2000). The variable names and the data sources are depicted in Table 1. A very popular variable in the regional growth regression literature is the initial level of income. Most studies include this variable and find it to be significant. Proxies for human capital are also widely considered as a key determinant of economic growth. We measure human capital by the skills of the workforce as given by the level of educational attainment of the population, and distinguish between lower and higher educated workers, where high and low education levels are defined by the ISCED (international standard classification of education) levels 1-2 and 5-6, respectively. We also included physical capital stocks constructed using the perpetual inventory method using a depreciation rate of ten per cent and investment data for the years 1990-2000.

There is substantial empirical evidence supporting the role of high-technology firms in technological change and economic growth. Despite the inherent difficulties in measuring the effects of technological progress on economic growth we rely on two candidate variables that capture different aspects of the process of innovation and technological change at the regional level. We consider the ratio of the number of high-technology patent applications at the European Patent Office (EPO) to gross-value added per capita as a proxy for the output of high-technology invention activities in each region. Another candidate variable, the share of human resources employed in science and technology, represents a technology input measure. To account for the industrial mix, we also consider the shares of employment in agriculture, mining-manufacturing-energy, construction, and market services. Harris (1954), and LeSage and Fischer (2008) argue that market access is also important for regional income. We therefore include an index of market potential that measures the export demand each region faces given its spatial location and that of its trading partners.

Theoretical as well as empirical studies show that the age-structure of the population might exert a decisive effect on economic growth (see Azomahou and Mishra 2008, Boucekkine et al. 2002). We rely on two measures to proxy the demographic structure of the regions. First, the child-dependency ratio of a region which is defined as the number of people aged 0-14 as a ratio to the number of people aged 15-64. Second, the old-age dependency ratio of a region which is given by the ratio between the number of people aged 65 and over and the number of people aged 15-64. Both variables capture the burden of the economic productive part of the population to maintain the economically dependent.

Moreover, we follow Fingleton (2001) and consider population density, employment density and output density as candidate explanatory variables to control for urban agglomerations. Urban agglomerations are typically equipped with larger human capital stocks as a repository of knowledge, which facilitates innovation creation and adoption and thus accelerates technological progress and economic growth.

Finally, we also include candidate explanatory variables in the regressions with the purpose of accounting for likely differences in the access to sea, roads, air and rail transport, and a series of dummy variables suggested by Crespo Cuaresma et al. (2013), and Crespo Cuaresma and Feldkircher (2013).

Alternative spatial weights

In order to illustrate the ability of our approach to both identify model covariates and unveil spatial structures present in the data, we restrict our space of potential (row-normalized) spatial weight matrices to three different classes: binary queen contiguity matrices, (binary) k-nearest neighbor matrices and (binary) distance-based matrices. Queen contiguity matrices consider regions as neighbors if they share a common border (including cases where the common border is just a vertex). We will consider first-order and second-order contiguity definitions for neighbors in this class. k-nearest neighbor matrices constrain the neighbor structure to the k-nearest neighbors and thereby precluding islands and forcing each region to have the same number of neighbors. For this class of weighting matrices, we consider $k = 5, \dots, 14$. Finally, distance-based matrices are based on a distance criterion, such that two regions i and j are defined as neighbors when the distance between them is less than a given critical value d. Critical distance is defined here by the first and second quintile of the entire distribution, respectively.

k-nearest neighbor and distance-based spatial weight matrices are used with three alternative distance metrics that reflect different aspects of spatial connectivity: (i) geodesic distances, (ii) road travel time distances for cars, and (iii) drive time distances for heavy goods vehicles. LeSage and Fischer (2008) argue that drive time measures of distance reflect economic distance which may introduce important aspects to connectivity. The structure of the road networks, presence of mountains, rivers, landlocked areas, national car and lorry speed limit, as well as statutory rest periods for drivers may lead to considerable differences between geodesic and drive time distances. The travel time spatial weight matrices are based on information on road infrastructure from the European transport network database of the Institute of Spatial Planning in Dortmund (IRPUD) based on reference year 2005.

A comparison of alternative spatial weight matrices

An important point to note about spatial model comparison is that the performance will depend on the strength of the spatial dependence in the sample data. LeSage and Pace (2009) illustrate this for spatial weight matrix comparisons in the case of conventional SAR models using data generated experiments. They show that values for the spatial dependence parameter close to zero make it difficult to distinguish between alternative spatial weight matrices. Since the spatial dependence in our growth model is moderately strong, with $\rho \approx 0.64$, this should not present a problem here.

[Table 2 about here]

Posterior probabilities of inclusion for the 38 spatial weight matrices are shown in Table 2. We see support for a spatial weight matrix based on 10 (probability of inclusion: 0.396) and 11 nearest neighbors (probability of inclusion: 0.237) where distance is measured in terms of lorry travel times or geodesic distances, respectively. Since the average number of first-order contiguous neighbors for the European regions in our sample is near five, and the average number of second-order contiguous neighbors is 13, this suggests a spatial connectivity structure that extends beyond first-order contiguous regions, but not to all second-order contiguous neighbors.

High probability matrix exponential spatial growth regression models
Running the MC^3 sampler for 30 million draws and discarding the first 5 million iterations produced 114,864 unique models. Note that there are $2^{64} \approx 1.84\ 10^{19}$ possible models based on alternative ways to combine the 32 candidate explantory variables and their spatial lags, and for each of these another 38 possible spatial weight matrices that can be used with each of these models. As a test for convergence of the MC^3 procedure, we produced several runs of the sampler using different starting models which resulted in correlations between posterior model probabilities above 0.99. In all cases, the results are nearly identical, suggesting that the MC^3 procedure is converging sufficiently.[5]

[Table 3 about here]

Table 3 shows the variables appearing in the ten highest posterior probability models, along with the model probabilities. The posterior probabilities for these models are (0.0994, 0.0524, 0.0448, 0.0349, 0.0255, 0.0229, 0.0130, 0.0121, 0.0107, 0.0104) accounting for 33.9 percent of the posterior probability mass. Variables that appear in the respective models are designated with a '1', and those that do not appear with a '0'. The bottom rows of the table show the number of variables included, the particular spatial weight matrix employed, and the posterior model probability. Fernández et al. (2001a) provide details on calculations of posterior inclusion probabilities of individual variables. We find that two of the 32 variables (initial income and lower education workers) appear in all ten highest probability models, and two variables (capital city regions and Objective 1 regions) appear in one model and not in others. Another 27 of the 32 variables do not appear in one of the top ten models.

Model averaged parameter and impact estimates
Table 4 depicts the posterior inclusion probabilities and model-averaged parameter estimates of the variables and their spatial lags. In addition to posterior standard deviations, Table 4 also reports conditional sign certainty probabilities as another measure of the significance of variables and their spatial lags (see Sala-i-Martin

[5] For alternative convergence diagnostics, see Fernández et al. (2001b).

et al. 2004). Conditional sign certainty probabilities are calculated from the marginal posterior distribution which only consists of models where the respective variable or its spatial lag is included. Conditional on the inclusion of a variable or its spatial lag this metric measures the probability that a coefficient has the same sign as its posterior mean.

[Table 4 about here]

The posterior mean (i. e. the model averaged estimate) of the spatial autocorrelation parameter ρ ($\rho \approx 1 - \exp(\alpha)$) amounts to 0.64. With a corresponding posterior standard deviation of 0.11, α is estimated very precisely. The high magnitude of the estimated spatial autocorrelation parameter stresses the importance of accounting for spatial dependence in the observations, since it is well-known that an erroneously omitted spatial lag in the dependent variable results in biased and inconsistent parameter estimates (LeSage and Pace 2009).

With posterior inclusion probabilities close to unity, we identify the variable lower education attainment and its spatial lag as well as the variable initial income as the most important growth determinants. The posterior mean of initial income has a negative sign with a sign certainty probability of unity. Our results thus suggest that poorer regions grow, on average, faster than richer regions – after controlling for other factors – highlighting income convergence among the regions in the sample. Interestingly, Table 4 reveals a positive posterior mean of the spatial lag of initial income giving rise to positive growth spillovers emanating from the intital income of neighboring regions. Regions thus may benefit from being close to rich neighbors. But the spatial lag of initial income receives with a probability of inclusion of 64 percent only moderate posterior support. The negative conditional income convergence effect appears to outweight positive growth spillovers from neighboring regions.

Both lower educational attainment measured in terms of primary and lower secondary education representing the highest degree obtained by population aged 25 and over, and its spatial lag exhibit a posterior probability of inclusion of unity. As expected, the variable lower education workers has a negative posterior mean. However, our results also suggest a positive posterior mean of its spatial lag. While a poorly educated labor force hampers income growth in the same region, our results suggest, however, positive effects on income growth rates to neighboring regions.

[Table 5 about here]

In Table 5 we report model averaged direct and indirect impact estimates based on the 114,864 models found by the MC^3 algorithm. From the estimates we see that the impact estimates of 30 from the 32 variables fell within two standard deviations of zero. This leaves us with two variables from the set of candidate explanatory variables that exerted significant total impact on growth with probabilities of inclusion above 99 percent. These variables were initial income and educational attainment measured by primary and lower secondary education representing the highest degree obtained by population aged 25 and over. Initial income was found to exert both a negative direct and total impact on growth. The respective spatial spillover effect, however, was estimated very imprecisely. The educational attainment variable had a negative direct, but positive indirect impact, which implies that a region, on average, benefits from a marginal increase in the share of low-educated in working age population in all other regions. Both estimates were found to be highly significant.

In concluding we note that our approach allowed us to provide estimates and inference, which variables from a set of 32 candidate explanatory variables exerted a significant impact on economic growth rates. Only two of these variables (initial income and lower education) were found to exhibit a posterior probability of inclusion close to unity. Both initial income and lower education exert a negative influence. But it is worth noting that the direct impact of the latter variable is -0.058 and the spillover impact 0.020 revealing the total negative influence. Moreover, we see support for spatial weight matrices based on 10 and 11 nearest neighbors where distance is measured in terms of lorry travel times and geodesic distances, respectively.

Closing remarks

The problem of model uncertainty can arise from several sources. First, the selection of appropriate variables is a difficult issue in growth empirics and involves a trade-off between the arbitrary selection of a small number of variables, which may imply some omitted variables bias, and the introduction of a larger set of variables with a number of econometric problems such as endogeneity or multicollinearity. A second source of model uncertainty arises in a spatial setting. One also has to specify the spatial weight matrix that defines connectivity between regions. In other words, the estimates and inferences in spatial growth regressions are not only conditional on the set of explanatory variables used but also on the selected spatial weight matrix.

LeSage and Fischer (2008) derive a Bayesian model averaging approach that simultaneously specifies the spatial weight structure and the explanatory variable in spatial Durbin models with an application to European regions. This paper departs from this previous research in two respects: First, it employs a spatial Durbin model based on the matrix exponential which replaces the geometric pattern of decay in the conventional spatial Durbin model with one of exponential decay. This specification has theoretical and computational advantages, since it eliminates the need to calculate the log-determinant when producing maximum likelihood estimates. Second, our solution to model uncertainty involves using a Bayesian information criterion approximation to the marginal likelihood for the models of interest to calculate model weights and form an averaged model. This greatly simplifies the task of accounting for the above motivated sources of model uncertainty, and enables to consider not only k-nearest neighbor, but also contiguity-based and distance-based spatial weight matrices within one framework.

Appendix: A simulation study

We tested our approach using a generated vector of growth rates constructed from our sample data for 273 European regions. We draw 10 potential independent variables $\tilde{\boldsymbol{X}} = \begin{bmatrix}\tilde{\boldsymbol{x}}_1 & \tilde{\boldsymbol{x}}_2 & \cdots & \tilde{\boldsymbol{x}}_{10}\end{bmatrix}$ using $N = 273$ draws from a standard normal distribution for each covariate, so as to match the sample size of our empirical application. The spatial autoregressive parameter ρ is fixed at $\rho \approx 0.63$ (i. e. $\alpha = -1$), a typical level of spatial dependence in growth data sets. Data on the dependent variable are generated according to:

$$\tilde{\boldsymbol{y}} = \exp(1\tilde{\boldsymbol{W}})(\tilde{\boldsymbol{y}} + \boldsymbol{\iota}_N + 1.5\tilde{\boldsymbol{x}}_1 + 2\tilde{\boldsymbol{x}}_2 - 0.5\tilde{\boldsymbol{x}}_3 + 0.5\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_1 + 0.25\tilde{\boldsymbol{\varepsilon}}) \tag{25}$$

where $\tilde{\boldsymbol{\varepsilon}}$ is a standard normal variable. $\tilde{\boldsymbol{W}}$ is an $N \times N$ row-stochastic spatial weight matrix. We restrict the space of potential weighting matrices to three different matrices: a first order-queen contiguity, a 10-nearest neighbor and a distance-based spatial weight matrix with critical distance being the first quantile of the entire distribution. All these alternative spatial weight matrices belong to the class of binary weight matrices and differ only with respect to the definition of the set of neighbors. Distance is measured in terms of geodesic distances. For the simulation study, matrix $\tilde{\boldsymbol{X}}$ is used as the set of potential regressors, along with the entire set of 38 potental spatial weight matrices used in our applied illustration in order to unveil the data generating process given in Eq. (25).

[Table 6 about here]

Table 6 depicts the outcomes of the simulation study. The results refer to averages over $1,000$ simulated datasets for each spatial weight matrix. For each simulated dataset we used $10,000$ iterations after discarding the first $5,000$ draws (burn-ins) of our MC^3 sampler. The posterior mean estimates for all parameters are very close to their true values. Moreover, the posterior inclusion probability of the correct spatial weight matrix is very close to unity. The results thus indicate that our approach is able to identify the correct model parameterization with very high precision.

References

Anselin, L. (1995): Local indicators of spatial association – LISA, *Geographical Analysis* 27 (2), 93–115

Azomahou, T., Mishra, T. (2008): Age dynamics and economic growth: Revisiting the nexus in a nonparametric setting, *Economic Letters* 99 (1), 67–71

Boucekkine, R., de la Croix, D., Licandro, O. (2002): Vintage human capital, demographic trends, and endogenous growth, *Journal of Economic Theory* 104 (2), 340–375

Burnham, K., Anderson, D. (2004): Multimodel inference: Understanding AIC and BIC in model selection, *Sociological Methods & Research* 33 (2), 261–304

Claeskens, G., Hjort, N. (2008): *Model selection and model averaging*. Cambridge: Cambridge University Press

Crespo Cuaresma, J., Doppelhofer, G., Feldkircher, M. (2013): The determinants of economic growth in European regions, *Regional Studies*, forthcoming

Crespo Cuaresma, J., Feldkircher, M. (2013): Spatial filtering, model uncertainty and the speed of income convergence in Europe, *Journal of Applied Econometrics* 28 (4), 720–741

Fernández, C., Ley, E., Steel, M. (2001a): Model uncertainty in cross-country growth regressions, *Journal of Applied Econometrics* 16 (5), 563–576

Fernández, C., Ley, E., Steel, M. (2001b): Benchmark priors for Bayesian model averaging, *Journal of Econometrics* 100 (2), 381–427

Harris, C. (1954): The market as a factor in the localization of industry in the United States, *Annals of the Association of American Geographers* 44 (4), 315–348

Hubert, L., Golledge, R., Constanzo, C., Gale, N. (1985): Measuring association between spatially defined variables: An alternative procedure, *Geographical Analysis* 17 (1), 36–46

Kass, R., Wasserman, L. (1995): A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion, *Journal of the American Statistical Association* 90 (431), 928–934

Koop, G. (2003): *Bayesian Econometrics*. Chichester: John Wiley & Sons

LeSage, J., Fischer, M. (2008): Spatial growth regressions, model specification, estimation, and interpretation, *Spatial Economic Analysis* 3 (3), 275–304

LeSage, J., Pace, K. (2009): *Introduction to Spatial Econometrics*. Boca Raton: CRC Press

LeSage, J., Pace, K. (2007): A matrix exponential spatial specification, *Journal of Econometrics* 140 (1), 190–214

LeSage, J., Parent, O. (2007): Bayesian model averaging for spatial econometric models, *Geographical Analysis* 39 (3), 241–267

Ley, E., Steel, M. (2009): On the effect of prior assumptions in Bayesian model averaging with applications to growth regression, *Journal of Applied Econometrics* 24 (4), 651–674

Madigan, D., York, J. (1995): Bayesian graphical models for discrete data, *International Statistical Review* 63 (2), 215–232

Fingleton, B. (2001): Theoretical economic geography spatial econometrics: Dynamic perspectives, *Journal of Economic Geography* 1 (2), 201–225

Moral-Benito, E. (2012a): Determinants of economic growth: A Bayesian panel data approach, *The Review of Economics and Statistics* 94 (2), 566–579

Raftery, A. (1995): Bayesian model selection in social research, *Sociological Methodology* 25, 111–163

Sala-i-Martin, X., Doppelhofer, G., Miller, I. (2004): Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach, *American Economic Review* 94 (4), 813–835

Schwarz, G. (1978): Estimating the dimension of a model, *Annals of Statistics* 6 (2), 461–464

Tables

Table 1: The variables used in the analysis

Variable	Description
Initial income	Gross-value added divided by population, 2000. *Source*: Cambridge Econometrics
Physical capital	Gross fixed capital formation, 2000. *Source*: Cambridge Econometrics
Higher education workers (share)	Share of population (aged 25 and over, 2000) with higher education (ISCED levels 1-2). *Source*: Eurostat
Lower education workers (share)	Share of population (aged 25 and over, 2000) with lower education (ISCED levels 5-6). *Source*: Eurostat
High-technology invention activities	Measured in terms of the ratio of the number of high-technology EPO (European Patent Office) patent-applications to gross-value added per capita, 2000. High-technology is defined to include the ISIC sectors of aerospace (ISIC 3845), electronics and telecommunication (ISIC 3832), computers and office equipment (ISIC 3825), and pharmaceuticals (ISIC 3522). *Source*: European Patent Office
Technology resources	Human resources in science and technology, share in persons employed, 2000. *Source*: Eurostat
Agricultural employment	Share of NACE A and B (agriculture) in total employment, 2000. *Source*: Cambridge Econometrics
Manufacturing employment	Share of NACE C to E (mining, manufacturing and energy) in total employment, 2000. *Source*: Cambridge Econometrics
Construction employment	Share of NACE F (construction) in total employment, 2000. *Source*: Cambridge Econometrics
Market services employment	Share of NACE G to K (market services) in total employment, 2000. *Source*: Cambridge Econometrics
Market potential	For a region defined in terms where the size of the regional economy is proxied by gross-value added, and the distance is the interregional great circle distance. *Source*: gross-value added data from Cambridge Econometrics
Output density	Gross-value added per square km, 2000. *Source*: Eurostat
Employment density	Employed persons per square km, 2000. *Source*: Eurostat
Population density	Population per square km, 2000. *Source*: Eurostat
Population growth	Average growth rate of the population for 1996-2000. *Source*: Eurostat
Unemployment rate	Average unemployment rate for 1996-2000. Unemployment rate is defined as the share of unemployed persons of the economically active population *Source*: Eurostat
Labor force participation rate	Employed and unemployed persons as a share of total population, 2000. *Source*: Eurostat
Child dependency ratio	The ratio of the number of people aged 0-14 to the number of people aged 15-64, 2000. *Source*: Eurostat
Old-age dependency ratio	The ratio of the number of people aged 65 and over to the number of people aged 15-64, 2000. *Source*: Eurostat
Peripheriality	Measured in terms of distance to Brussels
Accessibility road	Potential accessibility road, ESPON space=100. *Source*: ESPON
Accessibility rail	Potential accessibility rail, ESPON space=100. *Source*: ESPON
Region with a seaport	Dummy variable, 1 denotes region with seaport, 0 otherwise. *Source*: ESPON
Region with an airport	Dummy variable, 1 denotes region with airport, 0 otherwise. *Source*: ESPON
Coastal region	Dummy variable, 1 denotes region with coast, 0 otherwise. *Source*: ESPON
Capital city region	Dummy variable, 1 denotes region with capital city, 0 otherwise. *Source*: ESPON
Region with a large city	Dummy variable, 1 denotes region with a city larger than $300,000$ inhabitants, 0 otherwise. *Source*: ESPON
Rural region	Dummy variable, 1 denotes region with a population density lower than 100 and without a city larger than $125,000$ inhabitants, 0 otherwise. *Source*: ESPON
Objective 1 region	Dummy variable, 1 denotes region eligible under Objective 1 for 2000-2006, 0 otherwise. *Source*: ESPON
Border region	Dummy variable, 1 denotes region with country borders, 0 otherwise. *Source*: ESPON
EU-15 region	Dummy variable, 1 denotes region belonging to the 15 pre-2004 EU member states, 0 otherwise
Pentagon region	Dummy variable, 1 denotes region belonging to the Pentagon shaped by London, Paris, Munich, Milan and Hamburg, 0 otherwise. *Source*: ESPON

Table 2: Comparison of alternative spatial weight matrices

Spatial weight matrix	Probability of inclusion
First-order contiguity	0.0000
Second-order contiguity	0.0000
Geodesic 5-nearest neighbors	0.0000
Geodesic 6-nearest neighbors	0.0000
Geodesic 7-nearest neighbors	0.0002
Geodesic 8-nearest neighbors	0.0005
Geodesic 9-nearest neighbors	0.0053
Geodesic 10-nearest neighbors	0.0087
Geodesic 11-nearest neighbors	0.2371
Geodesic 12-nearest neighbors	0.0218
Geodesic 13-nearest neighbors	0.0240
Geodesic 14-nearest neighbors	0.0166
Car travel time 5-nearest neighbors	0.0000
Car travel time 6-nearest neighbors	0.0000
Car travel time 7-nearest neighbors	0.0000
Car travel time 8-nearest neighbors	0.0000
Car travel time 9-nearest neighbors	0.0112
Car travel time 10-nearest neighbors	0.0050
Car travel time 11-nearest neighbors	0.0106
Car travel time 12-nearest neighbors	0.0009
Car travel time 13-nearest neighbors	0.0032
Car travel time 14-nearest neighbors	0.0042
Lorry travel time 5-nearest neighbors	0.0000
Lorry travel time 6-nearest neighbors	0.0000
Lorry travel time 7-nearest neighbors	0.0000
Lorry travel time 8-nearest neighbors	0.0009
Lorry travel time 9-nearest neighbors	0.0306
Lorry travel time 10-nearest neighbors	0.3963
Lorry travel time 11-nearest neighbors	0.0285
Lorry travel time 12-nearest neighbors	0.0525
Lorry travel time 13-nearest neighbors	0.0427
Lorry travel time 14-nearest neighbors	0.0987
Geodesic distance-based, first quintile	0.0000
Geodesic distance-based, second quintile	0.0000
Car travel time-based, first quintile	0.0000
Car distance-based, second quintile	0.0000
Lorry travel time-based, first quintile	0.0004
Lorry travel time-based, second quintile	0.0000

Table 3: High probability models

Variable name	Model 10	Model 9	Model 8	Model 7	Model 6	Model 5	Model 4	Model 3	Model 2	Model 1
Initial income	1	1	1	1	1	1	1	1	1	1
Physical capital	0	0	0	0	0	0	0	0	0	0
Higher education workers	0	0	0	0	0	0	0	0	0	0
Lower education workers	1	1	1	1	1	1	1	1	1	1
High-technology invention activities	0	0	0	0	0	0	0	0	0	0
Technology resources	0	0	0	0	0	0	0	0	0	0
Agricultural employment	0	0	0	0	0	0	0	0	0	0
Manufacturing employment	0	0	0	0	0	0	0	0	0	0
Construction employment	0	0	0	0	0	0	0	0	0	0
Market services employment	0	0	0	0	0	0	0	0	0	0
Market potential	0	0	0	0	0	0	0	0	0	0
Output density	0	0	0	0	0	0	0	0	0	0
Employment density	0	0	0	0	0	0	0	0	0	0
Population density	0	0	0	0	0	0	0	0	0	0
Population growth	0	0	0	0	0	0	0	0	0	0
Unemployment rate	0	0	0	0	0	0	0	0	0	0
Labor force participation rate	0	0	0	0	0	0	0	0	0	0
Child dependency ratio	0	0	0	0	0	0	0	0	0	0
Old-age dependency ratio	1	0	1	0	0	1	1	0	0	0
Peripheriality	0	0	0	0	0	0	0	0	0	0
Accessibility road	0	0	0	0	0	0	0	0	0	0
Accessibility rail	0	0	0	0	0	0	0	0	0	0
Region with a seaport	0	0	0	0	0	0	0	0	0	0
Region with an airport	0	0	0	0	0	0	0	0	0	0
Coastal region	0	0	0	0	0	0	0	0	0	0
Capital city region	0	1	0	0	0	0	0	0	0	0
Region with a large city	0	0	0	0	0	0	0	0	0	0
Rural region	0	0	0	0	0	0	0	0	0	0
Objective 1 region	1	0	0	0	0	0	0	0	0	0
Border region	0	0	0	0	0	0	0	0	0	0
EU-15 region	0	0	0	0	0	0	0	0	0	0
Pentagon region	0	0	0	0	0	0	0	0	0	0
Number of variables	4	3	3	2	2	3	3	2	2	2
Spatial weight matrix $\boldsymbol{W}$	12 nearest lorry	10 nearest lorry	14 nearest lorry	11 nearest geodesic	14 nearest lorry	10 nearest lorry	11 nearest geodesic	10 nearest lorry	11 nearest geodesic	10 nearest lorry
Posterior model probability	0.0104	0.0107	0.0121	0.0130	0.0229	0.0255	0.0349	0.0448	0.0524	0.0994

Table 4: Model averaged estimates

Variable	Inclusion prob.	Mean	Standard dev.	Sign prob.
α, $\rho \approx 1 - \exp(\alpha)$		-1.0223	0.1084	1.0000
Initial income	0.9877	-0.8239	0.2558	1.0000
Physical capital	0.0273	-0.0034	0.0205	1.0000
Higher education workers	0.0042	0.0000	0.0002	0.8153
Lower education workers	1.0000	-0.0602	0.0037	1.0000
High-technology invention activities	0.0229	-0.2164	1.4103	1.0000
Technology resources	0.0043	0.0000	0.0002	0.6158
Agricultural employment	0.0139	-0.0002	0.0016	0.9989
Manufacturing employment	0.0046	0.0000	0.0004	0.6912
Construction employment	0.0371	0.0019	0.0100	0.9998
Market services employment	0.0152	0.0002	0.0016	0.9998
Market potential	0.0047	-0.0001	0.0020	0.7473
Output density	0.0065	0.0003	0.0054	0.8893
Employment density	0.0100	-0.0006	0.0105	0.9625
Population density	0.0131	-0.0012	0.0132	0.9886
Population growth	0.0054	0.0001	0.0065	0.6517
Unemployment rate	0.0056	0.0000	0.0007	0.9485
Labor force participation rate	0.0080	0.0001	0.0010	0.9871
Child dependency ratio	0.0079	-0.0001	0.0021	0.8130
Old-age dependency ratio	0.3151	-0.0128	0.0188	1.0000
Peripheriality	0.0051	0.0002	0.0035	0.8621
Accessibility road	0.0148	0.0000	0.0002	0.9992
Accessibility rail	0.0107	0.0000	0.0002	0.9908
Region with a seaport	0.0056	0.0005	0.0066	0.9856
Region with an airport	0.0053	-0.0004	0.0055	0.9998
Coastal region	0.0061	0.0006	0.0078	0.9919
Capital city region	0.0881	0.0415	0.1321	1.0000
Region with a large city	0.0051	0.0003	0.0056	0.9482
Rural region	0.0050	0.0004	0.0072	0.9021
Objective 1 region	0.1042	0.0427	0.1258	1.0000
Border region	0.0055	0.0004	0.0062	0.9995
EU-15 region	0.0158	-0.0053	0.0445	1.0000
Pentagon region	0.0094	-0.0015	0.0164	0.9934
$\boldsymbol{W}$ initial income	0.6419	0.4841	0.3735	0.9999
$\boldsymbol{W}$ physical capital	0.0066	0.0002	0.0112	0.4629
$\boldsymbol{W}$ higher education workers	0.0056	0.0000	0.0007	0.9415
$\boldsymbol{W}$ lower education workers	0.9936	0.0464	0.0056	1.0000
$\boldsymbol{W}$ high-technology invention activities	0.0214	-0.4415	3.1287	0.9979
$\boldsymbol{W}$ technology resources	0.0103	0.0001	0.0016	0.9250
$\boldsymbol{W}$ agricultural employment	0.0093	-0.0001	0.0016	0.6054
$\boldsymbol{W}$ manufacturing employment	0.0105	-0.0003	0.0027	0.9829
$\boldsymbol{W}$ construction employment	0.0155	0.0015	0.0152	0.9931
$\boldsymbol{W}$ market services employment	0.0147	0.0002	0.0022	0.9902
$\boldsymbol{W}$ market potential	0.0059	0.0001	0.0055	0.5341
$\boldsymbol{W}$ output density	0.0110	0.0012	0.0144	0.8967
$\boldsymbol{W}$ employment density	0.0054	0.0002	0.0118	0.7117
$\boldsymbol{W}$ population density	0.0051	0.0000	0.0119	0.3665
$\boldsymbol{W}$ population growth	0.0104	0.0037	0.0393	0.9935
$\boldsymbol{W}$ unemployment rate	0.0055	0.0001	0.0014	0.8079
$\boldsymbol{W}$ labor force participation rate	0.0338	0.0011	0.0062	1.0000
$\boldsymbol{W}$ child dependency ratio	0.0118	0.0006	0.0066	0.9877
$\boldsymbol{W}$ old-age dependency ratio	0.0111	-0.0005	0.0052	0.9939
$\boldsymbol{W}$ peripheriality	0.0055	-0.0001	0.0067	0.5047
$\boldsymbol{W}$ accessibility road	0.0100	0.0000	0.0002	0.9750
$\boldsymbol{W}$ accessibility rail	0.0069	0.0000	0.0001	0.8730
$\boldsymbol{W}$ region with a seaport	0.0052	0.0005	0.0094	0.9085
$\boldsymbol{W}$ region with an airport	0.0098	0.0033	0.0360	0.9896
$\boldsymbol{W}$ coastal region	0.0053	0.0006	0.0103	0.9078
$\boldsymbol{W}$ capital city region	0.0048	0.0011	0.0271	0.8008
$\boldsymbol{W}$ region with a large city	0.0284	0.0174	0.1035	1.0000
$\boldsymbol{W}$ rural region	0.0060	0.0012	0.0208	0.9130
$\boldsymbol{W}$ Objective 1 region	0.0150	0.0065	0.0605	0.9112
$\boldsymbol{W}$ border region	0.0047	-0.0003	0.0074	0.7548
$\boldsymbol{W}$ EU-15 region	0.0042	0.0000	0.0002	1.0000
$\boldsymbol{W}$ Pentagon region	0.0108	-0.0023	0.0233	0.9937

Table 5: Model averaged impact estimates

Variable	Average direct impacts			Average indirect impacts			Average total impacts		
	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.	Mean	Std. dev.	Sign prob.
Initial income	-0.8083	0.1631	1.0000	-0.1102	0.4905	0.4954	-0.9185	0.3847	0.9900
Physical capital	-0.0036	0.0213	0.9003	-0.0052	0.0503	0.8939	-0.0088	0.0676	0.8944
Higher education workers	-0.0000	0.0002	0.2886	0.0001	0.0019	0.7132	0.0001	0.0020	0.7132
Lower education workers	-0.0583	0.0029	1.0000	0.0207	0.0085	0.9896	-0.0376	0.0084	1.0000
High-technology invention activities	-0.2856	1.5602	0.9981	-1.6439	9.0263	0.9980	-1.9295	9.8028	0.9980
Technology resources	0.0000	0.0003	0.8357	0.0004	0.0044	0.8407	0.0004	0.0046	0.8407
Agricultural employment	-0.0002	0.0018	0.8574	-0.0006	0.0057	0.8560	-0.0008	0.0068	0.8561
Manufacturing employment	-0.0000	0.0003	0.7061	-0.0006	0.0064	0.7096	-0.0006	0.0066	0.7096
Construction employment	0.0023	0.0109	0.9992	0.0077	0.0421	0.9992	0.0100	0.0483	0.9992
Market services employment	0.0002	0.0016	0.9960	0.0010	0.0063	0.9960	0.0012	0.0073	0.9960
Market potential	-0.0001	0.0023	0.5830	0.0001	0.0146	0.4174	0.0001	0.0158	0.4174
Output density	0.0004	0.0045	0.9150	0.0049	0.0469	0.9160	0.0053	0.0495	0.9160
Employment density	-0.0006	0.0105	0.7457	-0.0004	0.0408	0.7431	-0.0011	0.0473	0.7434
Population density	-0.0013	0.0134	0.8261	-0.0023	0.0431	0.8224	-0.0035	0.0521	0.8238
Population growth	0.0005	0.0080	0.8736	0.0104	0.1028	0.8741	0.0109	0.1077	0.8741
Unemployment rate	0.0001	0.0007	0.8758	0.0003	0.0042	0.8756	0.0004	0.0046	0.8756
Labor force participation rate	0.0002	0.0012	0.9972	0.0034	0.0167	1.0000	0.0036	0.0174	1.0000
Child dependency ratio	-0.0001	0.0019	0.3243	0.0013	0.0160	0.6960	0.0013	0.0167	0.6956
Old-age dependency ratio	-0.0141	0.0202	1.0000	-0.0254	0.0370	1.0000	-0.0395	0.0561	1.0000
Peripheriality	-0.0000	0.0000	0.6319	-0.0000	0.0001	0.6322	-0.0000	0.0001	0.6321
Accessibility road	-0.0000	0.0003	0.9928	-0.0001	0.0007	0.9907	-0.0001	0.0009	0.9921
Accessibility rail	-0.0000	0.0003	0.9542	-0.0001	0.0006	0.9505	-0.0001	0.0008	0.9526
Region with a seaport	0.0005	0.0067	0.9390	0.0021	0.0258	0.9390	0.0026	0.0299	0.9390
Region with an airport	-0.0001	0.0069	0.3575	0.0087	0.1004	0.6458	0.0086	0.1047	0.6458
Coastal region	0.0006	0.0080	0.9470	0.0024	0.0291	0.9470	0.0031	0.0340	0.9470
Capital city region	0.0295	0.1127	0.9888	0.0550	0.2145	0.9882	0.0845	0.3221	0.9885
Region with a large city	0.0019	0.0108	0.9950	0.0405	0.2419	0.9950	0.0424	0.2511	0.9950
Rural region	0.0005	0.0074	0.9109	0.0044	0.0630	0.9109	0.0049	0.0670	0.9109
Objective 1 region	0.0471	0.1310	0.9927	0.0960	0.2763	0.9906	0.1430	0.3887	0.9918
Border region	0.0005	0.0069	0.6733	0.0003	0.0224	0.6720	0.0007	0.0271	0.6720
EU-15 region	-0.0039	0.0336	1.0000	-0.0067	0.0600	1.0000	-0.0106	0.0935	1.0000
Pentagon region	-0.0020	0.0182	0.9892	-0.0097	0.0742	0.9896	-0.0117	0.0854	0.9896

Table 6: Simulation results

Variable	Incl. prob.	Mean	Incl. prob.	Mean	Incl. prob.	Mean
$\tilde{\boldsymbol{x}}_1$	1.0000	1.4999	1.0000	1.5012	1.0000	1.5032
$\tilde{\boldsymbol{x}}_2$	1.0000	2.0010	1.0000	1.9998	1.0000	2.0005
$\tilde{\boldsymbol{x}}_3$	1.0000	-0.4996	1.0000	-0.5002	1.0000	-0.5035
$\tilde{\boldsymbol{x}}_4$	0.0222	-0.0004	0.0109	-0.0001	0.0354	-0.0005
$\tilde{\boldsymbol{x}}_5$	0.0123	0.0001	0.0108	-0.0002	0.0172	-0.0007
$\tilde{\boldsymbol{x}}_6$	0.0247	0.0004	0.0172	0.0000	0.0101	0.0000
$\tilde{\boldsymbol{x}}_7$	0.0155	0.0000	0.0117	-0.0002	0.0113	0.0000
$\tilde{\boldsymbol{x}}_8$	0.0270	0.0009	0.0161	0.0001	0.0132	0.0002
$\tilde{\boldsymbol{x}}_9$	0.0225	-0.0001	0.0053	0.0000	0.0302	0.0010
$\tilde{\boldsymbol{x}}_{10}$	0.0232	0.0007	0.0147	0.0000	0.0267	0.0001
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_1$	0.9993	0.4998	0.9334	0.4851	0.9048	0.4494
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_2$	0.0126	-0.0008	0.1021	-0.0237	0.1063	-0.0449
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_3$	0.0188	-0.0001	0.0322	0.0017	0.0411	0.0074
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_4$	0.0225	-0.0004	0.0058	-0.0001	0.0190	0.0003
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_5$	0.0143	-0.0003	0.0180	-0.0011	0.0170	-0.0005
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_6$	0.0310	0.0004	0.0119	-0.0003	0.0046	0.0001
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_7$	0.0322	-0.0011	0.0123	0.0003	0.0123	-0.0010
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_8$	0.0214	0.0000	0.0195	-0.0014	0.0146	0.0003
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_9$	0.0178	-0.0008	0.0043	0.0002	0.0137	0.0009
$\tilde{\boldsymbol{W}}\tilde{\boldsymbol{x}}_{10}$	0.0121	0.0001	0.0251	0.0019	0.0149	-0.0014
α	1.0000	-0.9995	1.0000	-1.0126	1.0000	-1.0257
σ^2	1.0000	0.0610	1.0000	0.0625	1.0000	0.0629
$\tilde{\boldsymbol{W}}$	first-order contiguity		10-nearest neighbor		distance-based	
Correct $\tilde{\boldsymbol{W}}$	0.9999		0.9999		0.9999	