Habilitation Thesis

# Spatial econometric analysis with applications to regional macroeconomic dynamics

Tomáš Formánek

January 2019

Department of Econometrics
Faculty of Informatics and Statistics
University of Economics, Prague
nám. W. Churchilla 4
130 67 Praha 3

# Contents

# List of Figures

# List of Tables

# Acknowledgements

I would like to thank my long-term mentor prof. Ing. Roman Hušek, CSc. for his support and guidance.

I express deep gratitude to my colleagues, prof. RNDr. Ing. Michal Černý, Ph.D. and Ing. Miroslav Rada, Ph.D. Their valuable comments and suggestions truly helped to improve this text.

Tobler's first law of geography:

Everything is related to everything else, but near things are more related than distant things.

# Data source and copyright acknowledgement

## GISCO NUTS 2010, GISCO NUTS 2013

GISCO NUTS is a geographical dataset developed by the European Commission based on EuroBoundary Map (EBM) from EuroGeographics, complemented with the Global Administrative Units Layer (GAUL) from UN-FAO (for the Former Yugoslav Republic of Macedonia) and geometry from Turkstat (for Turkey).

When the GISCO NUTS geographical dataset is used in any printed or electronic publication, the source data set shall be acknowledged in the legend of the map and in the introductory pages of the publication:

Data source: GISCO – Eurostat (European Commission)
Administrative boundaries: © EuroGeographics, UN-FAO, Turkstat.

The GISCO NUTS 2010 & GISCO NUTS 2013 data were used as a source of all maps and geo-coded information used in this thesis.

Source:

`https://ec.europa.eu/eurostat/web/gisco/`

# Abstract

Regional macroeconomic processes may not be properly analyzed without accounting for their spatial nature: geographic distances and interactions between neighbors. State level and regional macroeconomic policy actions should be prepared, implemented and evaluated while accounting for the spatial nature of their effects: positive or negative spillovers may influence the desired outcome. Spatial econometrics is a versatile tool for a broad range of quantitative analyses performed with geo-coded (spatially defined) data. Over the last few years, both cross-sectional and panel data methods of spatial analysis have gained considerable attention in literature. While both types of data provide valuable insight and improve relevancy of quantitative analyses, spatial panels often bring useful advantages over cross-sectional spatial data in terms of tackling the temporal aspects of (macroeconomic or other) dynamics as well as by allowing to account for unit's individual effects.

This contribution to spatial analysis provides both methodological background and empirical applications of spatial econometric methods. The theoretically and methodologically focused chapters contain a comparative summary of estimation methods, corresponding tests and model interpretations. In contrast with most publications in the field, great emphasis is given to model robustness evaluation with respect to possible changes in the underlying spatial structure (both conceptual and parametric differences in spatial definitions are discussed). The empirical/application part of this contribution is focused on regional dynamics in macroeconomic processes within selected EU countries, with major emphasis on the Czech Republic and its neighbors.

**Keywords:** Spatial dependency, spatial econometric model, spatial panel data

# Preface

Spatial econometrics is a field of econometrics that explicitly deals with spatial interactions and spatial dependencies among geographically determined units. This contribution features both methodological background and empirical applications of spatial analysis and spatial econometric methods. Chapters 1 to 5 are theoretically and methodologically focused, encompassing a comparative summary of estimation methods, corresponding tests and model interpretations. Chapters 6 to 8 draw from published (submitted) empirical papers by the author and provide empirical analyses focused on regional dynamics in macroeconomic processes within selected EU countries.

## Structure of the thesis

Chapter 1 motivates spatial analysis methods and outlines basic terms and definitions. Stochastic spatial process (random field) is defined along with specific variability statistics (covariogram, variogram and semivariogram) and different types of stationarity assumptions. Also, a generalization of the spatial stochastic process and its features for panel data (spatio-temporal data) is included.

Chapter 2 focuses on basic tools, methods and underlying concepts of spatial econometrics. Spatial dependency (spatial autocorrelation) processes are motivated. Spatial structure & its setup are extensively discussed, with focus on various neighborhood definition methodologies. Main spatial dependency statistics and tests are introduced (Moran's $I$, Geary's $C$ and Getis' $G$).

Chapter 3 provides a detailed discussion of cross-sectional spatial econometric models: their setup, estimation methods and testing. Interpretation of an estimated spatial model is provided, along with specific topics (direct effects and spill-overs) and their caveats. Robustness of the estimators with respect to varying spatial structure is addressed in detail. Both fully parametric and semi-parametric approaches to spatial analysis are included.

Chapter 4 generalizes the topics and concepts introduced in chapter 3 to spatial panel data and spatial panel models. Different types of static spatial panel models are addressed here and both random effects and fixed effects assumptions are used.

Chapter 5 provides a brief overview of complex and advanced spatial panel models. Dynamic spatial panel models are described, as well as spatial panel data featuring a hierarchical spatial structure.

Chapter 6 uses Getis' spatial filtering to provide an empirical analysis of unemployment dynamics and its major constituent factors. 2015 data from 10 EU countries at the NUTS2 level are used and model robustness is evaluated. This chapter is largely based on a published contribution by Formánek and Hušek [38].

Chapter 7 focuses on macroeconomic convergence processes. Spatial panel data (years 2000 – 2015, 111 NUTS2 regions in 10 EU states) are used for "$\beta$-convergence" evaluation. Significant emphasis is given to the analysis of model robustness with respect to the underlying spatial structure. This chapter draws from a published contribution by Formánek [37].

Chapter 8 uses spatial panel data to analyze macroeconomic growth dynamics in selected EU countries. Regional (NUTS2) interdependencies and macroeconomic factors are considered and model robustness is systematically evaluated against changes in neighborhood definitions. Chapter 8 is based on a paper submitted by Formánek to the Journal of International Studies.

Chapter 9 contains brief final remarks.

# Spatial econometrics: methodology

# 1. Introduction

Spatial econometrics is a specialized field of econometrics that explicitly accounts for spatial interactions among geographical units. Spatial econometric models are often used to analyze regional macroeconomic processes and interactions. However, geographical units may be studied at multiple aggregation levels (from state-level to counties, etc.). Both cross-sectional and panel data can be used for quantitative spatial analyses.

Although the principle of spatial autocorrelation may resemble autocorrelation in time series (expanded to two dimensions), there are major differences: unlike the unidirectional dependency in time series, spatial units can affect each other mutually. Also, spatial arrangements often defy precise identification: there is seldom a unique, "right" way to define the correct spatial pattern for a given dataset or econometric model.

## 1.1. Brief history of spatial analysis

Spatial analysis can be traced back to early attempts at land surveying and cartography. Many science fields have contributed to establishing and developing the modern form of spatial analysis (astronomy, botanical studies, ecology, epidemiology, geology, etc.). However, as we focus on the statistical aspects of spatial analysis, John Snow and his disease mapping of 1854 is usually cited as the first historical application.

In August 1854, there was a major cholera outbreak in the Soho neighborhood of London, UK. There were 127 cholera related deaths around the area. At the time, germ theory (microorganisms causing disease) was not a generally accepted paradigm. Dr. John Snow (a physician and promoter of medical hygiene) spoke to local residents and mapped where the cholera cases occurred (see figure 1.1). Based on his map, he was able to pinpoint the public water pump on Broad Street as the source of contaminated water causing the cholera outbreak. Dr. Snow used statistics to find a relationship between water sources and cholera cases and subsequently found out that the waterworks company supplying water to Broad Street pump was taking water from a sewage polluted area of the Thames river.


Figure 1.1.: Map by Dr. John Snow, showing clusters of cholera cases in London. Source: Wikipedia.

More recently, in 1935, R.A. Fisher was the first to recognize the statistical implications of spatial dependency. In his work on design of experiments in agricultural science [35], he wrote:

> "After choosing the area we usually have no guidance beyond the widely verified fact that patches in close proximity are commonly more alike, as judged by the yield of crops, than those which are further apart."

Spatial variability, i.e. field-to-field changes in yields are largely due to physical properties of the soil and changes in environmental properties of the fields. In his analysis, Fisher devised a way to avoid confounding treatment effect with plot effect by introducing field randomization: his solution was to eliminate spatial dependency bias by localizing the crops under scrutiny into randomly assigned blocks.

Spatial econometrics has its roots in two texts, published in the 1950s: *Notes on continuous stochastic phenomena* by Moran [73] and *The contiguity ratio and statistical*

*mapping* by Geary [44]. However, the actual framework for contemporary applied spatial econometrics was provided by Cliff and Ord: many of their joint publications were published, starting in the late 1960s – see e.g. [21] or [22].

Most of the fundamental topics and methods relevant for spatial econometrics can be found in specialized textbooks, such as *Introduction to Spatial Econometrics* by LeSage and Pace [67]. As we look at contemporary methods and advancements in spatial econometrics, Elhorst [28] defines three generations of spatial econometric models: The first generation consists of models based on cross-sectional data (see also [39], [41] or [67]). The second generation comprises models based on static spatial panel data (see [71], [72] or [79]). Finally, the third generation encompasses dynamic spatial panel data models (see [28] or [33]).

## 1.2. Basic terms and topics of spatial analysis

The main focus of this text resides in spatial econometrics. However, it seems necessary to briefly outline some key underlying terms, concepts and definitions.

### Measuring spatial variables

Spatial measurements and measurement scales form a persistent issue in spatial analysis. In many practical applications, data vary continuously over space, but are measured only at discrete locations. Therefore, to characterize spatial features of the variables studied, some form of spatial aggregation is necessary. In economics (and other fields as well), the aggregation of observed spatial variables may be just another source of bias and potential data mis-manipulation. The area-specific summary values (unemployment rates, population densities, etc.) are influenced by both the shape and scale of the aggregation units. For example, a choropleth map showing euro-zone unemployment rates would look radically different if we plot individual counties instead of state-wide unemployment rates. Furthermore, districts and other administrative boundaries may change over time. Hence, scale, consistency and relevance should be carefully considered when collecting and analyzing spatial data.

In EU countries, consistency in geographic partitioning is well addressed by the Classification of Territorial Units for Statistics (NUTS) standards, developed by the EU (see `http://ec.europa.eu/eurostat/web/nuts`). Besides NUTS, smaller Local administrative units (LAUs) are also defined for EU member states and candidates. As NUTS and LAUs are EU-specific, their classification only covers the member states of the EU

in detail.

Beyond EU, multiple geographic & mapping datasets are freely available to the researchers (e.g. from Google Maps, `www.gadm.org`, etc.). However, consistency of the partitioning (hierarchical structuring into comparable areas) is often problematic. Finally, as we take into account repeated measurements and spatial panel data econometric models, temporal frequency may also play a significant role in determining data "quality" and relevance.

### Measuring spatial distances

With time series, there is only one dimension and its direction is set. In contrast, with spatial analysis, we usually use a two-dimensional space. For two given spatial units $s_i$ and $s_j$, direction can matter as well their distance. Distances ($d$) can be defined in a variety of ways, yet the following technical conditions should always apply (invariant to spatial translation, i.e. "shift"):

1. $d(s_i, s_j) = d(s_j, s_i)$ (symmetry)
2. $d(s_i, s_i) = 0$ (distance between a point and itself is zero)
3. $d(s_i, s_j) \leq d(s_i) + d(s_j)$ (triangle inequality; $d(s_i)$ is the distance from origin)

**Euclidean distances** are measured along straight lines between two point in the "ordinary" Euclidean space. In two dimensions, the Euclidean distance ($L_2$ norm) is defined as

$$d(s_i, s_j) = \sqrt{(s_{ix} - s_{jx})^2 + (s_{iy} - s_{jy})^2}\,, \tag{1.1}$$

where the $x$ and $y$ subscripts are used to handle planar coordinates. For many applications in spatial statistics and spatial econometrics, the computational simplicity of distances in the two-dimensional Euclidean space is remarkably attractive. For smaller distances, the non-planar latitude and longitude geographic coordinates are often projected to an Euclidean space as the so called local projection preserves distances [24].

For larger distances, planar projection accumulates non-negligible errors. In such circumstances, the **great circle distances** (shortest path between two points on a sphere given, their longitudes and latitudes) may be calculated using the haversine formula:

$$d(s_i, s_j) = 2r \arcsin \sqrt{\sin^2\left(\frac{\phi_j - \phi_i}{2}\right) + \cos(\phi_i)\cos(\phi_j)\sin^2\left(\frac{l_j - l_i}{2}\right)}\,, \tag{1.2}$$

where $r$ is the radius of the sphere, $\phi_1$ and $\phi_2$ are the latitudes of points $s_i$ and $s_j$ in radians, $l_i$ and $l_j$ are the longitudes (in radians). The formula (1.2) is only an approximation when applied to the Earth, which is not a perfect sphere. However, this approximation is sufficient as long as we can accept distance values that are correct within a 0.5% margin. For more accurate results that take Earth's ellipticity into consideration, we can revert e.g. to the Vicenty's formulae (see [24] for details).

**Manhattan distance** is a concept based on the well-known grid-like street geography of New York's Manhattan district. The Manhattan distance ($L_1$ norm) is a function on a fixed grid: it's the sum of horizontal and vertical components. For example, this is the driving distance between two points that a car has to cover while driving on streets, orthogonally intersecting at residential blocks.

Upon relevance in practical or theoretical applications, many additional definitions of distance are available for use in spatial statistics – see [24] for an exhaustive list and additional details.

### Spatial stochastic process – random field

A typical time-series based stochastic process may be denoted as $\{Z(t) : t \in T\}$ where $t$ is a time index from a one-dimensional index set $T$. Spatial stochastic process (random field) is a generalization of a stochastic process where the index set is not one-dimensional (e.g. time), but a higher dimensional Euclidean space (or a part of it). For a generic location $\boldsymbol{s}$ given by a vector of $d$ coordinates in a $d$-dimensional Euclidean space, spatial stochastic process is often denoted as

$$Z(\boldsymbol{s}) : \boldsymbol{s} \in D \subseteq \mathbb{R}^d\,. \tag{1.3}$$

Typically, $d = 2$ for most economic and econometric applications, $d = 3$ is often used in fields such as geology or astronomy. $D$ is a fixed finite set of $N$ spatial locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N$. Individual $\boldsymbol{s}_i$ units are points in space (say, with GPS-based latitude and longitude coordinates). Sometimes, such points can be associated with non-zero surface area elements – basically, they can serve as representative locations for anything from agricultural crop-fields to districts, counties, regions or even states. Much like in time-series analysis, the individual realization of a spatial stochastic process – random field – (1.3) are often denoted $z(\boldsymbol{s}_i)$ or, simply, $z_i$.

Observed spatial stochastic data may be either discrete or continuous, they may be observations at a given point in space or as spatial aggregations. The underlying spa-

tial structure may also be either continuous or discrete and both regular (chess-board like) and irregular spatial structures (NUTS regions) are commonly used in empirical applications.

### Stationarity of spatial processes, covariogram and (semi)variogram

The notion of stationarity in spatial stochastic processes may somewhat resemble stationarity in time series analysis. A common simplifying assumption that is made in spatial analysis is that the spatial process under scrutiny repeats itself over the domain $D$. Such process is said to be stationary. For a stationary process, the absolute coordinates at which we observe the process are unimportant. Orientated distances between the observed points provide sufficient information for analysis: if we translate the entire set of coordinates by a specific distance in a specified direction, the stochastic process and its features remain unchanged.

**Strong stationarity** of a spatial stochastic process may be formally defined as follows: We start with a given spatial stochastic process $Z(\boldsymbol{s})$ and data observations $\{z(\boldsymbol{s}_i) : i = 1, \ldots, m\}$, forming the following finite-dimensional distribution:

$$F_{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_m}(z_1, \ldots, z_m) = P[Z(\boldsymbol{s}_1) \leq z_1, Z(\boldsymbol{s}_2) \leq z_2, \ldots, Z(\boldsymbol{s}_m) \leq z_m] \,.$$

Strong stationarity means that $F$ is invariant under spatial translation $\boldsymbol{h}$. Unlike $d_{ij}$ (Euclidean distance between two spatial units $i$ and $j$), $\boldsymbol{h}$ is an orientated distance "shift" (spatial translation) vector. For strong stationarity,

$$\begin{aligned} &P\left[Z(\boldsymbol{s}_1) \leq z_1, Z(\boldsymbol{s}_2) \leq z_2, \ldots, Z(\boldsymbol{s}_m) \leq z_m\right] \\ &= P\left[Z(\boldsymbol{s}_1 + \boldsymbol{h}) \leq z_1, Z(\boldsymbol{s}_2 + \boldsymbol{h}) \leq z_2, \ldots, Z(\boldsymbol{s}_m + \boldsymbol{h}) \leq z_m\right] , \end{aligned} \tag{1.4}$$

has to hold for any spatial translation $\boldsymbol{h}$. This assumption is often too restrictive for real applications. For different types of empirical analyses, alternative (relaxed) types of stationarity are often sufficient. Before discussing weaker forms of spatial stationarity, we need to define the covariogram and (semi)variogram. Covariogram $C$ is the covariance between two spatial units: $C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \text{cov}[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)]$, semivariogram $\gamma$ is defined as $\gamma(\boldsymbol{s}_i, \boldsymbol{s}_j) = \text{var}[Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_i)]$ and the variogram is defined as $2\gamma$. Please note that $C$ and $\gamma$ definitions are very general – they do not require a stationary random field.

**Weak stationarity** (also called second order stationarity) assumes that the first two moments exist, are invariant (and finite) and covariance only depends on spatial translation (orientated distance) $\boldsymbol{h}$:

$$\begin{aligned} E[Z(\boldsymbol{s})] &= \mu\,, \\ \mathrm{var}[Z(\boldsymbol{s})] &= \sigma^2\,, \\ \mathrm{cov}[Z(\boldsymbol{s}+\boldsymbol{h}), Z(\boldsymbol{s})] &= C(\boldsymbol{s}+\boldsymbol{h}, \boldsymbol{s}) = C(\boldsymbol{h})\,. \end{aligned} \tag{1.5}$$

Here, because autocovariance is a function of $\boldsymbol{h}$ only (under weak stationarity), it follows that for any two spatial points $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ such that $\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}$, we can write:

$$\mathrm{cov}\left[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)\right] = C(\boldsymbol{s}_i - \boldsymbol{s}_j) = C(\boldsymbol{h})\,. \tag{1.6}$$

To summarize – under weak stationarity, the covariogram (spatial autocovariance) depends only on the difference between locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ and not on the locations themselves. Also, for $\boldsymbol{h} = \boldsymbol{0}$, the expression (1.6) simply describes variance:

$$\mathrm{cov}\left[Z(\boldsymbol{s}+\boldsymbol{0}), Z(\boldsymbol{s})\right] = C(\boldsymbol{0}) = \mathrm{var}\left[Z(\boldsymbol{s})\right].$$

A closely related "weak dependency" (by analogy to time series analysis) assumption is often used for empirical analysis of weakly stationary random fields. Under weak dependency, covariance between observations disappears with growing distance: $C(\boldsymbol{h}) \to 0$ as $||\boldsymbol{h}|| \to \boldsymbol{\infty}$.

**Intrinsic stationarity** is less restrictive than weak (second order) stationarity and it is defined in terms of first differences. A spatial process is intrinsically stationary if the difference between two observed spatial points is weakly stationary:

$$\begin{aligned} E[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})] &= 0\,, \\ \mathrm{var}[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})] &= 2\gamma(\boldsymbol{h})\,, \end{aligned} \tag{1.7}$$

where $2\gamma(\boldsymbol{h}) \geq 0$ is the variogram. For intrinsically stationary spatial processes, $2\gamma(\boldsymbol{h})$ is defined as:

$$\begin{aligned} 2\gamma(\boldsymbol{h}) &= \mathrm{var}[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})] \\ &= E\left\{\left([Z(\boldsymbol{s}+\boldsymbol{h}) - E(Z(\boldsymbol{s}+\boldsymbol{h}))] - [Z(\boldsymbol{s}) - E(Z(\boldsymbol{s}))]\right)^2\right\}. \end{aligned} \tag{1.8}$$

Generally, the value of (1.8) increases with growing oriented distance $\boldsymbol{h}$.

The two types of relaxed stationarity are related: weak stationarity implies intrinsic stationarity but not vice versa. Importantly, for weakly stationary spatial processes (where $E(Z(\boldsymbol{s}+\boldsymbol{h})) = E(Z(\boldsymbol{s})) = \mu$) the variogram (1.8) simplifies to:

$$2\gamma(\boldsymbol{h}) = E\left[(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2\right], \tag{1.9}$$

i.e. to the expected squared difference between two observed realizations of a spatial stochastic process.

The semivariogram is denoted as $\gamma(\boldsymbol{h})$ and it equals to half the variogram (i.e. expected squared difference between two spatial observations). Since (1.8) and (1.9) are expectations of a square, $\gamma(\boldsymbol{h}) \geq 0$ for both weakly and intrinsically stationary random fields. Also, at $\boldsymbol{h} = \mathbf{0}$, $\gamma(\mathbf{0}) = 0$ because

$$E\left[(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_i))^2\right] = 0 \text{ for } \forall\, i\,.$$

Also, it can be shown [23] that the variogram (semivariogram) is a generalization of the covariance function (1.6) and under weak stationarity, the two functions are related by expression (see Appendix A.2):

$$\gamma(\boldsymbol{h}) = C(\mathbf{0}) - C(\boldsymbol{h})\,. \tag{1.10}$$

If a stationary stochastic process has no spatial dependency at all (i.e. $C(\boldsymbol{h}) = 0$ for $\boldsymbol{h} \neq \mathbf{0}$), the semivariogram (1.10) is constant: $\gamma(\boldsymbol{h}) = \text{var}[Z(\boldsymbol{s})]$ everywhere, except for $\boldsymbol{h} = \mathbf{0}$, where $\gamma(\mathbf{0}) = 0$.

**Isotropic spatial process** may be defined through a semivariogram: $\gamma(\boldsymbol{h}) = \gamma(||\boldsymbol{h}||) = \gamma(d)$. Isotropy means that the semivariogram depends only on the distance $d$ between two points and not on direction. The lack of isotropy – anisotropy – means the semivariogram depends on direction as well as distance. To assess and test anisotropy, we can estimate and plot directional semivariograms (see [23] for discussion of this topic and corresponding tests).

### Empirical semivariogram

Although expression (1.6) carries useful information, most statisticians tend to favor semivariogram over the covariogram. The reasons are historical and – more importantly – the intrinsic stationarity conditions required for an empirical semivariogram are less restrictive than the weak stationarity required for empirical covariogram calculations

[69]. To perform empirical analysis of distance-based data correlations, we construct the so called empirical semivariogram as follows: First, we divide the distances observed over the domain $D$ into $K$ conveniently chosen intervals:

$$I_1 = (0, d_1],\ I_2 = (d_1, d_2],\ \ldots\ ,\ I_K = (d_{K-1}, d_K]\,.$$

Here, $d_1$ is the maximum distance within the $I_1$ interval and $d_K$ is the maximum distance observed over the field of data. The intervals can be proportional in terms of distance or in terms of sets of observation pairs allocated to each interval (to adjust for unevenly spaced observations). Although the interval setup may seem rather arbitrary, generally accepted rules (concerning interval setup, etc.) are available [57]. Please note that distances are determined by $d$ (distance magnitudes) only – here, we do not use the orientated distances $\boldsymbol{h}$.

Next, the empirical semivariogram is calculated using the following formula:

$$\hat{\gamma}(d_k) = \frac{1}{2N(d_k)} \sum_{N(d_k)} [Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)]^2\,, \tag{1.11}$$

where $N(d_k)$ is the number of distinct observation pairs in the interval $I_k$ and $\hat{\gamma}(d_k)$ is the semivariogram estimate for its corresponding group (interval) of distances.

Finally, we can fit a convenient parametric function (exponential, spherical, Gaussian, etc.) to the estimated $\hat{\gamma}(d_k)$ values – see [23] for details and examples. The main goal of empirical semivariogram construction is to estimate and visualize the spatial autocorrelation structure of the observed stochastic process. From figure 1.2, we can see three main features of an estimated empirical semivariogram:

- *nugget* (nugget effect) describes the micro-scale variations or measurement errors in data. Theoretically, at zero separation distance, $\gamma(0) = 0$. However, two factors play a role here: First, $\gamma(d_1)$ is estimated over the $N(d_1)$ set of pairs, i.e. for the first interval where $d_{ij} \in (0, d_1]$. Second, fitting the empirical semivariogram curve to observed values often causes the non-zero nugget.

- *sill* amounts to $\lim_{d\to\infty} \gamma(d)$. The sill corresponds to variance of the stochastic field at distances where spatial dependency (which reduces $\gamma(d)$) no longer applies. Using (1.10), we can see that $\lim_{d\to\infty} \gamma(d) = C(\mathbf{0}) = var[Z(\boldsymbol{s})]$.

- *range* is the spatial distance (if any) beyond which the data are not autocorrelated. In a way, range describes the strength of spatial structure – based on where the

semivariogram "reaches" its asymptote (sill).

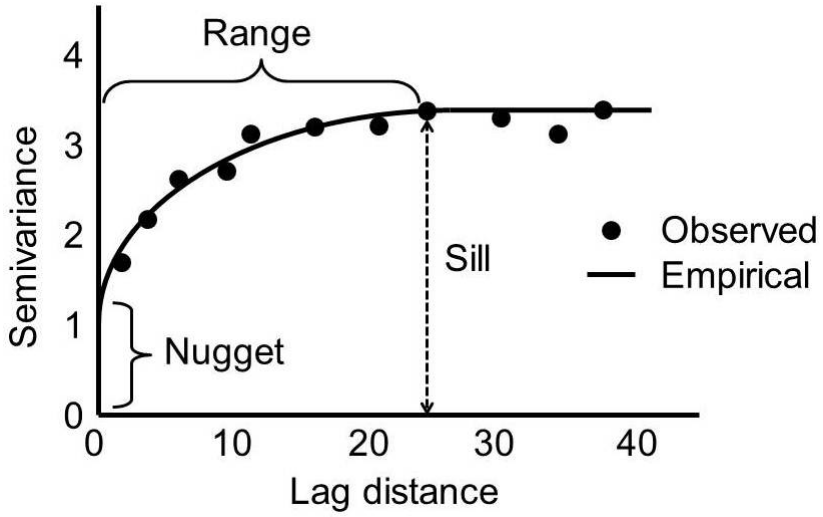


Figure 1.2.: Empirical semivariogram example. Source: [15].

Empirical semivariograms such as figure 1.2 reflect the fact that observations located close together are more alike than those far apart: increasing "variance" (semivariogram values) in pairwise differences along increasing $d$ means decreasing spatial dependency (i.e. there is an inverse relationship between distance and spatial autocorrelation).

## 1.3. Spatio-temporal data & analysis tools

This section generalizes the cross-sectional data & topics discussed in section 1.2 to accommodate processes that have both spatial and temporal dimensions.

Environmental, geophysical and socio-economic processes are often observed repeatedly over time. Such observations usually exhibit both spatial and temporal dependency and variability. Given the frequency and density limitations of empirical measurements of variables in continuous space and time, we often model our observations as realizations of a spatio-temporal random function (random field)

$$Z(\boldsymbol{s}, t), \qquad (\boldsymbol{s}, t) \in \mathbb{R}^d \times \mathbb{R}\,, \tag{1.12}$$

where the spatio-temporal domain is indexed in space by $\boldsymbol{s} \in \mathbb{R}^d$ and in time by $t \in \mathbb{R}$. The separation between spatial and time dimensions is substantial, which is reflected in the notation used in (1.12).

Assuming that second moments of of the spatio-temporal random field $Z(\boldsymbol{s}, t)$ exist and are finite, the covariance function for two arbitrary observations $Z(\boldsymbol{s}_1, t_1)$ and $Z(\boldsymbol{s}_2, t_2)$ is defined as

$$\begin{aligned} C(\boldsymbol{s}_1, \boldsymbol{s}_2; t_1, t_2) &= \operatorname{cov}[Z(\boldsymbol{s}_1, t_1); Z(\boldsymbol{s}_2, t_2)], \\ &= E\left[\{Z(\boldsymbol{s}_1, t_1) - E[Z(\boldsymbol{s}_1, t_1)]\}\{Z(\boldsymbol{s}_2, t_2) - E[Z(\boldsymbol{s}_2, t_2)]\}\right], \\ &\quad (\boldsymbol{s}_1, t_1), (\boldsymbol{s}_2, t_2) \in \mathbb{R}^d \times \mathbb{R}. \end{aligned} \tag{1.13}$$

For empirical analyses of spatio-temporal data, some simplifying assumptions are often necessary. Here, we shall briefly discuss the notions of stationarity, separability and (full) symmetry.

**Weak (second-order) stationarity:** $Z(\boldsymbol{s}, t)$ is weakly stationary in space and time if its mean function $E[Z(\boldsymbol{s}, t)]$ is constant for all $(\boldsymbol{s}, t)$ and its covariance function (1.13) only depends on (orientated) spatial and temporal distances:

$$\operatorname{cov}[Z(\boldsymbol{s}_1, t_1), Z(\boldsymbol{s}_2, t_2)] = C(\boldsymbol{s}_1 - \boldsymbol{s}_2, |t_1 - t_2|), \tag{1.14}$$

for all spatio-temporal coordinates $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ in $\mathbb{R}^d \times \mathbb{R}$. Hence, under weak stationarity, we may rewrite (1.14) for any arbitrarily chosen origin $(\boldsymbol{s}_0, t_0) \in \mathbb{R}^d \times \mathbb{R}$ as

$$C(\boldsymbol{s}_0, \boldsymbol{s}_0 + \boldsymbol{h}\,;\, t_0, t_0 + t) = C(\boldsymbol{h}, t), \tag{1.15}$$

where $\boldsymbol{h}$ and $t$ are the spatial and temporal distances, respectively [47, 69].

**Intrinsic stationarity:** is based on the traditional approach of differencing observed data in order to achieve a stationary process. The random field $Z(\boldsymbol{s}, t)$ as in (1.12) is intrinsically stationary in space and time (has stationary increments in space and time) if

$$Z(\boldsymbol{s}_0 + \boldsymbol{h}\,;\, t_0 + t) - Z(\boldsymbol{s}_0; t_0), \qquad (\boldsymbol{h}, t) \in \mathbb{R}^n \times \mathbb{R} \tag{1.16}$$

is a weakly (second-order) stationary spatio-temporal process – i.e. stationary in space and time for any origin $(\boldsymbol{s}_0, t_0)$. For an intrinsically stationary process $Z(\boldsymbol{s}, t)$, covariance might not be well defined (see [69]), but spatio-temporal semivariograms (STSV) is:

$$\gamma(\boldsymbol{h}; t) = \frac{1}{2}\operatorname{var}[Z(\boldsymbol{s}_0 + \boldsymbol{h}\,;\, t_0 + t) - Z(\boldsymbol{s}_0; t_0)], \qquad (\boldsymbol{h}, t) \in \mathbb{R}^d \times \mathbb{R}. \tag{1.17}$$

STSV (1.17) does not depend on the selection of origin $(\boldsymbol{s}_0, t_0) \in \mathbb{R}^d \times \mathbb{R}$ (under intrinsic stationarity). Also, for intrinsically stationary random fields $Z(\boldsymbol{s}, t)$, the STSV $\gamma(\boldsymbol{h}; t)$

is non-negative and $\gamma(\mathbf{0}; 0) = 0$. Empirical STSVs and corresponding fitting algorithms are described e.g. in [40] or [48], where both the theoretical derivation and empirical application are provided.

**Separability:** This additional simplifying assumption states that a random field $Z(\boldsymbol{s}, t)$ has a separable covariance if its spatio-temporal covariance can be factored into purely spatial and purely temporal components and

$$C\left[Z(\boldsymbol{s}_1, t_1), Z(\boldsymbol{s}_2, t_2)\right] = C_S(\boldsymbol{s}_1, \boldsymbol{s}_2) \cdot C_T(t_1, t_2) \tag{1.18}$$

holds for all spatio-temporal coordinates $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ in $\mathbb{R}^d \times \mathbb{R}$. As discussed in [47], this assumption allows for computationally efficient estimation (simple interpolation, i.e kriging, or other types analyses) and inference. For this reason, separable covariance assumption is commonly used even in situations where it isn't fully justifiable (see [69] for detailed discussion of separable spatio-temporal random fields).

**Full symmetry:** Spatio-temporal process $Z(\boldsymbol{s}, t)$ has a fully symmetric covariance if

$$\operatorname{cov}\left[Z(\boldsymbol{s}_1, t_1), Z(\boldsymbol{s}_2, t_2)\right] = \operatorname{cov}\left[Z(\boldsymbol{s}_1, t_2), Z(\boldsymbol{s}_2, t_1)\right] \tag{1.19}$$

for all spatio-temporal coordinates $(\boldsymbol{s}_1, t_1)$ and $(\boldsymbol{s}_2, t_2)$ in $\mathbb{R}^d \times \mathbb{R}$. Separability forms a special case of full symmetry and spatio-temporal processes that are not fully symmetric are not separable. Hence, tests for full symmetry can be used to reject separability [47].

**Compactly supported covariance** can be described as a spatio-temporal generalization of "weak dependency" as in time series analysis or "range" in a semivariogram (see Figure 1.2). $Z(\boldsymbol{s}, t)$ has a compactly supported covariance, if

$$\operatorname{cov}\left[Z(\boldsymbol{s}_1, t_1), Z(\boldsymbol{s}_2, t_2)\right] = 0\,,$$

whenever $||\boldsymbol{s}_1 - \boldsymbol{s}_2||$, $|t_1 - t_2|$ or both are sufficiently large. Compactly supported covariances allow for computationally efficient spatio-temporal analyses and predictions over large datasets. Additional in-depth aspects of spatio-temporal data analysis are covered in [47] or [69] and the estimation toolbox for R is provided in [77].

# 2. Spatial econometrics: basic tools and methods

Econometrics has evolved as a separate discipline of statistics (mathematical statistics) and it typically uses non-experimental (natural, empirical) data for estimating economic & socio-economic relationships, testing economic theories and other less structured assumptions, forecasting and for evaluating economic policies. Spatial econometrics is a specialized field of econometrics that combines methods of spatial statistics with economic theories and observed (geo-coded) data in order to perform quantitative analyses and related tasks.

The empirical existence of simultaneous spatial dependencies in observed data is the central driving factor that justifies the use of spatial autoregressive models. Spatial econometric models account for the presence of spatial effects (such as economic spillovers) when analyzing the relationships between variables through regression models and other related estimation methods. Spatial quantitative models play an ever more important role in regional macroeconomic and social analyses, real estate studies, agricultural and ecological applications, epidemiology and in many other non-economic fields of research.

This chapter provides a thorough overview of key topics and selected methods in spatial econometrics. Besides including core methods, the selection is tailored to provide theoretical background for subsequent (empirically oriented) chapters. The overview provided is not exhaustive – for additional topics, please refer to [16], [28], [64] or [67] and the literature referenced therein.

## 2.1. Spatial dependency

In spatial econometrics, data are associated with a particular position in space. Data are geo-coded using the latitude/longitude geographic coordinates system, distances and common borders are used for estimation of spatial dependencies. Spatial data can be observed either at point locations (housing data, air pollution measurements, street

traffic, etc.) or aggregated over regular or irregular areas (e.g., countries, regions, states, counties). Besides geo-coded cross sections, we often use spatial panel data and methods, provided that cross sectional data observations can be consistently repeated over time.

As observed variables are combined with spatial definition, we may draw conclusions about similarities or dissimilarities between spatially close objects. Fotheringham et al. [41] define spatial dependency as follows:

> "Spatial dependency is the extent to which the value of an attribute in one location depends on the values of the attribute in nearby locations."

Similarly, Legendre [65] defines spatial autocorrelation as

> "... the property of random variables taking values, at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations."

Various descriptions and definitions of spatial dependencies in observed data exist, often with emphasis on different aspects of the phenomena, due to different research scopes (say, ecology vs. housing prices). However, for many empirical applications – including microeconomic and macroeconomic analyses – spatial autocorrelations play an important role and we need to adjust our theories and models to incorporate spatial aspects. Figure 2.1 provides a simple illustration of spatial autocorrelation in observed 2014-unemployment data: for the six countries considered (Czechia, Slovakia, Poland, Germany, Austria and Hungary), NUTS2-level unemployment rates are clearly "clustered", with distinctive regional patterns and with noticeable spatial autocorrelation.

It may be argued that much of the spatial effects and dependencies are attributable to omitted variable factors. However, spatial autocorrelation may be conveniently interpreted as a proxy for numerous real (theoretically sound), yet practically unobservable spatial effects. Many spatial interactions and their dynamic features are very difficult to explicitly define and properly structure in a way that would facilitate informative and harmonized quantification. This applies to tasks such as consistently measuring knowledge and skills diffusion (in labor-productivity models) or cross-border work commuting intensity and preferences (in unemployment-describing models), accounting for administrative/qualification employment barriers between countries, quantifying the impact of language differences, considering aerial distances vs. means of transportation, etc. For many such variables, even if measurements are possible, they would inherently

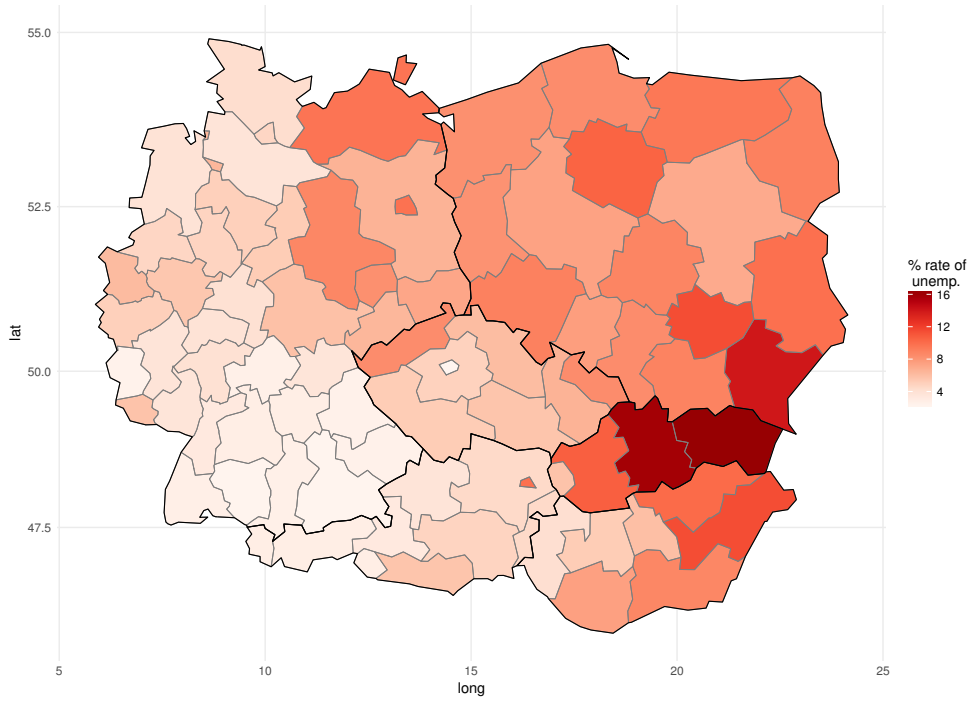


Figure 2.1.: Unemployment rates, 2014, NUTS2 regions. Source: Own calculation using GISCO – Eurostat data.

introduce many subjective choices and – in practical terms – many disputable features to quantitative models.

Spatial models often provide an intuitive, easily interpretable and functional approach towards regional (macroeconomic) data analysis. Different authors postulate diverse motivations and theoretic grounds for studying spatial effects, spatial dynamics and dependencies. Some of the most common factors [67] driving spatial correlation may be summarized as follows:

- *Omitted variables* motivation has been discussed in the preceding paragraph. Many unobservable (latent) factors and location-related features such as highway accessibility or neighborhood prestige may significantly influence the observed geo-coded variables. In practice, it is unlikely that appropriate observable explanatory variables would be available to accurately describe such influences.

- *Time-dependency* motivation is based on the premise that agents make decisions that are influenced by the behavior of other agents in previous periods. For example, local/regional/state authorities may set taxes or subsidies that reflect such policy actions taken by their neighbors in previous periods. Similarly, at the individual level, house selling prices are often influenced by past selling prices of neighboring houses (after controlling for other important factors such as surface area and the number of bedrooms).

- *Spatial heterogeneity* motivation is largely based on panel data methods and regression models. Within the panel data framework, we use individual effects (individual heterogeneity) that may be treated and interpreted as separate intercepts for each cross-sectional unit. For spatial panel data (where geo-coded units are observed for a number of repeated time periods), we may often conclude that spatially close units exhibit more similar individual effects as compared to non-neighboring units.

- *Externalities-based* motivation comes from a well-established economic concept: individuals and regions may be subject to (both positive and negative) consequences of economic activities exercised by unrelated third parties. Air pollution emitted by a factory that spoils the surrounding environment affects life quality in nearby residential areas and reduces property values is an example of a negative externality. On the other hand, beautifully landscaped parks may have a positive effect on the values of houses in the neighborhood.

- *Model uncertainty* motivation: spatial autocorrelation may be used in circumstances where we face uncertainty in terms of specifying a proper data generating process (DGP). For example, in a regression model environment, estimation and forecasting efficiency may often be improved by introducing spatial autocorrelation to the regression – this applies to both the dependent variable and regressors as well as to model errors.

In most empirical applications, finding the correct (most appropriate) motivation for an observed spatial dependency is complicated. Partly, this is due to the fact that different motivations are not mutually exclusive. Fortunately, this "identification problem" rarely causes complications in empirical analyses. For a detailed overview and spatial dependence taxonomy, see e.g. LeSage and Pace [67].

## 2.2. Neighbors: spatial and spatial weights matrices

Two spatial units are considered neighbors if they are "close" enough in space (see discussion next) to interact in terms of the associated (spatially defined) stochastic processes. Spatial connectivity matrices $\boldsymbol{S}$ are based on dummy variables: the $s_{ij}$ elements of $\boldsymbol{S}$ equal 1 if the two spatial units $i$ and $j$ are neighbors and 0 otherwise. Diagonal elements of $\boldsymbol{S}$ are set to zero by definition: units are not neighbors to themselves. Individual elements of the symmetric spatial matrix $\boldsymbol{S}$ may be formally outlined as follows:

$$s_{ij} = s_{ji} = \begin{cases} 0 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \neq j \text{ and regions } i \text{ and } j \text{ are not neighbors,} \\ 1 & \text{if} \quad i \neq j \text{ and regions } i \text{ and } j \text{ are neighbors.} \end{cases} \tag{2.1}$$

The elements of $\boldsymbol{S}$ are co-determined by the ordering of the data (spatial units), which can be arbitrary. A simple 4-unit (4×4) example is provided next:

$$\boldsymbol{S} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}. \tag{2.2}$$

From the first row (and column) of $\boldsymbol{S}$ in (2.2), we may observe that the first unit (say, region or city) is a neighbor of units 2, 3 and 4; the second row shows that unit 2 is a neighbor of units 1 and 3 (not a neighbor of unit 4), etc.

Cliff and Ord in [21], [22] have introduced a relatively flexible toolbox for spatial weights specification. Spatial weights are usually calculated in a two-step approach: First, a square spatial connectivity matrix $\boldsymbol{S}$ is established for a given set of $N$ spatial (geocoded) units. Next, a corresponding spatial weights matrix $\boldsymbol{W}$ is constructed by row standardization (scaling to unity), for use in spatial models such as (3.2) or (4.1). For example, from the spatial matrix $\boldsymbol{S}$ in (2.2), we may construct the spatial weights matrix $\boldsymbol{W}$ as follows:

$$\boldsymbol{W} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}. \tag{2.3}$$

Individual $\boldsymbol{W}$ elements $w_{ij}$ reflect the relationship intensity between cross sectional units $i$ and $j$. This topic is described in detail along equation (2.6).

As the number of spatial units and the dimension of $\boldsymbol{S}$ increase, we need to limit geo-dependencies to a manageable (computable) degree. This can be done through a simple stability condition stating that the correlation between two spatial units should converge to zero as their distance increases to infinity.

Elhorst [28] provides two alternatives stability conditions that may be restated as follows: (a) The row (and column) sums of any $\boldsymbol{S}$ matrix should be uniformly bounded in absolute value as the number of spatial units goes to infinity. (b) The row (and column) sums of $\boldsymbol{S}$ should not diverge to infinity at a rate equal to or faster than the rate of sample size growth. Condition (b) is more general – if (a) hold, (b) is implied but not vice-versa. Section 3.1 contains a formal discussion of this topic.

Different neighborhood definitions can be used for establishing $\boldsymbol{S}$ matrices and the corresponding weights matrices $\boldsymbol{W}$. The most common approaches to defining neighbors for spatial units are outlined next.

### Contiguity-based neighbors

Contiguity approach is a theoretically simple (yet computationally convoluted) rule, defining two units as neighbors if they share a common border. A generalization of this approach is based on the premise that a "second order" neighbor is the neighbor of a first order neighbor (the actual contiguous neighbor). With this type of approach, we can define a maximum neighborhood lag (order) to control for the highest accepted number of neighbors traversed (not permitting cycles) while determining the neighborhood of the spatial unit under scrutiny.

Computational convolutions of the contiguity approach are due to small, yet frequent topological inaccuracies in empirical maps (geo-data): spatial polygons may suffer from different types of errors such as intersecting or diverging boundaries. Various discrepancies may arise when spatial data are collected from different sources. Also, if a boundary between two units lies along a median line of a river channel, then the polygons of each unit would likely stop at the channel banks on each side. As a result, borders of such river-separated regions are not actually contiguous – there is a non-zero distance between them, corresponding to the width of the separating river (creek, lake, etc.). There are many such minor factors that complicate the unambiguous and automated (i.e. programmable) contiguity evaluation process. Therefore, heuristic approaches to

evaluation of contiguity are often required [16].

If regular spatial patterns are used (chessboard-like tiles), we often distinguish between queen and rook contiguity definitions (their names come from the movements of chess pieces). Rook is a more stringent definition of polygon contiguity than queen. For rook, the shared border must be of some non-zero length, whereas for queen the shared border can be as small as one point. This contiguity is sometimes generalized to natural map patterns: for example, using the queen rule, Arizona and Colorado are neighbors. Also, contiguity-based neighborhood evaluation is somewhat specific for "hole" regions. In the EU, there are several such NUTS2 level regions. In figure 2.1, we can see that the region DE30 (Berlin) lies inside the region DE40 (Brandenburg). Similarly, CZ01 (Praha) is located within CZ02 (Stredni Cechy) and AT13 (Wien) is encompassed by AT12 (Niederosterreich). For such regions, we either use the generalized (lag) approach to contiguity or we turn to distance-based neighborhood definitions.

### Distance-based neighbors

By adopting the distance-based approach, we construct spatial matrices by defining two units as neighbors if their distance does not exceed some ad-hoc predefined threshold. Formally, individual elements of $\boldsymbol{S}$ may be defined as follows:

$$s_{ij} = s_{ji} = \begin{cases} 0 & \text{if} \quad i = j, \\ 0 & \text{if} \quad h_{ij} > \tau, \\ 1 & \text{if} \quad h_{ij} \leq \tau, \end{cases} \tag{2.4}$$

where $h_{ij}$ is some adequate measure of distance between units' representative location points (centroids) and $\tau$ is an ad-hoc defined maximum neighbor distance threshold.

Distances between regions as in (2.4) are measured using centroids – conveniently chosen representative positions. Depending on model focus, data availability and researcher's individual preferences, centroids may be pure geographical center points (as in figure 2.2), locations of main cities, population-based weighted positions, transportation network based (highway/railway infrastructure, work-commuting intensities), etc.

Centroid-based distances are usually easy to calculate and evaluate against a chosen threshold $\tau$ . However, there are two important issues that need to be considered: This approach can generate "islands" (units with zero neighbors), unless the defined threshold is greater than the maximum of first nearest neighbor distances as measured across all units in the sample. Formally, the following conditions is sufficient to avoid islands (zero

rows/columns in the $\boldsymbol{S}$ matrix):

$$\tau \geq \max_i \min_j \{h_{ij} \,|\, h_{ij} > 0\}\,.$$

Also, the threshold-based approach might be less convenient for analysis of regions with uneven geographical density, i.e. with unequal sizes of units and distances between them.

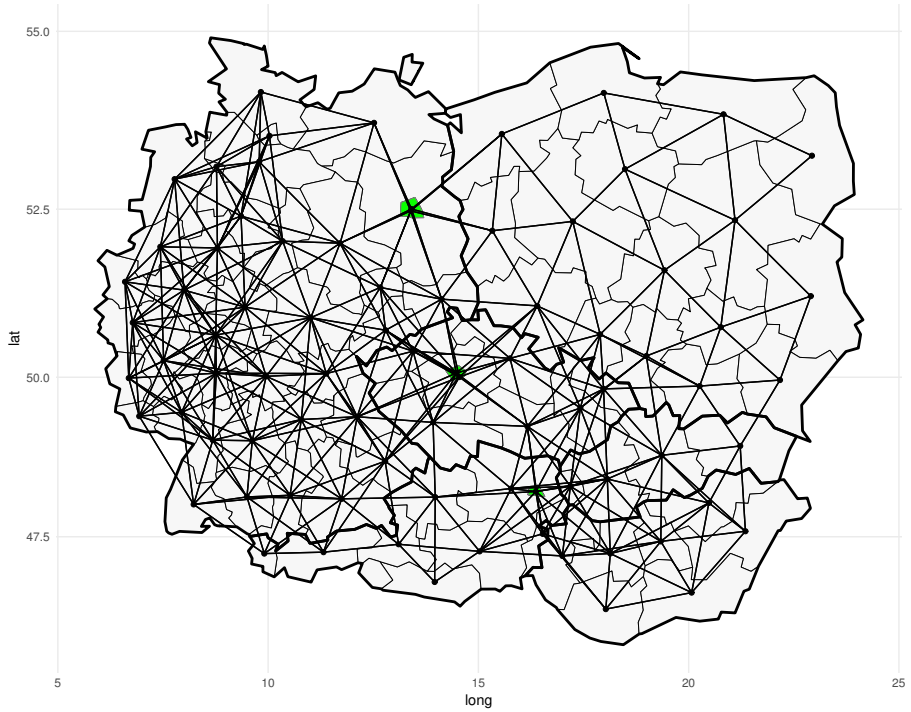


Figure 2.2.: Distance-based neighbors, 200 km threshold. Source: Own calculation using GISCO – Eurostat data.

Using the same region as in figure 2.1, an illustration showing NUTS2 regions and their neighborhoods is provided in figure 2.2 (pure geographical centroids of neighboring regions are connected by lines). We may observe the uneven regional density by comparing the complexity of the neighborhood connections in the western parts of Germany against the sparse north-eastern regions in Poland. Geographical heterogeneity of the regions in figure 2.2 is due to the fact that NUTS2 regions are bounded in terms of the number of their inhabitants (800,000 to 3 million) and there are prominent differences among geographical areas of Germany (densely populated small regions) on one hand and the

NUTS2 units in Poland and Hungary on the other hand.

The symmetric $\boldsymbol{S}$ matrix ($82 \times 82$) used to render figure 2.2 is omitted here, yet it may be briefly described as follows: given the maximum neighbor distance threshold of 200 km, the average number of neighbors is 8.85, PL34 (Podlaskie) is the the least connected region with 2 neighbors only while DE72 (Giessen) is the most connected with 17 neighbors.

### $k$-nearest neighbors

To define neighbors, we can apply the $k$-nearest neighbors ($k$NN) approach: for each spatial unit, we search for a preset number of $k$ nearest units that we define as its neighbors. This method conveniently solves for differences in areal densities ($k$ neighbors are ensured for each unit), yet it usually leads to asymmetric spatial matrices with potentially flawed neighborhood interpretation (simple transformation algorithms for asymmetric spatial matrices are available, e.g. from [16]). The symmetry of spatial matrix $\boldsymbol{S}$ has a strong impact on subsequent spatial econometric analysis. For a symmetric matrix, all eigenvalues are real. Importantly, this holds even after row standardization (i.e. for $\boldsymbol{W}$) – see chapter 3.3 for detailed discussion.

Also, it should be noted that under the $k$NN approach, individual $\boldsymbol{S}$ and $\boldsymbol{W}$ elements will depend on sample size. As we remove or add one or more spatial units to our sample (e.g. by including new country or region of interest), the group of $k$ nearest neighbors for each unit in the sample may change significantly – leading to potentially significant changes in the estimated spatial dynamics.

### Software and data for spatial analysis

Spatial matrix construction often requires extensive geographical datasets and specialized software. Fortunately, many such tools are freely available. Geodata for all countries and most of their administrative areas at different aggregation levels are available from GADM: `www.gadm.org`. For EU countries, a complete and consistent set of geodata may be obtained from Eurostat's GISCO: the Geographic information system of the commission: `http://ec.europa.eu/eurostat/web/gisco/`. Data analysis combining geographical and economic (environmental, epidemiology, etc.) information may be conveniently performed using the free and open source environments such as `R`: `www.r-project.org`, Python: `www.python.org` or Octave: `www.gnu.org/software/octave/`. From the category of comercially available software packages, MatLab: `www.mathworks.com/products/matlab` and Stata: `www.stata.com` feature tools

for estimation of spatial models. Unless stated otherwise, all examples, figures and empirical analyses presented here are produced using R software and Eurostat datasets (both geographic and macroeconomic).

## Spatial weights matrices

Construction of a spatial weights matrix $\boldsymbol{W}$ is based on row-standardizing the spatial connectivity matrix $\boldsymbol{S}$ (with $s_{ij}$ elements as binary neighborhood indicators), so that all rows in $\boldsymbol{W}$ sum to unity. As a direct consequence of this transformation, all elements of $\boldsymbol{W}$ in a given row lie within the $[0, 1]$ interval and can be used to calculate spatially determined expected values of $y_i$. The spatial lag (spatially determined expectation) for an $i$-th element of $\boldsymbol{y}$ is given by

$$\text{SpatialLag}(y_i) = \boldsymbol{w}_i \boldsymbol{y}\,, \tag{2.5}$$

where $\boldsymbol{w}_i$ is the $i$-th row of $\boldsymbol{W}$. Say, $\boldsymbol{y}$ is a 4-element vector with spatial properties determined by $\boldsymbol{S}$ and $\boldsymbol{W}$ as in expression (2.2). Hence, expanding on our sandbox example given by matrices (2.2) – (2.3), the spatial lags of $\boldsymbol{y}$ may be written as

$$\text{SpatialLag}(\boldsymbol{y}) = \boldsymbol{W}\boldsymbol{y} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \frac{1}{3}y_2 + \frac{1}{3}y_3 + \frac{1}{3}y_4 \\ \frac{1}{2}y_1 + \frac{1}{2}y_3 \\ \frac{1}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}y_4 \\ \frac{1}{2}y_1 + \frac{1}{2}y_3 \end{pmatrix}. \tag{2.6}$$

Note that the row elements of $\boldsymbol{W}$ display the impact on a particular spatial unit, constituted by all other units. The weighting operation shown in (2.6) can be interpreted as averaging across observations in neighboring units. Similarly, column elements in $\boldsymbol{W}$ describe the impact of a given unit on all other units. Because each row of $\boldsymbol{W}$ is normalized by a different factor, spatial weights are often asymmetric: the impact weight of unit $i$ on unit $j$ is not always the same as of unit $j$ on $i$.

## Moran plot

Figure 2.3 follows from the empirical example introduced in figures 2.1 and 2.2. Here, the observed values of unemployment are plotted against their spatial lags – this plot is often referred to as Moran plot (Moran scatter-plot). The $\boldsymbol{W}$ matrix (82×82) used in spatial lag (2.5) calculation of unemployment for figure 2.3 comes from the neighborhood definition as shown in figure 2.2. We can see that the scatter-plot "pairs" are well aligned along the

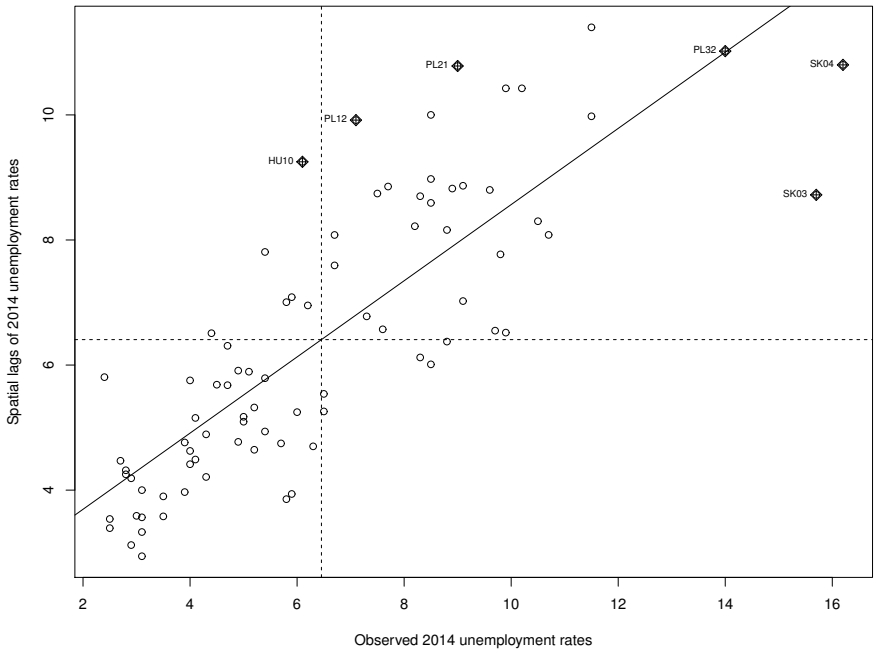


Figure 2.3.: Moran plot for unemployment rate, 2014: observed values vs. spatial lags. Source: Own calculation.

"regression line". This provides visual evidence for a significant spatial autocorrelation in the data. See [4] for additional empirical discussion and spatial lag evaluation.

### Generalized weights matrices

Spatial lag construction as in expressions (2.5) and (2.6) is straightforward. However, with increasing variance in units' neighbor-count (e.g. for distance-based neighbors with uneven geographical density), this widely adopted approach suffers from allocating uneven weights (influence), based on the number of neigbors of a given unit. To overcome this drawback, sometimes the non-zero elements in $\boldsymbol{W}$ are "generalized" before the row-standardization.

Distances to neighbors can be used to reflect some prior information concerning the spatial dependency processes: often we assume that spatial influences are inversely proportional to distances (linear, quadratic or other functional forms of influence decay may be used). For example, $\boldsymbol{W}$ construction may be based on a "truncated distance matrix"

$\boldsymbol{C}$, defined as

$$\boldsymbol{C} = \boldsymbol{S} \circ \boldsymbol{H} \,,$$

where $\boldsymbol{S}$ and its elements are defined in (2.4), $\boldsymbol{H}$ contains pairwise $h_{ij}$ distances and $\circ$ is the Hadamard (element-wise) product. Hence, $\boldsymbol{C}$ is a non-negative symmetric matrix with zeros on the main diagonal and its individual $c_{ij}$ elements equal either 0 or $h_{ij}$, depending on whether $s_i$ and $s_j$ are neighbors. Different transformations of $c_{ij}$ elements may be used to produce the $\boldsymbol{W}$ weights matrix. Using prior information regarding the inverse relationship between distance and interaction intensity, $w_{ij}$ elements may be based on transformed non-zero $c_{ij}$ elements: $(1/c_{ij})$, $(1/c_{ij}^2)$, $(1/\log c_{ij})$, etc. are often used for row-standardization while keeping the zero elements from $\boldsymbol{C}$. If $h_{ij}$ describes interaction intensity (e.g. commuting volume) instead of distance, $\boldsymbol{W}$ elements may be given as $w_{ij} = c_{ij} / \sum_j c_{ij}$.

The above described approach has been empirically verified in many applications [65]. For example, when analyzing employment/unemployment dynamics, labor force commuting habit dynamics in densely vs. sparsely populated areas may be modeled substantially better using this approach. However, the efficiency of any such $\boldsymbol{W}$ generalization crucially depends on the accuracy and validity of the prior information (decay pattern) used.

Many additional alternatives exist for the classical approach to $\boldsymbol{W}$ construction through individual row standardization of $\boldsymbol{S}$, described by expression (2.2). For example, Griffith et al. [51] use a single-factor normalization – see expression (3.27) and corresponding discussion for details.

## 2.3. Sample selection in spatial data analysis

Spatially autocorrelated processes are defined in terms of individual units and their interaction with corresponding neighbors. Clearly, we can only assess the impact of neighboring units if such units are part of our sample. Hence, in spatial econometrics, we usually do not draw limited samples from a particular area.

Instead, we work with cross-sectional (or spatial panel) data from adjacent units located in unbroken ("complete") study areas. Otherwise, $\boldsymbol{S}$ and $\boldsymbol{W}$ matrices would be misleading and we could not consistently estimate spatial interactions and effects. Generally speaking, spatial analysis should include the whole geographically defined area/region instead of using random sampling (from a "population" of regions within the relevant area).

## 2.4. Spatial dependency tests

Two basic types of spatial dependencies exist (as opposed to spatial randomness): positive spatial autocorrelation occurs if high or low values of a variable cluster in space. For negative spatial autocorrelation, spatial units tend to be surrounded by neighbors with very dissimilar observations. Sometimes, spatial dependency patterns are easy to discern visually using choropleths such as figure 2.1. However, a formal approach towards evaluation of spatial dependency is often required.

Before the actual estimation of spatially augmented econometric models, we should apply preliminary tests for spatial autocorrelation in the observed data. Many types of spatial autocorrelation test statistics are available, such as those presented by Anselin and Rey in [6]. Here, we only focus on the most used statistics for cross-sectional data as introduced by Moran, Geary and Getis.

### Moran's $I$

First introduced by Moran in [73], Moran's $I$ is a measure of global spatial autocorrelation that describes the overall clustering of the data:

$$I = \frac{N}{W} \boldsymbol{z}' \boldsymbol{W} \boldsymbol{z} (\boldsymbol{z}' \boldsymbol{z})^{-1}, \tag{2.7}$$

where $N$ describes the number of spatial observations (units) of the variable under scrutiny (say, $y$), $\boldsymbol{z}$ is the centered form of $\boldsymbol{y}$; it is a vector of deviations of the variable of interest with respect to its sample mean value such that $z_i = y_i - \bar{y}$. The standardization factor $W = \sum_i \sum_j w_{ij}$ corresponds to the sum of all elements of the spatial weights matrix $\boldsymbol{W}$. For row-standardized $\boldsymbol{W}$ matrices, $\frac{N}{W} = 1$. However, in its original form, Moran's $I$ does not require row-standardized weights. Instead of $\boldsymbol{W}$, we might use the spatial matrix $\boldsymbol{S}$ in expression (2.7) as well.

In most empirical circumstances, $I \in [-1, 1]$. The actual lower and upper bounds to $I$ are given by $(N/\boldsymbol{\iota}'\boldsymbol{W}\boldsymbol{\iota})\kappa_{\min}$ and $(N/\boldsymbol{\iota}'\boldsymbol{W}\boldsymbol{\iota})\kappa_{\max}$ where $\kappa_{\min}$, $\kappa_{\max}$ are extreme eigenvalues[1] of the double-centered connectivity matrix

$$\boldsymbol{\Omega} = (\boldsymbol{I}_N - \frac{1}{N} \boldsymbol{\iota}_N \boldsymbol{\iota}_N') \, \boldsymbol{S} \, (\boldsymbol{I}_N - \frac{1}{N} \boldsymbol{\iota}_N \boldsymbol{\iota}_N') \,,$$

[1] Here, as well as in equation (3.29), etc., we deviate from the the common notation $\lambda$ for eigenvalues as used in linear algebra and use $\kappa$ instead. Throughout this text, $\lambda$ is used as a spatial autocorrelation coefficient.

where $\boldsymbol{\iota}$ is a $(N \times 1)$ vector of ones. If $y_i$ observations follow *iid* normal distribution (i.e. under the null hypothesis of spatial randomness), Moran's $I$ is asymptotically normally distributed with the following first two moments (see [25] or [83] for derivation):

$$E(I) = -\frac{1}{N-1} \tag{2.8}$$

and

$$\text{var}(I) = \frac{N^2 W_1 - N W_2 + 3W^2}{(N^2-1)W^2}\,, \tag{2.9}$$

where $W$ comes from (2.7), $W_1 = \sum_i \sum_j (w_{ij} + w_{ji})^2$ and $W_2 = \sum_i (\sum_j w_{ij} + \sum_j w_{ji})^2$. Given the normality assumption in $y_i$, we can calculate a $z$-score

$$z = \frac{I - E(I)}{\sqrt{\text{var}(I)}}\,, \tag{2.10}$$

test for statistical significance of Moran's $I$ statistic (2.7): whether neighboring units are more similar ($I > E(I)$) or more dissimilar ($I < E(I)$) than they would be under the null hypothesis of spatial randomness.

Kelejian and Prucha [62] have demonstrated that standardized Moran's $I$ has an asymptotically normal distribution under various assumptions on $y_i$ variables: they provide a more general set of expressions (2.7) – (2.9) where sample normality of Moran's $I$ $z$-score holds for a variety of important variable types: $y_i$ can be dichotomous, polychotomous (multinomial) or count variable, as well as "corner-solution response" (see [85] for description of Tobit-type models).

Moran's $I$ spatial dependency analysis yields only one statistic that summarizes the nature of spatial dependency in the observed variable. In other words, Moran's $I$ as in (2.7) assumes geographical homogeneity (stationarity) in the data. If such assumption does not hold and the actual spatial dependency patterns vary over space, then Moran's $I$ test loses power and the "global" statistic (2.7) is non-descriptive.

The fact that Moran's $I$ is a summation of individual crossproducts (not outright apparent from the matrix notation in (2.7), see [2] for derivation) is exploited in an alternative spatial dependency test based on the Local Moran's $I$ statistic (row-standardized $\boldsymbol{W}$ assumed):

$$I_i = \frac{z_i N}{\boldsymbol{z}'\boldsymbol{z}} \boldsymbol{w}_i \boldsymbol{z}\,. \tag{2.11}$$

The expected value of Local Moran's $I$ under the null hypothesis of no spatial autocorrelation is: $E(I_i) = -w_i/(N-1)$. Here, $w_i$ is the sum of elements in the $i$-th row of $\boldsymbol{W}$.

For row-standardized weights matrices, $w_i = 1$. Values of $I_i > E(I_i)$ indicate positive spatial autocorrelation, i.e. that the $i$-th region is surrounded by regions that, on average, are similar to the $i$-th region with respect to the observed variable $y$. $I_i < E(I_i)$ would suggest negative spatial autocorrelation: on average, the $i$-th region is surrounded by regions that are different with respect to the observed variable. Local Moran's $I$ values as in (2.11) are calculated for each spatial unit and the statistical significance of spatial dependency is then evaluated using $var(I_i)$ and the corresponding $z$-score [2]. By comparing (2.7) and (2.11), we may see the global nature of Morans $I$ from

$$I = \frac{1}{N} \sum_{i=1}^{N} I_i \,. \tag{2.12}$$

Moran's $I$ (2.7) is often used for testing spatial dependency in regression model residuals. Please note that $z_i = y_i - \bar{y}$ from (2.7) may be recast as a residual part from a trivial regression model $y_i = \beta_0 + z_i$, where $\beta_0$ is the intercept ($\hat{\beta}_0 = \bar{y}$) and $z_i$ is the random element. Once the trivial model is expanded by a convenient set of regressors, Moran's $I$ can be used for testing regression residuals [21].

### Geary's $C$

Geary's $C$ is another test statistic for evaluation of spatial autocorrelation in geo-coded variables. It depends on the (absolute) difference between neighboring values of observed spatial variables. In principle, Geary's $C$ is a variance test similar to the Durbin-Watson test statistic for residuals' autocorrelation in time-series regressions [85]. For a spatially determined variable $y$, Geary's $C$ is calculated as:

$$C = \frac{N-1}{2W} \frac{\sum_i \sum_j w_{ij}(y_i - y_j)^2}{\sum_i (y_i - \bar{y})^2} \,, \tag{2.13}$$

where $N$, $W$, $w_{ij}$, etc. elements follow from previous section. Empirical Geary's $C$ values range from 0 to 2, however Griffith [52] shows that rare occurrences of $C > 2$ are possible. Under the null hypothesis of no spatial autocorrelation, the first two moments of Geary's $C$ are:

$$E(C) = 1 \quad , \quad \mathrm{var}(C) = \frac{(N-1)(2W_1 + W_2) - 4W^2}{2(N+1)W^2} \,, \tag{2.14}$$

where all elements have been introduced in (2.9). Positive spatial dependency leads to $C$ values lower than 1 and negative spatial autocorrelation is reflected in $C$ values

greater than 1. Similarly to Moran's $I$, the $z$-transformation of Geary's $C$ is asymptotically normally distributed. Therefore, $z(C)$ can be used for testing spatial randomness. Significant $z(C) < 0$ values lead to $H_0$ rejection in favor of positive spatial autocorrelation: there is evidence of "more similar" i.e. spatially clustered values of the variable $y$ than they would be by chance. Also, significant $z(C) > 0$ values provide statistical evidence for negative spatial autocorrelation: i.e. a "lack" of similar (high/low) values of $y_i$ observed across neighbors as compared to a random spatial distribution.

### Getis' $G$: spatial clusters and hotspot analysis

Clustering analysis by Getis can only be performed for positively autocorrelated spatial data (where spatial units with high values of a given variable tend to be surrounded by other high observations and vice versa).

*Local G*: $G_i(\tau)$ statistic measures the degree of spatial association – for each $y_i$ from a geo-coded sample, we can calculate a *Local G* statistic as

$$G_i(\tau) = \frac{\sum_{j=1}^{N} s_{ij}(\tau)\, y_j}{\sum_{j=1}^{N} y_j}\,, \qquad j \neq i\,, \tag{2.15}$$

where $s_{ij}(\tau)$ comes from (2.1) and $s_{ij}(\tau) = 1$ if the distance between distinct units $i$ and $j$ is below the (arbitrary) threshold $\tau$ – i.e. if $i$ and $j$ are neighbors – and it is zero otherwise. Observations of variable $y$ are assumed to have a natural origin and positive support [46]. For example, it would be innapropriate to use $G_i(\tau)$ for analysis of residuals from a regression. The numerator of (2.15) is the sum of all $y_j$ observations within distance $\tau$ of unit $i$, but not including $y_i$. The denominator is the sum of all $y_j$ in the sample, not including $y_i$. Hence, $G_i(\tau)$ is a proportion of the aggregated $y_j$ values that lie within $\tau$ of $i$ to the total sum of $y_j$ observations. For example, if we observe high values of $y_j$ within distance $\tau$ of unit $i$, then $G_i(\tau)$ would be relatively high compared to its expected value under the null hypothesis of full spatial randomness:

$$E\left[G_i(\tau)\right] = \frac{S_i}{N-1}\,, \tag{2.16}$$

where $S_i$ is the sum of elements in the $i$-th row of spatial matrix $\boldsymbol{S}$, i.e. the number of neighbors of $i$. Again, $N$ is the total number of spatial observations in the sample. Also, under the $H_0$ of spatial randomness, we can write

$$\mathrm{var}\left[G_i(\tau)\right] = \frac{S_i(N-1-S_i)}{(N-1)^2(N-2)} \left(\frac{Y_{i2}}{Y_{i1}^2}\right)\,, \tag{2.17}$$

where $Y_{i1} = \frac{\sum_j y_j}{N-1}$ and $Y_{i2} = \frac{\sum_j y_j^2}{N-1} - Y_{i1}^2$.

A common modification to the $G_i(\tau)$ statistic consists in dropping the $j \neq i$ restriction from (2.15). Such *Local G* statistic is usually denoted by $G_i^*(\tau)$ and the values of $y_i$ enter both its numerator and denominator expressions. Under spatial randomness, the expected value and variance of $G_i^*(\tau)$ are defined as:

$$E\left[G_i^*(\tau)\right] = \frac{S_i^*}{N}\,, \tag{2.18}$$

$$\text{var}\left[G_i^*(\tau)\right] = \frac{S_i^*(N - S_i^*)}{N^2(N-1)}\left(\frac{Y_{i2}^*}{(Y_{i1}^*)^2}\right)\,, \tag{2.19}$$

where $S_i^* = S_i + 1$, $Y_{i1}^* = \frac{\sum_j y_j}{N}$ and $Y_{i2}^* = \frac{\sum_j y_j^2}{N} - (Y_{i1}^*)^2$ ; the condition $j \neq i$ is dropped.

Usually, $G_i(\tau)$ or $G_i^*(\tau)$ statistics are not reported directly. Instead, a convenient $z$-transformation is used. For example, "*Getis-Ord Local G**": statistic $G_i^*$ is calculated ($i \neq j$ dropped here):

$$G_i^* = \frac{G_i^*(\tau) - E\left[G_i^*(\tau)\right]}{\sqrt{\text{var}\left[G_i^*(\tau)\right]}}\,, \tag{2.20}$$

We can see that $G_i^*$ is a "local" indicator. For an approximately normally distributed $G_i^*(\tau)$, (2.20) readily indicates the type and statistical significance of clustering: As $G_i^*$ statistics (2.20) are calculated for each spatial unit, high positive $G_i^*$ ($z$-score for an $i$-th unit) indicates a hot-spot – a significant concentration of higher-than-average values in the neighborhood of $i$, and vice versa. A $z$-score near zero indicates no such concentration.

To determine statistical significance for a given $N$ and significance level chosen, $G_i^*$ is compared to critical values as provided by Getis and Ord in [46]. Say, for $N = 100$ and $\alpha = 5\%$, the $z$-scores would have to be less than -3.289 for a statistically significant cold spot or greater than 3.289 for a statistically significant hot spot. As an example, we can use the 2014 unemployment data from figure 2.1 to search for hot spots and cold spots of unemployment. At the 5% significance level, we find one unemployment hot spot: an area with a statistically significant concentration of high unemployment values. This hot spot is shown as red-colored units in figure 2.4. Similarly, we identified one unemployment cold spot (low-unemployment cluster). This cold spot is marked blue in figure 2.4.

A general (i.e. not local) statistic of overall spatial concentration $G(\tau)$ can be constructed. $G(\tau)$ evaluates all pairs of values $y_i$ and $y_j$ such that units $i$ and $j$ are within

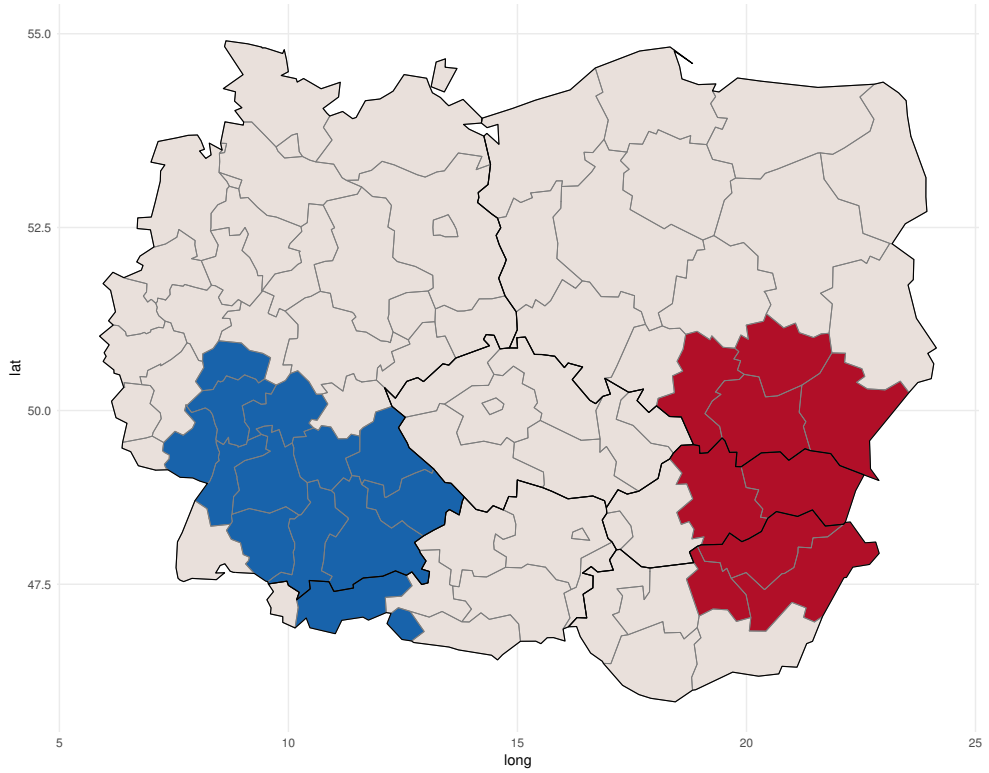


Figure 2.4.: Hot spots and cold spots: Unemployment rate, 2014. Source: Own calculation using GISCO – Eurostat data.

the $\tau$ distance of each other ($i \neq j$ condition is usually applied). $G(\tau)$ interpretation is well comparable to other global statistics, such as Moran's $I$ (see next paragraph for details). $G(\tau)$ is defined as

$$G(\tau) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij}(\tau)\, y_i \cdot y_j}{\sum_{i=1}^{N} \sum_{j=1}^{N} y_i \cdot y_j}, \qquad j \neq i\,. \tag{2.21}$$

Again, the test for statistical significance of overall spatial clustering is based on a $z$-score, where we use the expected mean value

$$E[G_i(\tau)] = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij}(\tau)}{N(N-1)}, \qquad j \neq i\,, \tag{2.22}$$

and variance:

$$\text{var}\left[G(\tau)\right] = E\left[(G_i(\tau))^2\right] - \left[\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij}(\tau)}{N(N-1)}\right]^2, \qquad j \neq i\,. \tag{2.23}$$

Expression (2.22) is the ratio of observed neighbors (actual count of neighbors) to all pairs of spatial units (all potential neighbors) in the dataset, given $\tau$-threshold and assuming that units are not neighbors to themselves. $G_i(\tau)$ in (2.22) comes from expression (2.15) and the derivation of variance formula (2.23) is provided e.g. in [46].

### Comparison of spatial dependency statistics

Despite the fact that all spatial statistics mentioned in this section reflect dependency and non-random patterns in observed spatial data and often provide similar test results, they are not entirely redundant. While Moran's and Geary's statistics concentrate on covariances, Getis' global indicator is based on sums of products. Carrying out different spatial dependency tests – i.e. focusing on different aspects of a (potentially unobservable) spatial dependency pattern can be informative: specific types of spatial settings may lead to disparities in spatial dependency test results. For example, Moran's $I$ does not discriminate between patterns that have high (or low) values concentrated within the $\tau$-defined neighborhood (i.e. among hot-spots and cold-spots under positive spatial autocorrelation), while Getis' $G_i(\tau)$ performs well in this respect. On the other hand – given $G_i(\tau)$ construction (natural origin and positive support of the underlying variable) – it is not suitable for evaluating spatial dependency in variables such as residuals from a regression (Moran's and Geary's statistics can be used for such purpose).

Some general limitations apply to all spatial tests discussed in this chapter:

- None of the statistics is well suited for discerning random observations from spatially dependent data with relatively small deviations from the mean.
- Transformations of the observed spatial variables (e.g. changing measurement units, log-transformation) can result in different values of the statistics.
- If $\tau$-threshold is too low or too high (relatively speaking), the normal approximation and $z$-score based tests may be inappropriate.

Various different spatial dependency statistics and tests have been developed in order to overcome the above general shortcomings – many such tools (usually specialized, i.e. not generally applicable) are available e.g. from [6].

# 3. Spatial econometric models for cross-sectional data

The basic linear regression model (no spatial interactions) for cross sectional data is often denoted as

$$\boldsymbol{y} = \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{3.1}$$

where $\boldsymbol{y}$ is a ($N\times1$) vector of dependent variable observations, $\alpha$ is the intercept and $\boldsymbol{\iota}$ is a ($N\times1$) vector of ones, $\boldsymbol{X}$ is a ($N\times k$) matrix of exogenous regressors, $\boldsymbol{\beta}$ is a ($k\times1$) vector of corresponding parameters and $\boldsymbol{\varepsilon}$ is a $(N\times1)$ vector of error elements. Assumptions and methods for model estimation, statistical inference and interpretation are available e.g. from [84] or [85]. Usually, model (3.1) is estimated using ordinary least squares (OLS).

Within the standard spatial model environment, three different spatial interaction types need to be considered: spatial interaction effects among observations of the endogenous (dependent) variable, interaction effects among regressors and interactions among error terms.

Using a modified notation from [28], we can generalize (3.1) into a fully spatial specification of a linear regression model (cross-sectional data) as follows:

$$\begin{aligned} \boldsymbol{y} &= \lambda\boldsymbol{W}\boldsymbol{y} + \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{u}\,, \\ \boldsymbol{u} &= \rho\boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}\,, \end{aligned} \tag{3.2}$$

where $\boldsymbol{W}\boldsymbol{y}$ is the spatial lag such as (2.6), $\boldsymbol{W}\boldsymbol{X}$ is the spatial lag for regressor matrix $\boldsymbol{X}$ and $\boldsymbol{W}\boldsymbol{u}$ describes spatial interactions (spatial lag) among disturbance elements. Scalars $\lambda$ and $\rho$ as well as the $(k\times1)$ vector $\boldsymbol{\theta}$ are the spatial parameters of the model to be estimated along with $\alpha$ and $\boldsymbol{\beta}$. Since (3.2) includes all the possible spatial interaction types, Elhorst [28] refers to this model as the generalized nesting spatial model (GNS model).

We should also note that there is only one weights matrix $\boldsymbol{W}$ $(N \times N)$ specification in the above GNS model. However, the GNS model may be generalized even further by allowing different $\boldsymbol{W}$ matrices for each of the $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{u}$ lag elements (say, denoted as $\boldsymbol{W^y}, \boldsymbol{W^X}$ and $\boldsymbol{W^u}$). This may be appropriate for applications where significantly diverse spatial interactions occur – see e.g. [62] and [67]. However, in most practical applications, we simply assume a common $\boldsymbol{W}$ for the whole model.

Using various assumptions, the GNS model may be simplified into more specific (nested) types of spatial models. A complete taxonomy is provided e.g. in [28] and reproduced in appendix A.1 for readers' convenience. Here, we only cover in detail three of the most common and empirically useful spatial model specifications.

### Spatial lag model

By assuming that spatial interactions affect only the dependent variable, i.e. by assuming $\boldsymbol{\theta} = \boldsymbol{0}$ and $\rho = 0$, we simplify the GNS model into a spatial lag model (SLM). Here, the endogenous variable is the only element with a significant spatial lag:

$$\boldsymbol{y} = \lambda \boldsymbol{W}\boldsymbol{y} + \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,. \tag{3.3}$$

The SLM specification is used commonly throughout empirical literature, e.g. in models describing taxes imposed by governments (see [28] for other examples). The reduced form of (3.3) is

$$(\boldsymbol{I}_N - \lambda \boldsymbol{W})\boldsymbol{y} = \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{3.4}$$

where $\boldsymbol{I}_N$ is an $(N \times N)$ identity matrix and the RHS regression coefficients explain the variability of individual $y_i$ observations that is not explained spatially. Also, if the inverse to $(\boldsymbol{I}_N - \lambda \boldsymbol{W})$ exists, we can simply transform (3.4) into

$$\boldsymbol{y} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}(\alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\,. \tag{3.5}$$

Equation (3.5) is often referred to as the data generating process (DGP) for $\boldsymbol{y}$ – see [67].

### Spatial Durbin model

If we drop the simplifying assumption $\boldsymbol{\theta} = \boldsymbol{0}$ from the SLM (3.3), thus allowing for spatial interactions in exogenous variables, we get the spatial Durbin model (SDM) specification

$$\boldsymbol{y} = \lambda \boldsymbol{W}\boldsymbol{y} + \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}\,. \tag{3.6}$$

The only difference between SDM and GNS models is the absence of spatial interactions in the error term. For SDM, we assume that the observed variable $y_i$ in a given unit $s_i$ is affected by the endogenous spatial lag $\boldsymbol{Wy}$ (i.e. values of $y$ in neighboring regions have effect on $y_i$), by exogenous regressors for the $i$-th region (the $i$-th row in $\boldsymbol{X}$) as well as by exogenous regressors in neighboring regions (through the $\boldsymbol{WX\theta}$ element). For example, if $y_i$ describes aggregate household income in a region $i$, then such income is influenced by incomes (say, wages) in neighboring regions and by both "domestic" and neighboring rates of unemployment, labor force productivities, etc. For an illustrative list of empirical applications of the SDM, see e.g. [78].

By analogy to the SLM case – and given $(\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}$ exists – we may re-formulate (3.6) in terms of the DGP as follows:

$$\boldsymbol{y} = (\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}(\alpha\boldsymbol{\iota} + \boldsymbol{X\beta} + \boldsymbol{WX\theta} + \boldsymbol{\varepsilon})\,. \tag{3.7}$$

### Spatial error model

The spatial error model (SEM) is another frequently used specification of the spatial model. SEM is obtained from the GNS model by assuming $\lambda = 0$ and $\boldsymbol{\theta} = \boldsymbol{0}$. Hence, spatial interactions take place only among the error terms:

$$\begin{aligned} \boldsymbol{y} &= \alpha\boldsymbol{\iota} + \boldsymbol{X\beta} + \boldsymbol{u}\,, \\ \boldsymbol{u} &= \rho\boldsymbol{Wu} + \boldsymbol{\varepsilon}\,. \end{aligned} \tag{3.8}$$

Theoretical (say, macroeconomic) reasoning of the spatial dependency is not required for SEMs – this approach can be used to model a situation where endogenous variables are influenced by exogenous factors that are omitted from the main equation and spatially autocorrelated. Alternatively, unobserved shocks may follow spatial pattern(s).

## 3.1. Estimation, testing and interpretation of cross sectional spatial models

### Model stability and stationarity conditions for $\lambda$, $\rho$ and $\boldsymbol{W}$

Elhorst [28] and Kelejian and Prucha [60, 61], provide formal assumptions (some of which were already mentioned in section 2.2), generally applicable to the three subtypes of our GNS model specification (3.2). Please note that stability assumptions concerning the spatial weights matrix $\boldsymbol{W}$ are usually based on the spatial (connectivity) matrix $\boldsymbol{S}$

(2.1). Also, for the SDM-related exogenous and spatially lagged element $\boldsymbol{WX\theta}$, we only need to observe stability conditions for $\boldsymbol{W}$ [28].

Spatial model stability conditions may be formalized as follow:

1. Spatial matrix $\boldsymbol{S}$ – such as (2.1), yet other connectivity definitions may be used as well – is a non-negative matrix of known constants with zeros on the diagonal. If $\boldsymbol{S}$ meets this condition, it holds for the row-standardized $\boldsymbol{W}$ matrix as well.

2. Spatial weak dependency holds. This means that correlation between two spatial units converges to zero as the distance between the two units increases. In terms of the $\boldsymbol{S}$ matrix environment, this condition is often formalized into one of the following two conditions: (a) The row and column sums of $\boldsymbol{S}$ should be uniformly bounded in absolute value as $N$ (the number of observed units) goes to infinity. (b) The row and column sums of $\boldsymbol{S}$ should not diverge to infinity at a speed equal to or faster than the growth of sample size $N$.

   Condition (b) is more general (relaxed) and (a) may be interpreted as its special case. Elhorst [28] provides detailed technical discussion and examples related to both alternatives.

3. Matrices $(\boldsymbol{I}_N - \lambda\boldsymbol{W})$ and $(\boldsymbol{I}_N - \rho\boldsymbol{W})$ – used for estimating parameters from (3.5), (3.7) or (3.8) – are non-singular. If the underlying $\boldsymbol{S}$ matrix is symmetric and non-negative, this condition is satisfied whenever $\lambda$ and $\rho$ lie within the $(1/\kappa_{min}, 1)$ interval, where $\kappa_{min}$ denotes the smallest (most negative) real eigenvalue of $\boldsymbol{W}$ and 1 is the largest eigenvalue for a row-standardized $\boldsymbol{W}$.

   Please note that for a symmetric non-negative $\boldsymbol{S}$ matrix, all eigenvalues are real. Even if such $\boldsymbol{S}$ matrix is subsequently row-standardized into the spatial weights matrix $\boldsymbol{W}$, the characteristic roots of this non-symmetric $\boldsymbol{W}$ would remain purely real. Even if the $\boldsymbol{S}$ matrix is not symmetric (say, in the $k$NN case discussed in section 2.2), the same conditions for $\lambda$ and $\rho$ apply: they should stay within the $(1/\kappa_{min}, 1)$ interval, where $\kappa_{min}$ is the most negative purely real eigenvalue of $\boldsymbol{W}$.

   Using a slightly different approach, Kelejian and Prucha [60, 61] argue that $\lambda$ and $\rho$ should lie within the $(-1, 1)$ interval.

### Maximum likelihood estimation of SLMs and SDMs

The RHS regressor element $\boldsymbol{Wy}$ in equation (3.3) is correlated with the error term. Hence, ordinary least squares (OLS) estimation of models with spatially lagged endoge-

nous variables yields biased and inconsistent estimates of regression parameters and standard errors. In contrast, maximum likelihood (ML) estimators for such models are consistent [63]. Therefore, we shall focus on ML estimators here. For other estimation methods (two-stage least squares, generalized method of moments) and related topics, including the Bayesian estimation of spatial models, see e.g. [18], [28] or [67].

We use a slightly modified notation from [67] to describe the single ML estimator used for both SLMs and SDMs specifications, as their likelihood functions coincide (SLM is a special case of SDM, with a restriction $\boldsymbol{\theta} = \mathbf{0}$ imposed). First, we expand the DGP (3.7) by *iid* normality assumption for residuals:

$$\begin{aligned} \boldsymbol{y} &= (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}(\alpha \boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon})\,, \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma_\varepsilon^2\, \boldsymbol{I}_N)\,, \end{aligned} \tag{3.9}$$

where $\sigma_\varepsilon^2$ is the variance of $\boldsymbol{\varepsilon}$. Using substitutions $\boldsymbol{Z} = [\boldsymbol{\iota}\,, \boldsymbol{X}\,, \boldsymbol{W}\boldsymbol{X}]$ and $\boldsymbol{\delta} = [\alpha\,, \boldsymbol{\beta}\,, \boldsymbol{\theta}]^{'}$, we can re-write the SDM equation (3.6) as

$$\boldsymbol{y} = \lambda \boldsymbol{W}\boldsymbol{y} + \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}\,. \tag{3.10}$$

Now, model (3.9) may be written as:

$$\begin{aligned} \boldsymbol{y} &= (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}\boldsymbol{Z}\boldsymbol{\delta} \;+\; (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}\boldsymbol{\varepsilon}\,, \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma_\varepsilon^2\, \boldsymbol{I}_N)\,. \end{aligned} \tag{3.11}$$

The above substitution allows us to use a single likelihood function for both SLM and SDM: for SDMs, we use $\boldsymbol{Z} = [\boldsymbol{\iota}\,, \boldsymbol{X}\,, \boldsymbol{W}\boldsymbol{X}]$. For SLMs, $\boldsymbol{Z} = [\boldsymbol{\iota}\,, \boldsymbol{X}]$ and analogous amendments are made to the vector of parameters $\boldsymbol{\delta}$. Following the approach derived in [3] or [67], the log-likelihood function for the SLM (and SDM) model may be outlined as

$$\begin{aligned} LL(\lambda, \boldsymbol{\delta}, \sigma_\varepsilon^2) &= -\frac{N}{2}\log(\pi\sigma_\varepsilon^2) + \log|\boldsymbol{I}_N - \lambda \boldsymbol{W}| - \frac{\boldsymbol{e}^{'}\boldsymbol{e}}{2\sigma_\varepsilon^2}\,, \\ \boldsymbol{e} &= \boldsymbol{y} - \lambda \boldsymbol{W}\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\delta}\,, \end{aligned} \tag{3.12}$$

where $N$ is the number of spatial units, $|\boldsymbol{I}_N - \lambda \boldsymbol{W}|$ is the determinant of this $N \times N$ matrix, $\boldsymbol{e}$ is a vector of residuals and the row-standardized spatial weights matrix $\boldsymbol{W}$ has real eigenvalues only. For $\lambda$, the above discussed assumption $\lambda \in (1/\kappa_{min}, 1)$ applies. However, in many practical applications, this range is reduced even further by allowing for positive spatial autocorrelation only: $\lambda \in (0, 1)$.

Direct estimation (maximization) of (3.12) is subject to multiple computational issues. Hence, alternative approach is used: technical description of the iterative ML maximization of (3.12) by means of concentrated log-likelihood functions is provided e.g. in [67].

### Maximum likelihood estimation of SEMs

To estimate SEM parameters by the ML method, we use a similar approach as in (3.9): after adding *iid* normality assumption to $\boldsymbol{\varepsilon}$ residuals of the model (3.8), we may write the DGP as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I}_N - \rho\boldsymbol{W})^{-1}\boldsymbol{\varepsilon}\,, \\ \boldsymbol{\varepsilon} &\sim N(\boldsymbol{0}, \sigma_\varepsilon^2\,\boldsymbol{I}_N)\,, \end{aligned} \tag{3.13}$$

where the intercept term has been incorporated into the $\boldsymbol{X}\boldsymbol{\beta}$ expression for simplicity. Now, the full (not concentrated) log-likelihood function for SEMs has the form

$$\begin{aligned} LL(\boldsymbol{\beta}, \rho, \sigma_\varepsilon^2) &= -\frac{N}{2}\log(\pi\sigma_\varepsilon^2) + \log|\boldsymbol{I}_N - \rho\boldsymbol{W}| - \frac{\boldsymbol{e}'\boldsymbol{e}}{2\sigma_\varepsilon^2}\,, \\ \boldsymbol{e} &= (\boldsymbol{I}_N - \rho\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\,. \end{aligned} \tag{3.14}$$

Again, for computational reasons, concentrated log-likelihood functions are calculated iteratively to maximize (3.14) and thus to obtain parameter estimates and corresponding standard errors.

The ML functions (3.12) and (3.14) may be amended to accommodate binomial, count, multinomial and other types of dependent variables – see [11] and [67] for technical discussion.

### Evaluation and comparison of estimated spatial models

Once a spatial model is estimated, we often need to evaluate its overall performance. Usually, models estimated by the ML approach are evaluated using information criteria such as the Akaike information criteria (AIC) or Bayesian information criteria (BIC) as discussed in [85]. Alternatively, the maximized log-likelihood values of (3.12) or (3.14) may be used for testing.

Likelihood ratio ($LR$) test (3.15) can be used to evaluate the relevance of spatial model specification through a set of conveniently chosen restrictions leading to two alternative nested models. For example, we can start with a regression model featuring spatial autocorrelation and compare its performance against a simplified (nested) specification,

where spatial interactions are excluded by zero restrictions on the corresponding coefficients. The $LR$ statistic has the following form:

$$LR = 2(\mathcal{L}_{ur} - \mathcal{L}_r) \underset{H_0}{\sim} \chi^2_q \,, \tag{3.15}$$

where $\mathcal{L}_{ur}$ is the maximized log-likelihood function of the estimated spatial model, (3.12), (3.14), etc. The null hypothesis is used to impute zero restriction on all parameters describing spatial autocorrelation ($\lambda$, $\boldsymbol{\theta}$ or $\rho$ – given specification of the unrestricted model). Therefore, $\mathcal{L}_r$ is the maximized log-likelihood of an estimated restricted (i.e. non-spatial) model such as (3.1). Parameter $q$ describes the degrees of freedom of the $\chi^2_q$ distribution and it equals to the number of parameter restrictions imposed. Under $H_0$ of insignificant spatial effects, the $LR$ statistic approximately follows $\chi^2_q$ distribution and the usual $p$-values may be used for testing $H_0$ against the alternative of significant spatial effects.

In principle, information criteria and $LR$ statistics might be used to compare the general spatial specification of GNS models with different nested specifications (say, SLM or SEM). However, more efficient techniques to choose between SLM and SEM specifications exist: "focused" Lagrange multiplier tests are discussed next.

### Model specification tests: SLM vs SEM

When testing spatial autocorrelation in regression models, Florax and Nijkamp [36] distinguish two basic types of tests: diffuse and focused. Diffuse tests simply reflect whether the residuals are spatially correlated. For example, Moran's $I$ (2.7) is a diffuse test and it can be used for testing spatial randomness in residuals from an estimated regression model such as (3.1):

$$I = \frac{N}{W} \hat{\boldsymbol{u}}' \boldsymbol{W} \hat{\boldsymbol{u}} (\hat{\boldsymbol{u}}' \hat{\boldsymbol{u}})^{-1} \,, \tag{3.16}$$

where $\hat{\boldsymbol{u}}$ is a $N \times 1$ vector of residuals (geo-coded data) from an estimated model. By analogy to Moran's $I$ (2.7), statistical inference is based on the asymptotic normality of (3.16) and the corresponding $z$-score (2.10) as in [3].

The following Lagrange multiplier (LM) tests are focused – have an informative alternative hypothesis where the null hypothesis of a non-spatial model is tested against the alternatives of SEM or SLM specification, respectively [36]. Lagrange multiplier test for SEM specification (3.17) evaluates the null hypothesis of no spatial autocorrelation of

residuals against the alternative of spatial autocorrelation in the residuals (i.e. against the SEM specification):

$$\textit{LM-SEM} = \frac{1}{T}\left(\frac{\hat{\boldsymbol{u}}'\boldsymbol{W}\hat{\boldsymbol{u}}}{\hat{\sigma}^2}\right)^2 \underset{H_0}{\sim} \chi_1^2 \,, \tag{3.17}$$

where $\hat{\sigma}^2$ is the ML-estimated variance of residuals $\hat{\boldsymbol{u}}$ and $T = \mathrm{tr}(\boldsymbol{W}'\boldsymbol{W}+\boldsymbol{W}^2)$ is trace of the matrix. Under the null hypothesis, the *LM-SEM* statistic asymptotically follows $\chi^2$ distribution with one degree of freedom. Please note that the *LM-SEM* statistic (3.17) is just a scaled version of Moran's $I$ (3.16).

Similarly, the *LM-SLM* statistic is used in OLS-estimated linear models to test $H_0$ of spatial independence in $\boldsymbol{y}$ against the alternative of its spatial autocorrelation (SLM specification):

$$\textit{LM-SLM} = \frac{1}{N\,J_{\lambda,\boldsymbol{\beta}}}\left(\frac{\hat{\boldsymbol{u}}'\boldsymbol{W}\boldsymbol{y}}{\hat{\sigma}^2}\right)^2 \underset{H_0}{\sim} \chi_1^2 \,, \tag{3.18}$$

where the term $J_{\lambda,\boldsymbol{\beta}} = \left[(\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{M}(\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}}) + T\hat{\sigma}^2\right]/N\hat{\sigma}^2$ is calculated using the vector of OLS-estimated parameters $\hat{\boldsymbol{\beta}}$ and the "residual maker" (orthogonal projection matrix) $\boldsymbol{M} = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Under $H_0$, *LM-SLM* (3.18) has the same $\chi_1^2$ asymptotic distribution as *LM-SEM* (3.17) – see [49] for technical discussion and derivation of the tests.

The above *LM-SEM* and *LM-SLM* tests are not robust to misspecification: when testing for spatial autocorrelation in the dependent variable, the *LM-SLM* statistic may be severely biased as a result of an autocorrelated error term and vice versa. To address this issue, Anselin [1, 5] introduced LM tests that are robust against local – as in expression (2.11) – misspecifications. The test for a spatial error process that is robust to local presence of a spatial lag is given as:

$$\textit{RLM-SEM} = \frac{1}{T - T^2(NJ_{\lambda,\boldsymbol{\beta}})^{-1}}\left(\frac{\hat{\boldsymbol{u}}'\boldsymbol{W}\hat{\boldsymbol{u}}}{\hat{\sigma}^2} - T(NJ_{\lambda,\boldsymbol{\beta}})^{-1}\frac{\hat{\boldsymbol{u}}'\boldsymbol{W}\boldsymbol{y}}{\hat{\sigma}^2}\right)^2 \underset{H_0}{\sim} \chi_1^2 \,, \tag{3.19}$$

where the subtraction of a correction factor that accounts for the local misspecification (potentially omitted spatial lag process) is clearly visible. In (3.19), we test the $H_0$ of no spatial dependency in residuals (OLS-estimated) against the alternative of SEM specification, while controlling for possible local spatial lag (SLM process).

Similarly, a test for a spatial lag process robust to local presence of spatial error autocorrelation is defined as:

$$RLM\text{-}SLM = \frac{1}{N\,J_{\lambda,\boldsymbol{\beta}} - T}\left(\frac{\hat{\boldsymbol{u}}'\boldsymbol{W}\boldsymbol{y}}{\hat{\sigma}^2} - \frac{\hat{\boldsymbol{u}}'\boldsymbol{W}\hat{\boldsymbol{u}}}{\hat{\sigma}^2}\right)^2 \underset{H_0}{\sim} \chi^2_1\,. \tag{3.20}$$

Under null hypotheses of non-spatial processes, both *RLM-SEM* and *RLM-SLM* asymptotically follow a $\chi^2_1$ distribution. The above tests (3.17) – (3.20) are implemented in R – see [18] for technical details and additional references. Heteroskedasticity-robust versions of statistics (3.17) and (3.18) are available [36]. However, heteroskedasticity-robust versions of the (3.19) and (3.20) tests are not easily accessible as accounting for heteroskedasticity leads to highly non-linear expressions [3].

## Marginal effects in spatial models

In spatial models (GNS), there are two basic types of marginal effects: direct and indirect effects. In presence of spatial autocorrelation among observed variables, if a given explanatory variable in some $i$-th unit changes, than not only the dependent variable in the $i$-th unit is expected to change (direct effect) but also the dependent variables in other units (neighbors of unit $i$) would change. Such effect across spatial units is the indirect effect, sometimes called "spillover". This topic can be conveniently illustrated using a slightly modified GNS model (3.2):

$$\boldsymbol{y} = (\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta}) + \boldsymbol{r}\,, \tag{3.21}$$

where $\boldsymbol{r}$ contains both the intercept and error term of (3.2) specification. Marginal effects for some arbitrary regressor $\boldsymbol{x}_k$ from (3.21) are given by a Jacobian matrix of first derivatives of the expected values of $\boldsymbol{y}$ with respect to the explanatory variable:

$$\frac{\partial E(\boldsymbol{y})}{\partial E(\boldsymbol{x}_k)} = \begin{pmatrix} \dfrac{\partial E(\boldsymbol{y})}{\partial x_{1k}} & \cdots & \dfrac{\partial E(\boldsymbol{y})}{\partial x_{Nk}} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial E(y_1)}{\partial x_{1k}} & \cdots & \dfrac{\partial E(y_1)}{\partial x_{Nk}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial E(y_N)}{\partial x_{1k}} & \cdots & \dfrac{\partial E(y_N)}{\partial x_{Nk}} \end{pmatrix}. \tag{3.22}$$

After some calculation and re-arranging [28], this can be expressed as:

$$\frac{\partial E(\boldsymbol{y})}{\partial E(\boldsymbol{x}_k)} = (\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}(\boldsymbol{I}_N\beta_k + \boldsymbol{W}\theta_k)\,. \tag{3.23}$$

For convenience and clarity, RHS of (3.23) may also be re-written as:

$$\frac{\partial E(\boldsymbol{y})}{\partial E(\boldsymbol{x}_k)} = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} \begin{pmatrix} \beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k \end{pmatrix}. \tag{3.24}$$

Recall that $w_{ij}$ denotes the $(i,j)$-th element of the $\boldsymbol{W}$ matrix and $w_{ij} > 0$ if two spatial units $s_i$ and $s_j$ are neighbors (and zero otherwise). $\beta_k$ and $\theta_k$ are parameters of model (3.21), corresponding to the $k$-th regressor. The RHS of (3.24) is a $N \times N$ matrix. From the RHS of (3.24), we may see several properties of marginal effects in spatial models:

- Each diagonal element of the partial derivatives matrix (3.24) represents a direct effect and every off-diagonal element represents an indirect effect.
- Direct effects and indirect effects differ across spatial units. Each element of the RHS matrix in (3.24) might be different. Individual direct effects differ because the diagonal elements of $(\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}$ are different for each unit (given $\lambda \neq 0$). Indirect effects are different because off-diagonal element of both $\boldsymbol{W}$ and $(\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}$ are different if $\lambda \neq 0$ and/or $\theta_k \neq 0$.
- In absence of spatial autocorrelation of $\boldsymbol{y}$ and $\boldsymbol{x}_k$, i.e. if both $\lambda = 0$ and $\theta_k = 0$, then all off-diagonal elements equal zero. In this case (non-spatial model), indirect effects are not present. Also, direct effects are constant (equal to $\beta_k$) across all spatial units as $(\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}$ simplifies to $\boldsymbol{I}_N$ if $\lambda = 0$.
- The indirect effects that occur if $\theta_k \neq 0$ and $\lambda = 0$ are referred to as **local effects**. The name arises from the fact that such effect only arise from the neighborhood of a given unit. For example, from (3.24) we can see that the effect of $x_{jk}$ ($k$-th regressor for the $j$-th unit) on $y_i$ is nonzero only if units $s_i$ and $s_j$ are neighbors (i.e. $w_{ij} > 0$). For non-neighboring units, $x_{jk}$ has no effect on $y_i$.
- The indirect effects that occur if $\lambda \neq 0$ and $\theta_k = 0$ are referred to as **global effects**. The name comes from the fact that effects on $y_i$ originate from units that lie within the neighborhood of $s_i$ as well as from units outside this neighborhood. Mathematically, this is due to the fact that matrix $(\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1}$ does not contain zero elements (given $\lambda \neq 0$) – even though $\boldsymbol{W}$ does contain (usually many) zero elements.

- For $\lambda \neq 0$ and $\theta_k \neq 0$, both local and global indirect effect are present and they cannot be separated from each other.
- The presence or absence of spatial autocorrelation $\rho$ in the error term of equation (3.21) has no impact on the marginal effects (3.24): as we take the first derivative of $E(\boldsymbol{y})$ with respect to $\boldsymbol{x}_k$, the $\boldsymbol{r} = (\alpha\boldsymbol{\iota} + \rho\boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon})$ element disappears because it is a "constant".

An alternative approach to describing (the same underlying) direct and indirect effects in cross-sectional models is formalized by equations (4.10) – (4.13) where it serves for derivation of marginal effects in spatial panel data models as in [67].

Considering the complexity of marginal effects (3.24) for one regressor, the problem of presenting estimation output from a spatial econometric model with multiple regressors may be severe: even if reliable estimates $\hat{\lambda}, \hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ are available, we have to deal with a $N \times N$ matrix of marginal (direct and indirect) effects for each regressor.

Therefore, estimated marginal effects are usually presented in an aggregated form. For each regressor, we usually report two (sometimes three) statistics: First, a summary indicator for direct effects is calculated as the average of all diagonal elements in (3.24). Second, indirect effects are reported as the average of all off-diagonal elements. The above statistics are usually reported along with their corresponding standard errors and statistical significance indicators ($p$-values) – see [28] or [67] for technical discussion.

In addition, total effects are sometimes reported. Total effect (total impact) is just a sum of the direct and indirect impacts. Total standard errors, $z$ scores and statistical significance levels are also calculated by aggregating the underlying direct impacts and spillovers [28]. The main reason for reporting total impacts can be summarized as follows: in many empirical applications, the direct and indirect effects may come with opposite signs. Therefore, at some higher level of spatial aggregation, direct impacts and spillovers could cancel out. For example, positive direct effects may come at the "price" of equally prominent negative spillovers. Therefore, total impacts are often reported along with their direct/indirect constituents – even if there are no contradicting signs of direct/indirect impacts.

There is a particular drawback to the above discussed marginal effects for SLM: the ratio between direct and indirect effect for a regressor $\boldsymbol{x}_k$ is independent of $\beta_k$. This is because the $\beta_k$ coefficients cancel out in the numerator and denominator of such ratio (direct/indirect effects). This ratio depends only on the parameter $\lambda$ and on the $\boldsymbol{W}$

matrix specification. Hence, it is the same for all regressors in a given spatial model. Unfortunately, this "behavior" (say, identical relative strengths of direct and indirect effects for all regressors) seems rather implausible in many types of empirical applications [28].

## 3.2. Robustness of spatial models with respect to neighborhood definition

Another major weakness of the spatial models described above is the fact that $\boldsymbol{W}$ matrices cannot be estimated along with model parameters. Rather, $\boldsymbol{W}$ needs to be specified prior to model estimation. There is little theoretical background for choosing the "right" $\boldsymbol{W}$ matrix specification. The variety of available neighborhood definitions and standardization methods implies that researchers usually evaluate several alternative spatial structure settings in order to verify model stability and robustness of the results.

On the other hand, not all researches consider the ambiguity in $\boldsymbol{W}$ specification as a problem. LeSage and Pace [68] argue that SDMs and other spatial specifications allow for accurate estimation of the spatial effects, even if both the spatial matrix $\boldsymbol{W}$ and the spatial regression model are misspecified. They argue that for a given model – estimated using two similar (highly correlated) weights matrices $\boldsymbol{W}^{(a)}$ and $\boldsymbol{W}^{(b)}$ – it would be unlikely to reach materially different coefficient estimates and partial derivatives as in (3.24). Their argument is supported by an empirical (micro-level) housing-prices example based on data from [55] (506 spatial units) and for three alternative $\boldsymbol{W}$ matrices generated using the $k$NN approach for $k = 5$, 6 and 7. Unfortunately, the conclusions presented in [68] would only hold for a relatively narrow class of spatial models and $\boldsymbol{W}$ settings. The presumed robustness does not easily extend from a $k$NN-based spatial structure to other types of structures, such as distance-based $\boldsymbol{W}$ matrices.

A theoretically simple yet computationally expensive approach to evaluating robustness of estimated models (regression parameters, direct and indirect effects and their confidence intervals) against changes in the pre-specified $\boldsymbol{W}$ matrices may be summarized as follows:

1. Start with a relatively sparse $\boldsymbol{W}$ distance matrix, that is generated by using a restrictive (i.e. low) maximum distance threshold for neighbor definition. Note that the threshold must be high enough to ensure at least one neighbor for each spatial unit in the sample. The existence of islands (units with zero neighbors) breaks down the ML estimation of spatial models. Using such sparse $\boldsymbol{W}$ matrix,

estimate your model and record all relevant estimates along with their confidence intervals and other model-related statistics (AIC, BIC, etc.)

2. Increase the maximum distance threshold by some relatively low amount (say, 10-km iterations can be used for modeling behavior in NUTS2 units). Estimate the model and record all relevant information. Note that unlike in the $k$NN case, threshold changes may lead to significantly uneven changes in neighbor sets and thus in the corresponding $\boldsymbol{W}$ matrices.

3. Repeat step 2 until a maximum neighborhood threshold distance (defined with respect to the spatial domain – "map" – used) is reached. Usually, this would happen in one of the following manners: (a) "range" (as in the semivariogram figure 1.2) is reached – there is no point of increasing the maximum neighbor distance threshold beyond a distance where data are no longer spatially autocorrelated. (b) Maximum neighbor distance threshold becomes so large that the assumption of spatial weak dependency (discussed in section 1.2) no longer applies. As a rule-of-thumb indicator, we often see that the SpatialLag($y_i$) as in (2.6) becomes nearly constant across spatial units and var($\boldsymbol{w}_i\boldsymbol{y}$) falls quickly beyond some ad-hoc (dataset-specific) distance threshold. (c) We have some theoretically/empirically based prior information limiting the plausible range of spatial interactions (i.e. maximum distance threshold).

4. Plot the estimated spatial parameters, direct and indirect effects of interest, etc. against the distance thresholds used. From such plots, stability of estimates and corresponding significance intervals can be studied. Also, the information criteria (or maximized log-likelihoods) obtained for models estimated using different $\boldsymbol{W}$ matrices may suggest ("identify") the best distance threshold (most supported by the data) to be used for subsequent model interpretation. See figures 6.2, 7.2 and 8.3 for empirical examples of this approach.

## 3.3. Spatial filtering and semi-parametric models

Parametric framework and models as discussed in section 3.1 are appropriate in multiple empirically relevant scenarios and applications. However, parametric methods are potentially not robust in a situation where the model suffers from a simultaneous presence of different sources of misspecification. Factors such as unaccounted nonlinear relationship among spatially correlated variables, spatially varying relationships (non-stationarity), uncontrolled common factors (spatial and time-related) and other instances of spatial

heterogeneity can disrupt spatial (cross-sectional) dependencies or even manifest themselves as such.

In such circumstances, spatial filtering methods may be used to remove global and/or local spatial dependencies among geo-coded variables. Unlike ML estimators, spatial filtering does not rely on distributional assumptions and it is fairly robust to model misspecification. Nonparametric filtering can be used to eliminate spatial autocorrelation from observed $y_i$ values by "spatial demeaning" through local autocorrelation measures.

In case we need to preserve some level of spatial properties within the model, spatial filtering can be implemented as a semi-parametric method [82]. For this approach, spatial information can be extracted from the underlying spatial structure through the Moran eigenvector approach [17].

### Univariate nonparametric spatial filtering by Getis

Getis' nonparametric filtering method can only be applied to non-negative and positively autocorrelated spatial observations. It is based on the *Local G* statistic $G_i(\tau)$ from (2.15). Pioneered by Getis [45], the ratio of $G_i(\tau)$ and its expected value $E\left[G_i(\tau)\right]$ from (2.16) can be used for multiplicative transformation of a spatial variable $y_i$ as follows:

$$\ddot{y}_i = \frac{E\left[G_i(\tau)\right]}{G_i(\tau)} \cdot y_i \,, \tag{3.25}$$

where $\ddot{y}_i$ is the spatially filtered value of $y_i$. The transformation outlined in (3.25) corrects for positive spatial autocorrelation in observed data by counterbalancing the clustering of below-average and above-average observations. Specifically, the filtering factor in (3.25) shrinks $y_i$ if the majority of observations $y_j$ within the $\tau$ distance of unit $i$ are above average. Similarly, $y_i$ is inflated if neighboring observations feature below-average values.

While this approach is computationally simple and intuitive, its underlying positive support assumption for $y_i$ can be a strong limitation. Also, the process of setting $\tau$ (maximum neighbor distance threshold) is rather arbitrary. However, the critical distance for statistically significant spatial interactions may be based on observed data – by fitting empirical semivariograms (1.11). Formula (3.25) is univariate. Therefore, if we aim to estimate regression models such as (3.1) using spatially filtered data:

$$\ddot{\boldsymbol{y}} = \alpha\boldsymbol{\iota} + \ddot{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \,, \tag{3.26}$$

then spatial filtering (3.25) has to be applied individually to each observed variable in the model – hence, positive support assumptions apply to all variables used in the regression. For additional discussion and an empirical application of Getis' univariate filtering to the analysis of EU's regional unemployment dynamics, see Chapter 6 or [38].

### Moran's eigenvector maps

Moran's eigenvector maps (MEM) belong to a wider class of spatial-filtering methods that seek to avoid the inconveniences involved in estimation and interpretation of spatial autoregressive parameters of SLM (3.3) and SDM (3.6) models [51, 82].

When spatial-filtering is used in the context of spatial econometric modeling, we work with two distinct types of regressors that are used in the regression model: we have a set of geo-coded variables (macroeconomic indicators) with a common underlying spatial structure and a spatial-filter element (e.g. a MEM), describing spatial dependency patterns. The eigenvector-based filtering as described in [51] can be summarized as follows: We start by determining pairwise geographic (Euclidean) distances $h_{ij}$ among all spatial units into a symmetric $N \times N$ matrix of distances. Next, $\boldsymbol{W}$ is constructed along the following rules for its individual $w_{ij}$ elements:

$$\boldsymbol{W} = [w_{ij}] = \begin{cases} 0 & \text{if} \quad i = j\,, \\ 0 & \text{if} \quad h_{ij} > \tau\,, \\ [1 - (h_{ij}/4\tau)^2] & \text{if} \quad h_{ij} \leq \tau\,, \end{cases} \tag{3.27}$$

where $\tau$ is a conveniently chosen threshold value that keeps spatial units connected – e.g. through a minimum spanning tree algorithm [51] on a graph corresponding to the $\boldsymbol{H}$ matrix of distances. This approach leads to a symmetric $\boldsymbol{W}$ matrix with uniformly bounded row and column sums (see section 2.2 or [28] for details). In literature ([25], [51], etc.), the $\boldsymbol{W}$ matrix from (3.27) is also called "truncated connectivity matrix" because not all spatial units are connected. For the same reason (not all units being connected), $\boldsymbol{W}$ connectivity matrices (3.27) have a non-Euclidean nature – unlike the distance matrices $[h_{ij}]$. The description provided above uses spatial distances to define neighbors, yet contiguity and $k$NN-based approaches are applicable as well.

Now, we use $\boldsymbol{W}$ from (3.27) to construct a double-centered connectivity matrix $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega} = (\boldsymbol{I}_N - \frac{1}{N}\,\boldsymbol{\iota}_N\boldsymbol{\iota}_N')\,\boldsymbol{W}\,(\boldsymbol{I}_N - \frac{1}{N}\,\boldsymbol{\iota}_N\boldsymbol{\iota}_N')\,, \tag{3.28}$$

where $\boldsymbol{I}_N$ is the identity matrix ($N \times N$) and $\boldsymbol{\iota}_N$ is a column vector of ones, with length $N$ (the number of spatial units considered). Row and column sums of $\boldsymbol{\Omega}$ equal zero by construction. See [51] for technical discussion of centered connectivity matrices and their applications in ecology, etc.

Next, we obtain "Moran's eigenvectors" $\boldsymbol{v}$ and eigenvalues $\kappa$ for $\boldsymbol{\Omega}$ from equation

$$\boldsymbol{\Omega}\,\boldsymbol{v} = \kappa\boldsymbol{v}\,. \tag{3.29}$$

By solving the characteristic equation system $(\boldsymbol{\Omega} - \kappa\boldsymbol{I}_N)\,\boldsymbol{v} = \boldsymbol{0}$ for $\kappa$, we obtain all solutions $\kappa_i$, satisfying the condition $|\boldsymbol{\Omega} - \kappa\boldsymbol{I}_N| = 0$ . Because the centered connectivity matrix $\boldsymbol{\Omega}$ is real and symmetric, its eigenvectors (a subset of unique eigenvectors of the rank-deficient matrix $\boldsymbol{\Omega}$) are orthogonal and linearly independent. Given the non-Euclidean nature of the underlying connectivity matrix $\boldsymbol{W}$ in (3.27), both positive and negative eigenvalues are produced by solving (3.29). Eigenvectors corresponding to positive $\kappa$ represent positive spatial association and negative eigenvalues represent negative spatial dependency processes.

For a well-defined spatial domain that provides adequate coverage of a given geographic area[1] and given positive spatial autocorrelation in data, eigenvectors $\boldsymbol{v}$ bear the following interpretation: MEMs (spaces spanned by single or multiple $\boldsymbol{v}$ eigenvectors) with associated large (positive) eigenvalues $\kappa_i$ represent global-scale spatial trends (say, landscape-wide core/periphery dynamics in observed EU data). Eigenvectors with medium eigenvalues represent medium scale dynamics (e.g. "regional", say NUTS1 and NUTS2 interaction patterns) and eigenvectors with small (positive) eigenvalues would represent small scale dependencies ("local" patchiness, e.g. at the NUTS3 or LAU levels).

MEM, which is a conveniently chosen subset of eigenvectors $\boldsymbol{v}$, can be used as a synthetic explanatory variable in semiparametric regression models. First, the selected eigenvectors are combined to constitute a spatial autocorrelation function. This semi-parametric part of the model (the spatial function furnishing the latent spatial autocorrelation in geo-coded variables) is then additively combined with an appropriate set of explanatory

[1] This necessarily non-rigorous statement reflects a general assumption that irregularly spaced patterns (e.g. NUTS2 regions and their centroids) can appropriately reflect geographical variability in the observed data. Regularly spaced (chessboard-like) patterns are not always required and/or appropriate. While both approaches have their empirical advantages and disadvantages, irregular design (adaptive sampling) is not fundamentally "worse".

variables in the regression model. Following the general MEM selection methodology described in [17] and [82], a relatively simple semiparametric spatial model based on MEMs is derived next.

### MEM-based semiparametric spatial regression models

To accommodate the paradigm used by Tiefelsdorf and Griffith [82], we re-cast the SDM (3.6) specification

$$\boldsymbol{y} = \lambda \boldsymbol{W}\,\boldsymbol{y} + \alpha\boldsymbol{\iota} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

using a common spatial autocorrelation coefficient $\delta = \lambda$ and assuming $\boldsymbol{\theta} = -\delta\boldsymbol{\beta}$:

$$\boldsymbol{y} = \delta \boldsymbol{W}\,\boldsymbol{y} + (\boldsymbol{I}_N - \delta\boldsymbol{W})\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{3.30}$$

where the intercept term $\alpha\boldsymbol{\iota}$ is absorbed into $\boldsymbol{X}$ matrix, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ and the common factor constraint assumption $\boldsymbol{\theta} = -\delta\boldsymbol{\beta}$ for the term $(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta})$ follows e.g. from [3]. In model (3.30), the spatial structure (spatial weights matrix) $\boldsymbol{W}$ is used generically: we can embrace symmetric matrices such as (3.27) or their row-standardized (non-symmetric) transformations, as well as other approaches to $\boldsymbol{W}$ specification.

The semiparametric model is established using a misspecification paradigm, where we assume an elementary regression model with spatially autocorrelated disturbances

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*\,, \\ \boldsymbol{\varepsilon}^* &= \boldsymbol{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\,, \end{aligned} \tag{3.31}$$

where $\boldsymbol{\varepsilon}^*$ are the spatially autocorrelated disturbances that may be decomposed into $\boldsymbol{\varepsilon}$ (white noise) and $\boldsymbol{E}$: a set of missing (unobservable, unspecified) exogenous variables that follow a common spatial dependency pattern given by $\boldsymbol{W}$ and $\boldsymbol{\gamma}$ is a vector of parameters.

It is important to note that the misspecification approach to spatial modeling – concentrated in the $\boldsymbol{E}\boldsymbol{\gamma}$ term – is not directly comparable with the preceding specification (3.30) nor with the seemingly conformable SEM (3.8) specification that is based on spatial dependency among random elements. Using the semiparametric approach described next, the misspecification term $\boldsymbol{E}\boldsymbol{\gamma}$ is approximated by a set of spatial proxy variables – conveniently chosen MEMs.

We start by rewriting the spatial model (3.30) as

$$\boldsymbol{y} - \delta \boldsymbol{W}\,\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} - \delta \boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,. \tag{3.32}$$

Solving the LHS of (3.32) for $\boldsymbol{y}$, we get

$$\boldsymbol{y} = (\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1}[\boldsymbol{X}\boldsymbol{\beta} - \delta \boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}]\,. \tag{3.33}$$

Using the equivalent expansion $(\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} = \sum_{k=0}^{\infty} \delta^k \boldsymbol{W}^k$, we can cast (3.33) as

$$\begin{aligned} \boldsymbol{y} &= \sum_{k=0}^{\infty} \delta^k \boldsymbol{W}^k [\boldsymbol{X}\boldsymbol{\beta} - \delta \boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}]\,, \\ &= \sum_{k=0}^{\infty} \delta^k \boldsymbol{W}^k (\boldsymbol{X}\boldsymbol{\beta}) - \sum_{k=0}^{\infty} \delta^{k+1} \boldsymbol{W}^{k+1} (\boldsymbol{X}\boldsymbol{\beta}) + \sum_{k=0}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon}\,, \end{aligned} \tag{3.34}$$

Because $\delta^0 \boldsymbol{W}^0 = \boldsymbol{I}_N$ and $\sum_{k=0}^{\infty} \delta^{k+1} \boldsymbol{W}^{k+1} = \sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k$, we can simplify (3.34) into

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \left[\sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon}\right] + \boldsymbol{\varepsilon}\,, \tag{3.35}$$

which incorporates white noise $\boldsymbol{\varepsilon}$ and the spatial misspecification term, thus following the structure of $\boldsymbol{\varepsilon}^*$ in model (3.31). Importantly, $\sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon}$ is not correlated to the regressors $\boldsymbol{X}$ (by OLS assumption of $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$ independence). Therefore, the OLS-estimated parameters $\hat{\boldsymbol{\beta}}$ of a modified model (3.31)

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*\,, \\ \boldsymbol{\varepsilon}^* &= \sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon} + \boldsymbol{\varepsilon} \end{aligned} \tag{3.36}$$

are unbiased estimators for the population parameters $\boldsymbol{\beta}$. On the other hand, the estimated standard errors $s.e.(\hat{\boldsymbol{\beta}})$ will be biased [82].

In empirical applications, the misspecification "problem" as in equation (3.31) is difficult to deal with because the exogenous term $\boldsymbol{E}\boldsymbol{\gamma}$ is unknown/unspecified/missing. However, taking advantage of the specification (3.36), we can design spatial proxy variables that satisfy model assumptions: We start by extracting eigenvectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ from the

quadratic form in

$$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\} \equiv \text{evec}\left[\boldsymbol{M}_{\boldsymbol{X}} \frac{1}{2}(\boldsymbol{W} + \boldsymbol{W}')\, \boldsymbol{M}_{\boldsymbol{X}}\right], \tag{3.37}$$

where $\boldsymbol{M}_{\boldsymbol{X}} = [\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$ is a projection matrix and the extracted eigenvectors are orthogonal to $\boldsymbol{X}$. The RHS part of expression (3.37) is designed to induce symmetry to generic (non-symmetric) neighborhood specifications. For symmetric $\boldsymbol{W}$ matrices, it holds that $\boldsymbol{W} = \frac{1}{2}(\boldsymbol{W} + \boldsymbol{W}')$ and thus (3.37) may be simplified accordingly. Please note that the RHS term in square brackets is a generalization of the centered connectivity matrix introduced in (3.28).

Given the properties of the quadratic form in (3.37), eigenvectors are mutually orthogonal and form a basis for a spatial proxy variable in a semiparametric spatial model. Now, if we follow [82] and associate the term $\boldsymbol{E}$ from (3.31) with a convenient parsimonious subset of eigenvectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$, we can approximate the misspecification term in (3.30) and its (3.35) form as

$$\boldsymbol{E}\boldsymbol{\gamma} \approx \sum_{k=1}^{\infty} \delta^k \boldsymbol{W}^k \boldsymbol{\varepsilon}\,, \tag{3.38}$$

where orthogonality to $\boldsymbol{X}$ is maintained so that OLS-based estimates $\hat{\boldsymbol{\beta}}$ remain unbiased. Once the approximate substitution (3.38) is performed in model (3.35), the resulting specification may be estimated using the formula

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\,, \tag{3.39}$$

where $\boldsymbol{y}$ is decomposed into a systematic component (featuring $\boldsymbol{X}$), stochastic spatial component and white-noise residuals. For conveniently specified $\boldsymbol{E}$, the estimated stochastic spatial term $\boldsymbol{E}\hat{\boldsymbol{\gamma}}$ removes a significant portion of the mean squared error (MSE) term attributable to spatial autocorrelation (i.e. $\boldsymbol{E}\hat{\boldsymbol{\gamma}}$ is often referred to as spatial filter). Overall, Tiefelsdorf and Griffith [82] conclude that this filtering approach is fairly robust to model specification errors when compared with fully parametric models and ML-based estimators. Also, they provide semiparametric approach to estimation of SLM-like models, where spatial autocorrelation affects the dependent variable only.

Choosing a convenient and parsimonious subset of eigenvectors for the $\boldsymbol{E}$ term is crucial for a successful application of the MEM-based semiparametric algorithm. First of all, we aim to choose such $\boldsymbol{E}$ so that the residuals $\hat{\boldsymbol{\varepsilon}}$ from model (3.39) become spatially random (independent with respect to the underlying spatial domain). Also, we aim to

find a parsimonious, (i.e. smallest possible) subset of eigenvectors leading to spatial independence of $\hat{\boldsymbol{\varepsilon}}$.

To establish the subset of eigenvectors $\boldsymbol{E}$ with desired properties, stepwise regression approach is often suggested [50], [82]. Although different search (stepwise) algorithms are available, most of them are based on a modified Moran's $I$ (2.7) coefficient, often denoted [25] as $MC$ and formalized by

$$MC_{\boldsymbol{v}_i} = \frac{N}{\boldsymbol{\iota}_N' \boldsymbol{Z}\, \boldsymbol{\iota}_N} \boldsymbol{v}_i' \boldsymbol{W}\, \boldsymbol{v}_i \,, \tag{3.40}$$

where $\boldsymbol{Z}$ is a distance-based similarity matrix with $z_{ij} = 1 - (\frac{h_{ij}}{\max(h_{ij})})^2$. Individual $z_{ij}$ values vary between zero for $h_{ij} = \max(h_{ij})$ and 1 for $h_{ij} = 0$. Given the real-valued and symmetric nature of spatial matrices, Moran's $MC$ for each eigenvector $\boldsymbol{v}_i$ is equal to its associated eigenvalue $\kappa_i$ if $\boldsymbol{W}$ is scaled to satisfy $[\boldsymbol{\iota}_N'(\boldsymbol{W} + \boldsymbol{W}_N')\boldsymbol{\iota}_N]/2 = N$ (see [82] for discussion on the Rayleigh quotient and for computational implications).

A forward stepwise selection method by Griffith's [50] may be described as follows: The first eigenvector $\boldsymbol{v}_1$ is chosen based on maximizing $MC_{\boldsymbol{v}_i}$ in expression (3.40). Using such $\boldsymbol{v}_1$ as a starting eigenvector subset for $\boldsymbol{E}$, equation (3.39) is estimated and corresponding residuals $\hat{\boldsymbol{\varepsilon}}$ are evaluated with respect to their spatial autocorrelation (e.g. using Moran's $I$). If residuals are spatially dependent, new eigenvector, $\boldsymbol{v}_2$ is added to $\boldsymbol{E}$ using the same $MC$-maximization criterion and spatial autocorrelation in residuals is tested again. Eigenvectors are iteratively added to $\boldsymbol{E}$, until the spatial autocorrelation in residuals $\hat{\boldsymbol{\varepsilon}}$ falls below a predetermined threshold (say, until we fail to reject the $H_0$ of no spatial autocorrelation at the 5 % significance level).

Please note that all eigenvectors (3.37) in $\boldsymbol{E}$ are mutually orthogonal by design, which has important implications to our semiparametric model: Estimated $\hat{\gamma}_i$ parameters already included in model (3.39) would remain unaffected after adding new eigenvector(s) to $\boldsymbol{E}$. Eigenvectors in $\boldsymbol{E}$ follow a strictly decreasing sequence, where each eigenvector explains a specific proportion of variance in residuals of the model (3.39) – the largest proportion of variance is explained by the first eigenvector selected into $\boldsymbol{E}$, the second largest amount of variance is explained by the second eigenvector, etc. These two implications combined together also lead to identical $\boldsymbol{E}$ obtained through forward and backward stepwise selections.

Computational efficiency of the above stepwise selection procedure may be significantly increased using some of the following approaches. For example, we can restrict the set of

eigenvectors entering our stepwise evaluation based on the sign of spatial autocorrelation in observed data (usually positive for geo-coded economic variables). This approach reduces the number of $MC_{\boldsymbol{v}_i}$ calculated and it is primarily relevant for large spatial domains, i.e. for large $N$ values). Tiefelsdorf and Griffith [82] provide a very effective search algorithm that is also implemented in `R` (`spdep` package, [17]).

The above discussed search methods for eigenvector components of $\boldsymbol{E}$ focus on minimizing spatial autocorrelation in $\hat{\boldsymbol{\varepsilon}}$. However, this is not the only possible paradigm and other approaches are discussed in literature as well as used empirically (see [51] or [66]). For example, we can disregard spatial autocorrelation in residuals of model (3.39) and use stepwise (or exhaustive / brute force) approach for selection of individual eigenvectors in $\boldsymbol{E}$ to minimize the total variance in residuals of the model (3.39).

# 4. Spatio-temporal data and econometric models

## 4.1. Static spatial panel models

As the geo-coded observations of variables are often repeated in time, spatial panel models can be used to depict interactions among variables across spatial units as well as over time [10, 71]. This section provides a brief spatial panel model taxonomy with corresponding estimation and testing methods. The general form of a static panel model that includes both the spatial effects (spatial lag) for the dependent variable and the spatially autocorrelated error terms may be outlined as

$$\begin{aligned} \boldsymbol{y} &= \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}\,, \\ \boldsymbol{u} &= \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N\right) \boldsymbol{\mu} + \boldsymbol{\varepsilon}\,, \\ \boldsymbol{\varepsilon} &= \rho \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \boldsymbol{\varepsilon} + \boldsymbol{\upsilon}\,, \end{aligned} \tag{4.1}$$

where $\boldsymbol{y}$ is a $NT \times 1$ column vector of dependent variable observations ($i = 1, 2, \ldots, N$ denotes cross-sectional units and $t = 1, 2, \ldots, T$ relates to the time dimension and $i$ is the "fast" index here). The spatial weights matrix $\boldsymbol{W}$ follows from section 2.2 and $\boldsymbol{X}$ is a $(NT \times k)$ full column rank matrix of $k$ exogenous regressors. Elements $\boldsymbol{I}_T$ and $\boldsymbol{I}_N$ are identity matrices (with dimensions given by their subscripts) and $\boldsymbol{\iota}_T$ is a $(T \times 1)$ vector of ones. For panel data models, regression equations and corresponding expressions often involve the $\otimes$ Kronecker product operator. Elements of vector $\boldsymbol{\beta}$ as well as $\lambda$ and $\rho$ are parameters of the model. The disturbance vector $\boldsymbol{u}$ $(NT \times 1)$ is a sum of two terms: the unobserved individual effects $\boldsymbol{\mu}$ and spatially autocorrelated innovations $\boldsymbol{\varepsilon}$. The $(N \times 1)$ vector $\boldsymbol{\mu}$ holds time-invariant and spatially uncorrelated individual effects. Innovations $\boldsymbol{\varepsilon}$ are spatially autocorrelated with a spatial error autoregressive parameter $\rho$ where $|\rho| < 1$. Finally, $\boldsymbol{\upsilon}' = (\boldsymbol{\upsilon}_1', \ldots, \boldsymbol{\upsilon}_T')$ is a vector of spatially independent innovations: $\upsilon_{it} \sim IID\left(0, \sigma_\upsilon^2\right)$.

Following the standard approach to panel data analysis (as in [10], [49] and [84]), individual effects $\mu_i$ are treated as either fixed or random. With the "random effects" (RE) model, we assume that unobserved individual effects $\boldsymbol{\mu}$ are not correlated to other regressors of the model. For "fixed effects" (FE) models, we relax this assumption: some level of correlation between individual effects and other regressors is acceptable. In the following sections, basic familiarity with non-spatial RE and FE models [49] is assumed.

### Extending tests from spatial cross-sections to panel data

Ou et al. [76] discuss the application and power of Moran's $I$ test in panel models. For spatial panel models, Moran's $I$ may be cast as

$$I = \frac{\boldsymbol{e}'\boldsymbol{\mathcal{W}}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{e}}\,, \tag{4.2}$$

where the OLS-based residual vector $\boldsymbol{e}$ is given as $\boldsymbol{e} = [\boldsymbol{I}_{NT} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y} = \boldsymbol{M}\boldsymbol{y}$. $\boldsymbol{I}_{NT}$ is an $(NT \times NT)$ identity matrix and $\boldsymbol{M}$ is a real symmetric matrix, the "residual maker". $\boldsymbol{\mathcal{W}} = (\boldsymbol{I}_T \otimes \boldsymbol{W})$ is a block-diagonal matrix $(NT \times NT)$ with $T$ blocks of spatial weights matrices $\boldsymbol{W}$ (row-standardized). As derived in [81], Moran's $I$ from equation (4.2) asymptotically (fixed $T$, $N \to \infty$) follows normal distribution and its mean and variance are

$$E(I) = \frac{\mathrm{tr}(\boldsymbol{M}\boldsymbol{\mathcal{W}})}{NT - k}\ , \tag{4.3}$$

and

$$\mathrm{var}(I) = \frac{\mathrm{tr}(\boldsymbol{M}\boldsymbol{\mathcal{W}}\boldsymbol{M}\boldsymbol{\mathcal{W}}') + \mathrm{tr}\left((\boldsymbol{M}\boldsymbol{\mathcal{W}})^2\right) + (\mathrm{tr}(\boldsymbol{M}\boldsymbol{\mathcal{W}}))^2}{(NT - k)(NT - k + 2)} - [E(I)]^2\ , \tag{4.4}$$

where $\mathrm{tr}(\cdot)$ is trace of a matrix. Finally, expressions (4.2), (4.3) and (4.4) are combined into a $z$-score statistic that can be used (asymptotically) for testing the null hypothesis of spatial randomness in $\boldsymbol{e}$, as follows:

$$z_I = \frac{I - E(I)}{\sqrt{\mathrm{var}(I)}} \underset{H_0}{\sim} N(0,1)\,. \tag{4.5}$$

Please note that $\boldsymbol{e}$ can be either residuals from regressing $\boldsymbol{y}$ on $\boldsymbol{X}$ or – alternatively – $z_I$ can be used in a univariate mode: for $\boldsymbol{X} = \boldsymbol{\iota}$, we test for spatial dependency in deviations of $\boldsymbol{y}$ from its mean value.

Interestingly, the above described Moran's $I$ test for panel data may be applied even if the spatial weights matrix is time-varying – i.e. if spatial interactions are allowed to change over time (please do not confuse this concept with "dynamic" models containing lagged

exogenous and/or endogenous regressors as in section 5.1). For this generalization, $\boldsymbol{\mathcal{W}}$ is simply re-cast as

$$\boldsymbol{\mathcal{W}} = \{w_{ij,t}\} = \begin{bmatrix} \boldsymbol{W}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{W}_2 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{W}_{T-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{W}_T \end{bmatrix},$$

where $\boldsymbol{W}_t$ denotes spatial weights matrix ($N \times N$) for period $t$. However, Ou et al. [76] point out that with increased time-variability of spatial interactions, test results become very sensitive to $\boldsymbol{W}_t$ misspecification.

Specialized Lagrange multiplier tests for spatial dependency identification can be used to simplify the all-encompassing general spatial panel model specification (4.1) and to identify proper model specification. Anselin et al. [7] provide extensions to the (3.17) – (3.20) cross-sectional LM-tests used to choose between SLM and SEM-type dynamics. Also, similar LM-based specification tests for identification of spatial dependency types are available in [71].

When testing for spatial randomness of choosing among different spatial dependency patterns, we need to keep in mind the inherent ambiguities emanating from $\boldsymbol{W}$ construction, where contiguity, $k$NN (with different $k$) or maximum distance thresholds (with varying $\tau$) may be applied. Robustness of the test (model specification choice) should always be assessed with respect to varying spatial patterns ($\boldsymbol{W}$ matrix).

## 4.2. Random effects (RE) model

This section deals with random effect specification of individual effects $\boldsymbol{\mu}$: when individual effects can be viewed as random and independent of regressors, the random effect (RE) model and RE estimator can provide relative efficiency to estimation [71, 85]. RE model can be used even under a generalizing assumption of permanent (i.e. time-invariant) spatial spillovers in individual effects – for detailed discussion of various RE spatial panel model specifications, please refer to [64] and to literature sources referenced therein.

Random effect properties of $\boldsymbol{\mu}$ are implicitly included in the following assumption concerning unobservable individual effects: $\mu_i \sim IID\left(0, \sigma_\mu^2\right)$. Also, the spatially autocorre-

lated error term $\boldsymbol{\varepsilon}$ from (4.1) can be equivalently expressed (see Appendix A.3) as:

$$\boldsymbol{\varepsilon} = \left(\boldsymbol{I}_T \otimes \boldsymbol{B}_N^{-1}\right) \boldsymbol{v} \;\text{ where }\; \boldsymbol{B}_N = \left(\boldsymbol{I}_N - \rho \boldsymbol{W}\right),$$

where $\boldsymbol{B}_N$ is assumed non-singular. Now, under RE assumptions and substituting for $\boldsymbol{\varepsilon}$, the general spatial panel specification (4.1) can be re-written as a RE spatial model:

$$\begin{aligned} \boldsymbol{y} &= \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}\,, \\ \boldsymbol{u} &= \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N\right) \boldsymbol{\mu} + \left(\boldsymbol{I}_T \otimes \boldsymbol{B}_N^{-1}\right) \boldsymbol{v} \end{aligned} \tag{4.6}$$

and its error variance (see [71]) may be outlined as

$$\operatorname{var}\left(\boldsymbol{u}\right) = \boldsymbol{\Omega}_{\boldsymbol{u}} = \sigma_{\boldsymbol{\mu}}^2 \left(\boldsymbol{\iota}_T \boldsymbol{\iota}_T' \otimes \boldsymbol{I}_N\right) + \sigma_{\boldsymbol{v}}^2 \left[\boldsymbol{I}_T \otimes \left(\boldsymbol{B}_N' \boldsymbol{B}_N\right)^{-1}\right]. \tag{4.7}$$

ML and GMM-based estimation method for RE spatial lag models (featuring spatial lag in the dependent variable $\boldsymbol{y}$ but not in the error term $\boldsymbol{\varepsilon}$) and RE spatial error models (where only $\boldsymbol{\varepsilon}$ is spatially autocorrelated) are available e.g. in [28, 64] and from numerous other sources. In this section, estimation of a full specification (both $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ spatially lagged) is briefly outlined, based on the approach in [71].

We start by scaling down the error-term variance matrix $\boldsymbol{\Omega}_u$ by the idiosyncratic error variance into

$$\boldsymbol{\Sigma} = \phi \left(\boldsymbol{\iota}_T \boldsymbol{\iota}_T' \otimes \boldsymbol{I}_N\right) + \boldsymbol{I}_T \otimes \left(\boldsymbol{B}_N' \boldsymbol{B}_N\right)^{-1},$$

where $\phi = \sigma_\mu^2 / \sigma_\varepsilon^2$. From [3, 71], the log-likelihood function for our RE model (4.6) is given as follows:

$$\begin{aligned} LL(\boldsymbol{\beta}, \sigma_v^2, \phi, \lambda, \rho) = &-\tfrac{NT}{2} 2\pi \;-\tfrac{NT}{2} \log \sigma_v^2 \;+ T \log \left|\boldsymbol{A}_N\right| \\ &-\tfrac{1}{2} \log \left|T\phi \boldsymbol{I}_N + \left(\boldsymbol{B}_N' \boldsymbol{B}_N\right)^{-1}\right| \\ &+ (T-1) \log \left|\boldsymbol{B}_N\right| \;- \frac{1}{2\sigma_v^2} \boldsymbol{u}' \boldsymbol{\Sigma}^{-1} \boldsymbol{u}\,, \end{aligned} \tag{4.8}$$

where $|\boldsymbol{A}_N|$ is the determinant of $\boldsymbol{A}_N = (\boldsymbol{I}_N - \lambda \boldsymbol{W})$ and $|\boldsymbol{B}_N| = \det(\boldsymbol{B_N})$. An iterative estimation procedure for parameters of (4.6) may be outlined as follows: We start with some initial values for $\lambda$, $\rho$ and $\phi$ and estimate $\boldsymbol{\beta}$ and $\sigma_v^2$ using the first order conditions

$$\begin{aligned} \boldsymbol{\beta} &= (\boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{A}_N \boldsymbol{y}\,, \\ \sigma_v^2 &= (\boldsymbol{A}_N \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' \, \boldsymbol{\Sigma}^{-1} (\boldsymbol{A}_N \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) / NT\,. \end{aligned} \tag{4.9}$$

Output from the GLS-based estimation in (4.9) is used for concentrating the log-likelihood function (4.8) and its maximization with respect to $\lambda$, $\rho$ and $\phi$. The estimated $\lambda$, $\rho$ and $\phi$ parameters are then used to update matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{A}_N$ for the GLS estimation of (4.9). This estimation – alternating between GLS estimation of (4.9) and ML estimation of the concentrated (4.8) – is performed repeatedly, until convergence criteria are satisfied. `R`-implementation of this estimation process along with alternative estimation algorithms are available e.g. from [71].

### Interpretation of the estimated RE model – marginal effects

Besides spatial panel RE model estimation and testing, specific interpretation issues need to be addressed for models featuring spatial lag in the dependent variable. For the same reasons as in cross-sectional models, the estimated $\boldsymbol{\beta}$ parameters of model (4.1) do not form a proper basis for description of model dynamics. For such purpose, we use the partial derivative approach to interpretation of the impacts from changes to the regressors. Direct and indirect effects need to be considered: as we simulate a change in $x_{it,r}$ – the $r$-th explanatory variable for spatial unit $i$ at time $t$ – we expect the dependent variable in the $i$-th unit to change (direct effect) and also, for $\lambda \neq 0$, we expect some non-zero effects on the dependent variables in neighboring units (indirect effects). Please note that spatial autocorrelation in the error term plays no role here – see discussion below expression (3.24).

Even with static spatial panel models, this type of dynamics is relatively complex to describe – although in principle it is similar to the cross-sectional case described in section 3.1. We can use a slightly modified notation from LeSage and Pace [67] to provide a simple overview, starting with the reduced form of a cross-sectional spatial model (we follow the explicit inclusion of intercept as in [67]):

$$(\boldsymbol{I}_N - \lambda \boldsymbol{W})\, \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \alpha \boldsymbol{\iota}_N + \boldsymbol{u}\,, \tag{4.10}$$

where $\boldsymbol{y}$ and $\boldsymbol{u}$ are $(N \times 1)$, $\boldsymbol{X}$ is $(N \times k)$ and $\alpha$ is the intercept. Equation (4.10) can be conveniently rewritten for subsequent interpretation as

$$\boldsymbol{y} = \sum_{r=1}^{k} \boldsymbol{S}_r(\boldsymbol{W})\, \boldsymbol{x}_r + \boldsymbol{A}_N^{-1}\, \boldsymbol{\iota}_N \alpha + \boldsymbol{A}_N^{-1}\, \boldsymbol{u}\,, \tag{4.11}$$

where

$$\begin{aligned} \boldsymbol{S}_r(\boldsymbol{W}) &= \boldsymbol{A}_N^{-1}\boldsymbol{I}_N\beta_r = (\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}\,\boldsymbol{I}_N\beta_r\,, \\ \boldsymbol{A}_N^{-1} &= (\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1} = \boldsymbol{I}_N + \lambda\boldsymbol{W} + \lambda^2\boldsymbol{W}^2 + \lambda^3\boldsymbol{W}^3 + \cdots\,. \end{aligned} \tag{4.12}$$

The direct effects (direct impacts) and spillover effects (indirect effects) for a cross-sectional spatial model (4.11) are given by

$$\begin{aligned} \frac{\partial y_i}{\partial x_{ir}} &= \boldsymbol{S}_r(\boldsymbol{W})_{ii} \qquad \text{(direct effect: } s_i \to s_i\text{)} \\ \frac{\partial y_i}{\partial x_{jr}} &= \boldsymbol{S}_r(\boldsymbol{W})_{ij} \qquad \text{(spillover: } s_j \to s_i\text{)} \end{aligned} \tag{4.13}$$

where $\boldsymbol{S}_r(\boldsymbol{W})_{ij}$ is a scalar term, element of the matrix $\boldsymbol{S}_r(\boldsymbol{W})$. Now, we may conclude our derivation of impacts for the spatial panel model (4.1): for spatial panel models, $\boldsymbol{S}_r(\boldsymbol{W})$ may be generalized to

$$\boldsymbol{S}_r(\boldsymbol{\mathcal{W}}) = (\boldsymbol{I}_{NT} - \lambda\boldsymbol{\mathcal{W}})^{-1}\,\boldsymbol{I}_{NT}\beta_r\,, \tag{4.14}$$

where $\boldsymbol{I}_{NT}$ and $\boldsymbol{\mathcal{W}}$ follow from equation (4.2), i.e. $\boldsymbol{\mathcal{W}} = (\boldsymbol{I}_T \otimes \boldsymbol{W})$. To express the impacts from an estimated spatial panel model, we only need to substitute $\boldsymbol{S}_r(\boldsymbol{\mathcal{W}})$ for $\boldsymbol{S}_r(\boldsymbol{W})$ in expression (4.13). For additional discussion of impacts' variance and statistical significance tests, see [67].

### Mundlak-Chamberlain approach to RE models

As already stated, we assume that the (time-invariant and spatially independent) unobserved individual effects $\boldsymbol{\mu}$ are not correlated with other regressors $\boldsymbol{X}$ of the spatial RE model. Although this is a very strong assumption – arguably unrealistic in most empirical applications – one can simply adopt the Mundlak-Chamberlain method of dealing with correlated random effects (CRE) – see [84] for basic (non-spatial) reference.

The Mundlak-Chamberlain CRE method allows us to relax the assumption of independence among $\boldsymbol{\mu}$ and $\boldsymbol{X}$ while keeping the "random" nature of $\boldsymbol{\mu}$: since individual effects $\boldsymbol{\mu}$ are time invariant, we can reasonably assume that they would be correlated with the time-invariant part of $\boldsymbol{X}$ – such as their individual means:

$$\bar{x}_{i,k} = \frac{1}{T}\sum_{t=1}^{T} x_{it,k}\,.$$

Here, for each spatial unit $i$, we calculate average values of a given regressor $k$ across all

corresponding time observations (this is done for all time-varying regressors in $\boldsymbol{X}$).

Based on the above CRE assumptions, the individual-specific effect $\boldsymbol{\mu}$ can be split up into two parts: one that is related to the time-averages of the explanatory variables and a second one that is unrelated to the explanatory variables. This may be done through the following equation:

$$\boldsymbol{\mu} = (\boldsymbol{I}_N \otimes \boldsymbol{\iota}_T)' \, \bar{\boldsymbol{X}} \boldsymbol{\pi}_\mu + \boldsymbol{r}_\mu \,, \tag{4.15}$$

where $\boldsymbol{\mu}$ comes from the RE model (4.6), $\bar{\boldsymbol{X}}$ contains the individual-specific averages of all regressors from equation (4.6) and its first column is $\boldsymbol{\iota}_{NT}$ as it relates to the intercept of the "auxiliary regression" (4.15). $\boldsymbol{\pi}_\mu$ is a vector of coefficients relating the individual means of $\boldsymbol{X}$ to $\boldsymbol{\mu}$. Finally, $\boldsymbol{r}_\mu$ are the residuals – i.e. the "remaining part" of $\boldsymbol{\mu}$ that is uncorrelated to explanatory variables. Basically, the Mundlak-Chamberlain CRE method consist in substituting the RHS of (4.15) for $\boldsymbol{\mu}$ into (4.6) and estimating the resulting expanded specification using the RE approach: we can safely use the RE estimator as the random effects $\boldsymbol{r}_\mu$ are uncorrelated to other $\boldsymbol{X}$-matrix regressors – once the averages of time-varying regressors are incorporated into the model. In [72], Miranda et al. derive a very general Mundlak-Chamberlain CRE model that even allows for spatial dependency in the individual effects $\boldsymbol{\mu}$ (identification, estimation and interpretation of such model is discussed therein).

## 4.3. Fixed effects (FE) model

Compared to RE models, FE assumptions are less restrictive, allowing for correlation between the (unobserved) individual effects and other regressors. At the same time, we cannot use time-invariant regressors in spatial FE models (also applies to non-spatial FE models, see [85]) as those would be eliminated from the model during the FE-based estimation process where individual averages of variables are subtracted from all regressors (and from the dependent variable as well).

Following the approach of [29] and [71] as well as drawing from model taxonomy introduced in chapter 3, this chapter focuses on two types distinct models: FE spatial lag (spatial dependency in $\boldsymbol{y}$ only) and FE spatial error (only the error-term is spatially dependent) models. Discussion relating to empirically less common spatial FE model specifications that combine both types of dependencies is available from [64].

### Fixed effects spatial lag model

For the FE spatial lag models that do not feature spatial autocorrelation in the error term, we can simplify and stack the general spatial panel model (4.1) into

$$\boldsymbol{y} = \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \boldsymbol{y} + \left(\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N\right) \boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\upsilon}\,, \tag{4.16}$$

which is the FE spatial lag model. Although all elements of (4.16) have been introduced already, we shall repeat that $\boldsymbol{\upsilon}$ is a vector of spatially independent and normally distributed innovations that vary both over cross-sectional units and across time.

As already mentioned, the first step in FE spatial lag model estimation consists of eliminating the individual effects (along with any other time-invariant regressors) from the model. The elimination is performed by means of time-demeaning (subtracting corresponding individual means from observed values). The transformed equation (4.16) can be written as follows:

$$\ddot{\boldsymbol{y}} = \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \ddot{\boldsymbol{y}} + \ddot{\boldsymbol{X}}\boldsymbol{\beta} + \ddot{\boldsymbol{\upsilon}}\,, \tag{4.17}$$

where $\ddot{\boldsymbol{y}} = \left[\left(\boldsymbol{I}_T - \frac{1}{T}[\boldsymbol{\iota}_T\boldsymbol{\iota}_T']\right) \otimes \boldsymbol{I}_N\right] \boldsymbol{y}$, i.e. it features $\boldsymbol{y}$ observations after time-demeaning (subtracting individual means). Please note that proper ordering of the observed variables $y_{it}$ in $\boldsymbol{y}$ is necessary, with $t$ being the "fast" index. Alternatively, we can define $\ddot{\boldsymbol{y}}$ observations as $\ddot{y}_{it} = y_{it} - \bar{y}_i$. Because individual effects are time-invariant, $\ddot{\mu}_i = \mu_i - \bar{\mu}_i = 0$ for all spatial units and $\boldsymbol{\mu}$ disappears from (4.17). The calculation of $\ddot{\boldsymbol{X}}$ and definition of $\ddot{\boldsymbol{\upsilon}}$ are analogous to $\ddot{\boldsymbol{y}}$ [71].

The log-likelihood function corresponding to (4.16) is

$$LL(\boldsymbol{\beta}, \lambda, \sigma_\upsilon^2) = -\frac{NT}{2} \log\left(2\pi\sigma_\upsilon^2\right) + T \log\left|\boldsymbol{A}_N\right| - \frac{NT}{2\sigma_\upsilon^2}\, \boldsymbol{e}'\boldsymbol{e}\,, \tag{4.18}$$

where $\boldsymbol{e} = \boldsymbol{y} - \lambda \left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{A}_N = \left(\boldsymbol{I}_N - \lambda\boldsymbol{W}\right)$ was already defined for (4.8). Following the approach in [71], expression (4.18) is maximized as follows: We start with two auxiliary regressions performed using the transformed variables from (4.17): $\ddot{\boldsymbol{y}}$ and $\left(\boldsymbol{I}_T \otimes \boldsymbol{W}\right) \ddot{\boldsymbol{y}}$ are separately regressed on $\ddot{\boldsymbol{X}}$ and we store the corresponding residuals as $\ddot{\boldsymbol{e}}_0$ and $\ddot{\boldsymbol{e}}_1$. Next, we can formulate a concentrated log-likelihood function

$$LL = c + T \log\left|\boldsymbol{A}_N\right| - \frac{NT}{2} \log\left[(\ddot{\boldsymbol{e}}_0 - \lambda\ddot{\boldsymbol{e}}_1)'(\ddot{\boldsymbol{e}}_0 - \lambda\ddot{\boldsymbol{e}}_1)\right]\,, \tag{4.19}$$

where $c$ is a constant (does not depend on $\lambda$). Numerical optimization is used to obtain

$\lambda$ that maximizes the expression (4.19). $\boldsymbol{\beta}$ and $\sigma_v^2$ are then calculated from the first order conditions for (4.18) as $\lambda$ is replaced by its (4.19) estimate. The asymptotic variance-covariance matrix for the estimated parameters $(\boldsymbol{\beta}, \lambda, \sigma_v^2)$ is provided in [71].

Using the estimated FE model (4.16) parameters, we may also obtain estimates of individual effects as follows:

$$\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} \left[ y_{it} - \hat{\lambda} \left( \sum_{j=1}^{N} w_{ij} y_{jt} \right) - \boldsymbol{x}_{it} \hat{\boldsymbol{\beta}} \right] . \tag{4.20}$$

However, for short panels ($N \gg T$), enough observations for reliable $\mu_i$ estimation often do not accumulate.

### Fixed effects spatial error model

Spatial lag of the dependent variable is absent in the FE spatial error model. The FE spatial error model may be outlined as a simplified version of the (4.1) specification:

$$\begin{aligned} \boldsymbol{y} &= (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N)\, \boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \\ \boldsymbol{\varepsilon} &= \rho\, (\boldsymbol{I}_T \otimes \boldsymbol{W})\, \boldsymbol{\varepsilon} + \boldsymbol{\upsilon}\,. \end{aligned} \tag{4.21}$$

The estimation of FE spatial error model (4.21) extends easily from the cross-sectional SEM case described by (3.13) and (3.14). The log-likelihood function corresponding to (4.21) can be written as follows:

$$LL(\boldsymbol{\beta}, \rho, \sigma_v^2) = -\frac{NT}{2} \log\left(2\pi\sigma_v^2\right) + T \log|\boldsymbol{B}_N| - \frac{1}{2\sigma_v^2}\, \boldsymbol{e}' \left[ \boldsymbol{I}_T \otimes \left( \boldsymbol{B}_N' \boldsymbol{B}_N \right) \right] \boldsymbol{e}\,, \tag{4.22}$$

where $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{B}_N = (\boldsymbol{I}_N - \rho\boldsymbol{W})$ follows from (4.6). Again, concentrated log-likelihood functions serve for iterative parameter estimation of the (4.21) model. Detailed estimation algorithm is provided e.g. in [71], along with corresponding asymptotic variance-covariance matrix of parameters. After estimation, individual effects can be retrieved from the model as

$$\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} \left( y_{it} - \boldsymbol{x}_{it} \hat{\boldsymbol{\beta}} \right), \tag{4.23}$$

for which identical reliability issues as in (4.20) would apply.

The above list of RE and FE spatial models and their estimation methods provides just a

brief overview. Technical details of the estimation routines are not included as those may differ slightly among diverse software packages. For detailed technical descriptions, see e.g. Elhorst [28] or [29], who also provides a non-matrix-notation overview of the above specifications of RE and FE spatial models, along with alternative estimation methods (if applicable), empirical illustrations and useful references to supplementary literature. Additionally, different GMM estimation methods and statistical inference procedures relevant for diverse types of RE and FE spatial panel specifications are covered e.g. in [71].

### RE vs. FE: Hausman test

The testing of random effects assumptions with respect to an estimated spatial panel model (RE vs. FE tests) is an essential part of RE model-estimation and verification. The original Hausman test was introduced in [56] for non-spatial panel data models. By comparing the RE and FE estimators for a given model specification, the Hausman statistic is used to test whether RE assumptions are supported by the data. Fortunately, Hausman test extends relatively easily to spatial data and models. Mutl and Pfaffermayr [74] provide a generalization of the Hausman test for spatial panel data. The test statistic can be written as

$$H = NT\left(\hat{\boldsymbol{\theta}}_{RE} - \hat{\boldsymbol{\theta}}_{FE}\right)'\left(\hat{\boldsymbol{\Sigma}}_{RE} - \hat{\boldsymbol{\Sigma}}_{FE}\right)^{-1}\left(\hat{\boldsymbol{\theta}}_{RE} - \hat{\boldsymbol{\theta}}_{FE}\right) \underset{H_0}{\sim} \chi^2_k \,, \tag{4.24}$$

where $\hat{\boldsymbol{\theta}}_{RE}$ and $\hat{\boldsymbol{\theta}}_{FE}$ are the RE and FE-based estimates of spatial panel model parameters, while $\hat{\boldsymbol{\Sigma}}_{RE}$ and $\hat{\boldsymbol{\Sigma}}_{FE}$ are their corresponding variance-covariance matrices. Under $H_0$ (RE assumptions hold), the test statistic is asymptotically $\chi^2$-distributed with $k$ degrees of freedom where $k$ is the number of regressors. For technical details and additional tests relevant for RE and FE spatial models (linear restrictions, etc.), see [71].

# 5. Advanced spatial panel models

This chapter provides a brief description of two types of advanced spatial panel data models: dynamic and hierarchical. As spatial models are subject to ongoing theoretical and methodological research, new types of models and approaches are constantly emerging. An overview of relatively advanced models, along with corresponding assumptions and estimation methods, can be drawn from a survey article by Lee and Yu [64].

## 5.1. Spatial panel data: dynamic models

Spatial dynamic panel data (SDPD) models encompass both spatial and time (dynamic) effects and allow researchers to quantify and evaluate spatial and time dependencies along with main effects (say, macroeconomically determined). Elhorst [28] provides a set of general features (DGP aspects) that SDPD models should account for:

- Serial dependency (time-autocorrelation) in observations for each spatial unit $s_i$.
- Spatial interactions (spillovers) among neighboring units at each time period.
- Unobservable effects (both individual and time-effects).
- Possible endogeneity of regressors (not limited to spatial and/or temporal lags of the dependent variable).

A very general SDPD model specification – a counterpart to the GNS model (3.2) for cross sectional data – may be cast as follows:

$$\begin{aligned}\boldsymbol{y}_t &= \tau_1 \boldsymbol{y}_{t-1} + \lambda_0 \boldsymbol{W}\boldsymbol{y}_t + \lambda_1 \boldsymbol{W}\boldsymbol{y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta}_0 + \boldsymbol{X}_{t-1}\boldsymbol{\beta}_1 \\ &\quad + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\theta}_0 + \boldsymbol{W}\boldsymbol{X}_{t-1}\boldsymbol{\theta}_1 + \boldsymbol{Z}_t\boldsymbol{\pi} + \boldsymbol{\nu}_t\,, \qquad (5.1a)\\ \boldsymbol{\nu}_t &= \rho_1 \boldsymbol{\nu}_{t-1} + \rho_s \boldsymbol{W}\boldsymbol{\nu}_t + \boldsymbol{\mu} + \gamma_t \boldsymbol{\iota}_N + \boldsymbol{\varepsilon}_t\,, \qquad (5.1b)\\ \boldsymbol{\mu} &= \phi \boldsymbol{W}\boldsymbol{\mu} + \boldsymbol{\xi}\,, \qquad (5.1c)\end{aligned}$$

where $\boldsymbol{y}_t$ is a $(N\times 1)$ vector of observations for all spatial units $s_i$ at a given time period $t$. $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$ are $(N\times k)$ and $(N\times \ell)$ matrices of exogenous and endogenous regressors

respectively. $\boldsymbol{W}$ is a spatial weights matrix that follows from previous chapters, e.g. from equation (3.2). $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are $(k \times 1)$ response parameter vectors corresponding to exogenous regressors and $\boldsymbol{\pi}$ is an $(\ell \times 1)$ vector of parameters for the endogenous regressors. $\boldsymbol{\nu}_t$ is the error term of model (5.1) and it is assumed to be correlated both in space and serially (in time).

Following the approach in [28] or [64], there is no $\boldsymbol{W}\boldsymbol{\nu}_{t-1}$ included in (5.1b) – we do not assume a spatial lag effect from temporally lagged error term $\boldsymbol{\nu}_{t-1}$ on $\boldsymbol{\nu}_t$. $\boldsymbol{\mu}$ contains individual effects $\mu_i$ – those are time-invariant and serve for controlling unobservable spatial effects (their exclusion from the model would lead to biased estimation of the response parameters).

Similarly, $\gamma_t \boldsymbol{\iota}_N$ is a $(N \times 1)$ vector of time effects specific for each time period $t$, used to control for unobservable time effects (unit-invariant). From (5.1c), we can see the spatially autocorrelated nature of $\mu_i$ elements. $\tau_1$, $\lambda_0$, $\lambda_1$, $\rho_1$, $\rho_s$ and $\phi$ are model parameters describing spatial, temporal and time-lagged spatial dependencies. Finally, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\xi}$ (which is time-invariant, hence no $t$ subscript) are i.i.d. disturbance terms with zero means and $\sigma^2_\varepsilon$, $\sigma^2_\xi$ variances.

Stability conditions – necessary for ML estimation of model (5.1) – may be achieved by imposing restrictions on model parameters and spatial weights matrix $\boldsymbol{W}$. In equation (5.1c), restrictions relevant for $\boldsymbol{W}$ and the corresponding parameter $\phi$ follow from stability conditions discussed in section 3.1: mainly, the row sums of $\boldsymbol{W}$ are uniformly bounded as $N$ goes to infinity and $(\boldsymbol{I}_N - \phi\boldsymbol{W})$ is non-singular. In equation (5.1b), stability holds if the characteristic roots of matrix $\rho_1(\boldsymbol{I}_N - \rho_s\boldsymbol{W})^{-1}$ lie within the unit circle. Similarly, for equation (5.1a), the characteristic roots of matrix $(\tau_1\boldsymbol{I}_N + \lambda_1\boldsymbol{W})(\boldsymbol{I}_N - \lambda_0\boldsymbol{W})^{-1}$ should also lie within the unit circle. Also, matrices $(\boldsymbol{I}_N - \lambda_0\boldsymbol{W})$ and $(\boldsymbol{I}_N - \rho_s\boldsymbol{W})$ have to be non-singular for model estimation. Stability conditions discussed here are derived and described in detail by Elhorst ([26] and [27]) who also shows the inherent trade-off between magnitudes of spatial and temporal autocorrelation coefficients in model (5.1).

## SDPD estimation

If model (5.1) can be reasonably simplified to dynamic non-spatial specification by setting $\lambda_0 = \lambda_1 = 0$ and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 = \boldsymbol{0}$, then model estimation can be performed by applying the Arellano-Bond GMM-based estimator – see e.g. [8] or [49] for detailed description. Similarly, if we are able to reasonably ignore all temporal autocorrelations by setting $\tau = \lambda_1 = 0$ and $\boldsymbol{\beta}_1 = \boldsymbol{\theta}_1 = \boldsymbol{0}$, then the resulting static spatial panel models can be estimated using methods discussed in sections 4.1 and 4.2.

In SDPD model specifications where time and space lags are present simultaneously, parameter estimation becomes rather complicated. Different small-sample and asymptotic properties apply, depending on panel dimensions and effect assumptions used (short vs. long panels and FE or RE). For example, [64] shows that even as time and individual dimensions go to infinity (at the same rate), there is an asymptotic bias present in the autoregressive parameter(s) for both ML and instrumental variable (IVR) estimation . To address this problem, Bun and Carree [19] provide a bias-corrected (analytically modified FE) estimator, while other possible approaches to bias correction in "large $N$ and $T$" SDPD models involve e.g. a Jacknife-based procedure discussed in [53].

Besides the above-mentioned issues in asymptotic behavior of SDPD model estimators for different panel dimensions and estimator types, the situation may be complicated even more by violation of different stability assumptions corresponding to individual equations of model (5.1). Different spatial arrangements and corresponding estimation approaches (including spatial cointegration and ECMs) are discussed in [64]. Finally, the case of temporal unit root in the dependent variable of model (5.1a) (i.e. $\tau_1 = 1$) is addressed e.g. in [86], along with model estimation methods and asymptotic behavior of the estimators.

## 5.2. Hierarchical spatial panel data model

Fingleton et al. [34] outline a "three-dimensional" (hierarchical) spatial panel model with complex spatial interaction effects. Besides a standard spatial lag in the dependent variable, their model features nested random errors that follow a spatial moving average process. Using a unit-specific notation, their model can be outlined as follows:

$$\begin{aligned} y_{ijt} &= \lambda \left( \sum_{g=1}^{N} \sum_{h=1}^{M_g} w_{ij,gh}\, y_{ght} \right) + \boldsymbol{x}_{ijt}\boldsymbol{\beta} + \varepsilon_{ijt}\,, \\ \varepsilon_{ijt} &= u_{ijt} - \rho \sum_{g=1}^{N} \sum_{h=1}^{M_g} m_{ij,gh}\, u_{ght}\,, \\ u_{ijt} &= \alpha_i + \mu_{ij} + \upsilon_{ijt}\,, \end{aligned} \tag{5.2}$$

where $y_{ijt}$ is the dependent variable and the three subscripts are designed to describe the hierarchical and panel data structure: $i = 1, 2, \ldots, N$ describes groups, $j = 1, 2, \ldots, M_i$ denotes the number of individuals in an $i$-th group ($M_i$ values may differ between groups) and $t$ is the temporal subscript. By analogy to equation (5.1), $\boldsymbol{x}_{ijt}$ is a $(1\times k)$ row vector

of regressors, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of regression coefficients and $\varepsilon_{ijt}$ is the disturbance element. $w_{ij,gh}$ are the elements of the $ij$-th row ($j$-th element within group $i$) of a row-standardized weights matrix $\boldsymbol{W}$ such as (2.2) and the subscripts $gh$ relate to $h$-th individual within group $g$. The symbol $\lambda$ denotes the usual spatial lag parameter (stability conditions apply).

Disturbances $\varepsilon_{ijt}$ are contemporaneously correlated through a moving average process at the individual $(ij)$ level. $m_{ij,gh}$ are elements of a spatial weights matrix $\boldsymbol{M}$ (in this model, $\boldsymbol{W}$ and $\boldsymbol{M}$ can represent different neighborhood patterns) and $u_{ijt}$ is assumed *i.i.d.* distributed with zero mean and constant variance $\sigma_u^2$. In model (3.26), spatial heterogeneity is described through $u_{ijt}$: we have a group-specific time-invariant element $\alpha_i$, a nested (time-invariant) unit-specific element $\mu_{ij}$ and a white-noise element $\upsilon_{ijt}$. All three elements ($\alpha_i, \mu_{ij}$ and $\upsilon_{ijt}$) are assumed *i.i.d.* distributed with zero means and variances $\sigma_\alpha^2, \sigma_\mu^2$ and $\sigma_\upsilon^2$ respectively.

## Estimation of the hierarchical spatial model

Parameters of the model (5.2) can be estimated by the ML approach described in [3]. Fingleton et al. [34] point out the problematic (restrictive) nature of distributional assumptions that need to be explicitly formulated for ML estimation, along with potentially severe complications from possible regressor endogeneity and computational complexity issues originating from ML estimation of complex models such as (5.2). Instead of ML estimation, they propose a three-stage IVR-based estimation method as follows: In the first stage, spatial lag panel model is estimated using IVR. In the second stage, a GMM approach is used to estimate the spatial moving averages parameter $\rho$ and the variance $\sigma_u^2$ (using residuals from the first stage). In the third stage, a transformation similar to Cochrane–Orcutt approach [49] is combined with an IVR estimation to disentangle the structure of random elements $u$. The derivation and detailed discussion of the three-stage estimator and its properties are provided in [34].

# Empirical applications

# 6. Analysis of regional unemployment dynamics using Getis' filtering approach

This chapter is mostly based on the application part of [38] by Formánek and Hušek.

## 6.1. Spatial analysis of unemployment dynamics

As quantitative macroeconomic analyses are often performed at the regional level, statistical effects of spatial dependency in observed data should be addressed. Spatial autocorrelation violates basic assumptions of independence among observations of variables used for estimation of regression models [85]. As discussed by [28, 39, 67] and by numerous other authors, uncontrolled spatial autocorrelation leads to biased model estimates and/or biased significance statistics. A common econometric practice is to directly (parametrically) estimate spatial dependencies – either by maximum likelihood methods as in [28] or through a Bayesian approach described in [67]. Unfortunately, ML-based parametric approaches rely heavily on pre-specified distributional assumptions [82]. An alternative approach to the parametric spatial-dependency modeling is used here: a distribution-free & non-parametric approach pioneered by Getis [45] is applied to "filter out" spatial dependencies from the data.

This section provides an empirical application focused on unemployment dynamics and its major constituent factors. Besides model estimation and interpretation, stability of the results is cautiously evaluated under varying neighborhood definitions. In this application example, regional unemployment dynamics at the NUTS2 level is analyzed for the following 10 countries: Austria, Belgium, Czechia, Denmark, Germany, Hungary, Luxembourg, the Netherlands, Poland and Slovakia. Figure 6.1 is provided for basic illustration of the regional unemployment modelled, its spatial distribution and regional dependencies.

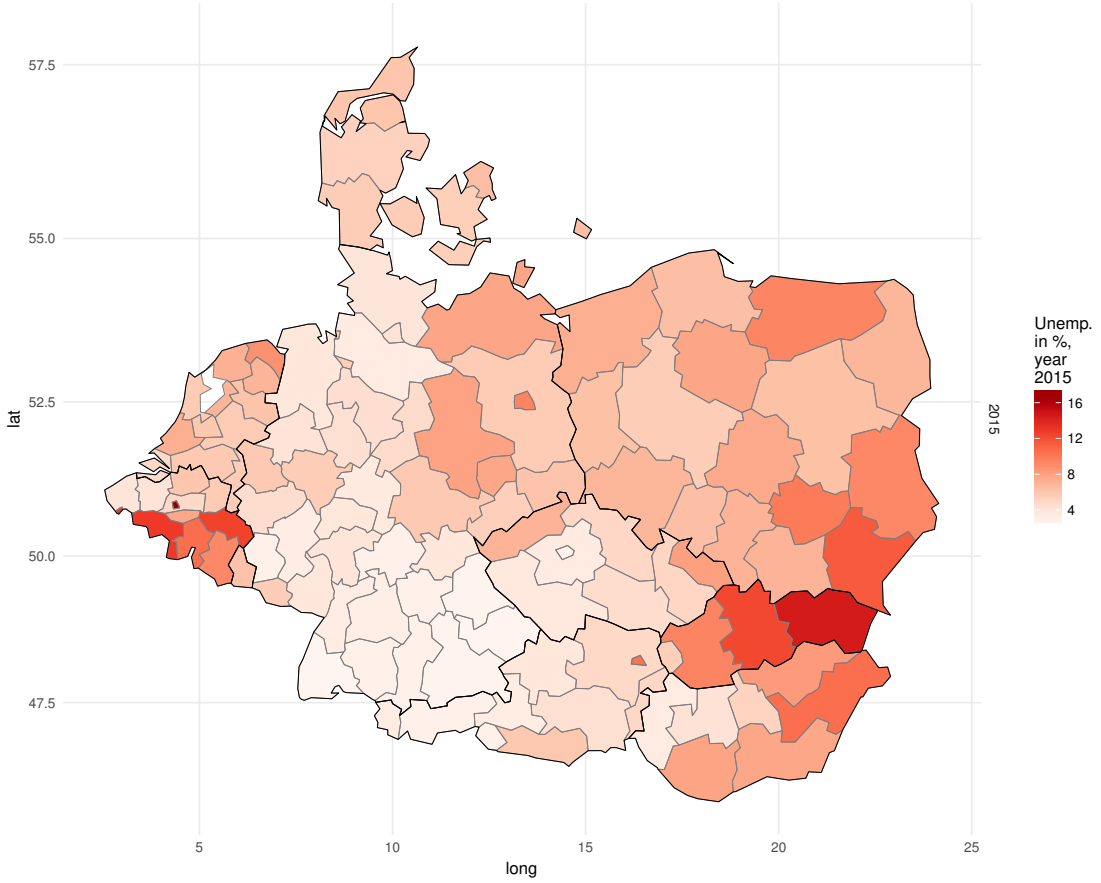


Figure 6.1.: Choropleth of 2015 unemployment rates – NUTS2 level. Source: Own calculation using GISCO – Eurostat data.

## 6.2. Methodology and data

The following empirical estimates are based on the Getis filtering approach. For detailed description of the underlying theory and data transformation (spatial filtering) process, see section 3.3.

Next, a linear regression model describing regional unemployment dynamics is presented. A relatively simple, yet informative and theoretically well-established model that draws from the "regional competitiveness' theory [39] is defined by equation (6.1), explaining unemployment dynamics in terms of its key determining factors: GDP per capita and two convenient labor-force structure and competitiveness indicators: the high-tech sector to total labor force ratio represents the relative predominance of this sector in economy.

Also, an analogous ratio for the sector of services (with a high proportion of unskilled labor input and predominant non-exportable i.e. non-tradable output) is presented. The model is defined as follows:

$$UN_i = \beta_0 + \beta_1\, GDP_i + \beta_2\, HTC_i + \beta_3\, Services_i + \varepsilon_i\,, \tag{6.1}$$

where $UN_i$ is the overall unemployment rate observed in the $i$-th spatial unit (specifically, we use NUTS2 unemployment data from the Eurostat's "lfst_r_lfu3rt" dataset), $GDP_i$ is the region's log-transformed GDP (euro per capita, "nama_10r_2gdp" dataset), $HTC_i$ describes the percentage of employees working in the high-tech industry (NACE r.2 code HTC, "htec_emp_reg2" dataset), $Services_i$ is the proportion of employees within the aggregated services sector (NACE r.2 code H-U, "htec_emp_reg2" dataset), $\beta_j$ are model parameters to be estimated and $\varepsilon_i$ is the random error. Equation (6.1) is estimated using observed 2015 data, based on a total of 111 NUTS2 regions (see figure 6.1) from Austria, Belgium, Czechia, Denmark, Germany, Hungary, Luxembourg, the Netherlands, Poland and Slovakia. Although observations in some spatial data series featured in model (6.1) may cover the period from 1999 to 2017, complete datasets for 2016 are not yet available from Eurostat at the NUTS2 level (as of March 2018).

## 6.3. Empirical results

Before actual model estimation, we begin by evaluating assumptions for the Getis filtering method: All the observed variables as per equation (6.1) exhibit positive support for our 111-unit sample. Also, all variables are positively spatially autocorrelated at the 5% significance level as tested using the Moran's $I$ statistic (2.7).

Table 6.1 shows key estimation outputs for model (6.1). Model estimation is provided in three versions: first, model (6.1) is estimated using spatially unfiltered data – output from a linear regression for a non-spatial model specification of (3.1) type is shown. Second, spatially filtered data are used – model (6.1) is estimated using spatially demeaned/filtered data as defined in equation (3.26), with maximum neighbor distance threshold $\tau$ set to 165 km. This threshold was set empirically, based on the maximized log-likelihood ($LL$) statistic ($LL$ has indicative properties only, please see following paragraphs for discussion of limits of the $LL$-based model comparison within the Getis filtering paradigm). The third estimation output also features spatially filtered data. Here, the $\tau = 217$ km threshold relates to an interesting locally optimal spatial setup (see figure 6.2 and – more importantly – it closely matches our previous findings [39] obtained

for a spatial lag model defined in terms of equation (3.2).

Table 6.1.: Estimated model, alternative spatial setups used

| Model type | Coefficients | Estimates | Std. Errors | t-values | Pr(>\|t\|) |
|---|---|---|---|---|---|
| Non-spatial | *(Intercept)* | 31.565 | 3.733 | 8.457 | 0.000 |
| linear | *GDP* | -4.046 | 0.503 | -8.041 | 0.000 |
| model | *HTC* | -0.208 | 0.139 | -1.498 | 0.137 |
| specification | *Services* | 0.240 | 0.036 | 6.659 | 0.000 |
| Data | *(Intercept)* | 7.513 | 6.740 | 1.115 | 0.268 |
| spatially | *GDP* | -1.088 | 0.782 | -1.390 | 0.167 |
| filtered with | *HTC* | -0.354 | 0.148 | -2.399 | 0.018 |
| $\tau = 165$ km | *Services* | 0.161 | 0.043 | 3.785 | 0.000 |
| Data | *(Intercept)* | 11.518 | 7.026 | 1.639 | 0.104 |
| spatially | *GDP* | -1.629 | 0.805 | -2.023 | 0.046 |
| filtered with | *HTC* | -0.311 | 0.141 | -2.210 | 0.029 |
| $\tau = 217$ km | *Services* | 0.180 | 0.043 | 4.224 | 0.000 |

Before discussing the estimation results as shown in table 6.1, we need to address spatial setups and the range of $\tau$ distance thresholds used for estimation. In order to evaluate model robustness and statistical properties under different spatial settings, equation (6.1) was estimated using spatially filtered data over a very extensive range of distance thresholds $\tau$: starting from a sparse spatial matrix $\boldsymbol{S}$, constructed for $\tau = 160$ km (lower $\tau$ values result in "islands" – regions without neighbors – that are incompatible with the logic of Getis-type filtering). Next, $\tau$ thresholds were iteratively increased by a 1-km step up to a maximum neighbor distance threshold of 1.000 km. Although the $\tau = 1.000$ km spatial setup is well beyond reasonably assumed spatial interactions of unemployment dynamics, it provides a nice illustration and comparison between spatial filtering and non-spatial estimation.

Overall, a total of 841 spatial specifications of our model (6.1) in its spatially filtered general form (3.26) were estimated. Those are summarized in figure 6.2 as follows: For each $\tau$, the corresponding Akaike information criteria (AIC) is shown, along with $LL$, $R^2$ and $\beta_j$ estimates (intercept excluded) along with their $\pm$ one standard error bands. Results from spatially filtered models are shown in blue, while the non-spatial estimation is shown in red for comparison (non-spatial estimate remains constant with respect to changing $\tau$ values).

Please note that the AIC, $LL$ and $R^2$ statistics are shown for illustrative purposes only – as spatially filtered (demeaned) values of the dependent variable differ across alternative

$\tau$ thresholds (i.e. between estimated models), statistics from different equations cannot be directly compared against each other. Finally, estimated values for $\tau = 160$ km and 217 km are highlighted by vertical dashed lines.

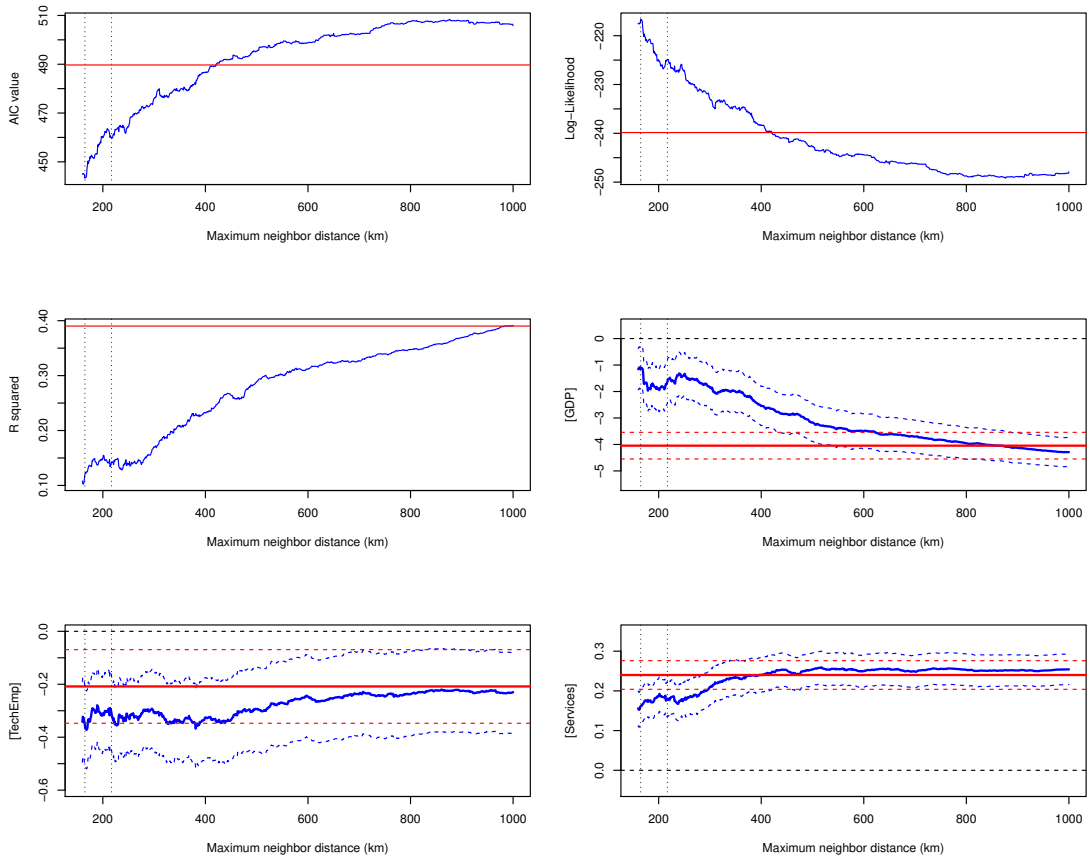


Figure 6.2.: Stability evaluation of the model under varying spatial setup. Source: Own calculation.

Results from table 6.1 and figure 6.2 provide several interesting insights into unemployment dynamics and its regional aspects. First of all, the expected effect of GDP and other explanatory variables on unemployment is significantly reduced once regional aspects are considered. This attenuation effect is most prominent for smaller $\tau$ values, where spatial filtering is concentrated to a compact neighborhood. This result reflects the fact that unemployment spillovers are often observed among contiguous (spatially adjacent) or otherwise very close neighbors.

Overall, most of the statistically significant differences between spatially filtered and unfiltered estimates disappear once $\tau$ approaches 500 km. Such results are in line with the theoretically presumed upper bound for unemployment spillover dynamics (spatial interdependence) where work commuting patterns and similar factors play a significant role. Please note that distances between regions are measured using centroids and actual inter-regional commuting distances may be significantly lower than the neighborhood threshold distance used in expression (2.4).

The estimated coefficients of model (6.1) are in line with theoretical expectations: we find evidence supporting an inverse relationship between unemployment and GDP growth. Also, prominent high-tech sector seems to lower the unemployment rate (at the 5% significance level and this effect is statistically significant in spatially augmented models only). Finally, we should stress out the fact that the estimated coefficients from a spatially filtered model such as (3.26) may be directly interpreted, unlike the $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ coefficients from the GNM specificatin (3.2), which do not constitute the usual marginal effects.

## 6.4. Discussion and conclusions

Spatially augmented models provide an important analysis framework where spatial patterns can be controlled for. As spatial dependencies are taken into account, we can see that many coefficients corresponding to "real world" regressors are attenuated (shrunk towards zero) in spatially augmented models. Along with the detected positive spatial dependence in observed data, such results have two main interpretations: First, there is a prominent influence of regional and potentially cross-border (international) factors in observed macroeconomic data that should be accounted for in quantitative analyses. Second, empirical results underline the importance of regional (cross-border) cooperation in macroeconomic policies.

While this analysis focuses on unemployment dynamics, regional interactions and spillovers are present in most macroeconomic variables and processes. Overall, Getis-type spatial filtering provides a relatively simple and interpretable toolbox for regional (spatial) analyses for a wide range of variables and research fields (macroeconomy, environmental studies, epidemiology, etc.).

# 7. Spatio-Temporal Analysis of Macroeconomic Convergence

This chapter is mostly based on the application part of [37] by Formánek.

## 7.1. Macroeconomic convergence: introduction

Macroeconomic convergence is often studied in terms of GDP per capita dynamics. Such approach is based on the neoclassical Solow-Swan model of long run growth and the corresponding analysis framework provided by Mankiw et al. [70]. Their approach leads to a convenient and empirically testable "$\beta$-convergence" model that estimates and evaluates the presumed inverse relationship between the growth rate of per capita output over a finite time period and the output level at the beginning of the period. This chapter focuses purely on $\beta$-convergence topics. However, other convergence analysis frameworks exist. For example, "$\sigma$-convergence" processes are described in [79].

The underlying hypothesis for $\beta$-convergence is quite simple: we assume that poorer economies take advantage of their potential and grow faster than the richer ones. In the long-run, this leads to wealth equalization among originally heterogeneous economies. Although the intuition behind $\beta$-convergence may be simple, empirically we are dealing with complex processes, prone to a continuous stream of diverse shock influences. Convergence-related topics can be approached and studied from many perspectives, e.g. focusing on different types of assumptions relevant for the convergence & growth dynamics, as discussed e.g. in [70]. In this chapter, the focus is on spatio-temporal aspects of macroeconomic convergence.

In recent GDP-growth literature, there is a prominent turn from cross-country analyses towards the sub-national scale; see Piras and Arbia [78] for examples and an exhaustive list of references. At the regional scale, closed-economy paradigms as in [70] are no longer appropriate – regional economies typically operate as prominently open and interconnected. Besides the Solow-Swan convergence mechanism, three main drivers of macroeconomic (NUTS2 level) convergence through regional interactions may be pointed

out: First of all, unification is institutionalized and incorporated in most EU policies. Also, factor mobility (labor, capital) and trade relations play an increasingly important role. Finally, technology & knowledge diffusion processes do provide a positive push to poorer regions.

Theoretically, the best way to control for such regional interactions would be to directly include labor, capital and goods movements, etc. into the growth models. In practical terms, such approach is impossible due to data availability issues, especially with variables such as inter-regional capital flows and technology diffusion. Here, spatial panel data methods may provide an indirect, yet feasible and reliable framework to regional growth and convergence analyses.

Spatial panel data methods – if properly applied – can correct for the inherent bias in classical cross-sectional growth models [78]. The bias generated by regional differences is controlled for by the explicit inclusion of individual (regional) effects that control for individual heterogeneities within the panel data paradigm [84]. Spatial interdependencies are also explicitly modeled in spatial models [28, 67]. As a consequence, spatial panel approach allows us to accurately differentiate between the two types of effects: individual and spatial.

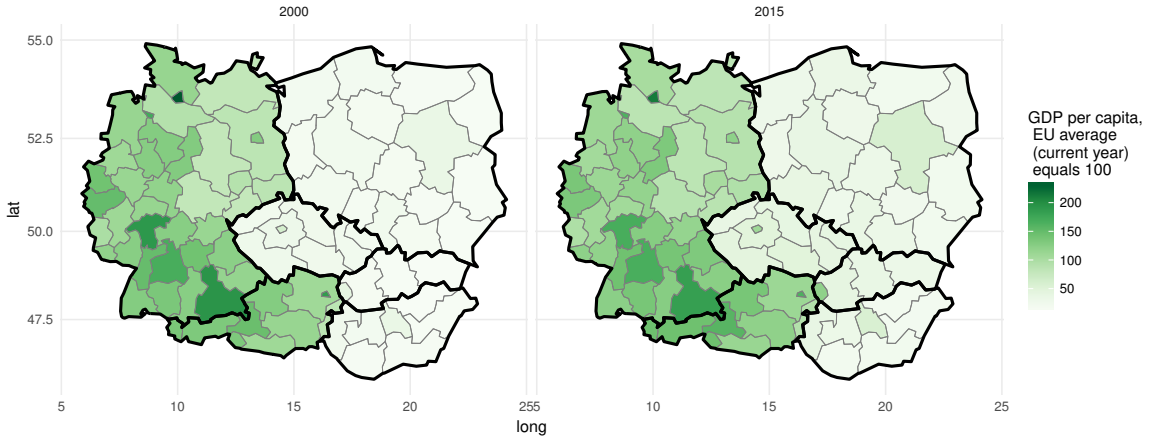


Figure 7.1.: Choropleths with 2000 & 2015 relative GDP per capita – NUTS2 level. Source: Own calculation using GISCO – Eurostat data.

This chapter provides a thorough integration of spatial modeling to the panel data-based analysis of regional convergence dynamics in terms of GDP per capita. In contrast with previous attempts in this field of research [78], proper interpretation of the ceteris-paribus

effects is used here (see section 7.2). The theoretical part of spatial panel analysis is accompanied by a $\beta$-convergence model describing regional growth dynamics at the NUTS2 level for the following six countries: Czechia, Slovakia, Poland, Hungary, Germany and Austria. For illustration of the convergence process, figure 7.1 compares relative GDP per capita levels as of 2000 and 2015: prominent & stable spatial patterns are apparent, while the presumed time convergence is not quite visually identifiable.

## 7.2. Macroeconomic convergence: empirical results

In this section, a relatively simple yet efficient spatial panel $\beta$-convergence model is established. Our model is based on annual growth dynamics and it follows the methodology used in [78]. However, it should be noted that the y-o-y dynamics is not ideal for addressing convergence processes that are a long-term phenomenon by nature. There is an ongoing discussion related to the estimation of $\beta$-convergence models using datasets covering relatively short time periods [78]. Wider time spans would increase the accuracy of tackling true convergence dynamics (instead of adjustments towards some trend after random shocks).

In fact, we face a trade-off here: while evaluating growth ($\beta$-convergence) over a longer time period (as compared to $t$-1 lags) is better for capturing true long-term growth dynamics, such approach considerably limits the number of observations available for model estimation.

Although data availability restrictions cannot be circumvented, we may take advantage of some rather non-restrictive assumptions: Our dataset features regions at diverse development stages – compare the GDP per capita among South-German regions, spatial units in the former East-Germany, Czechia and the Eastern parts of Poland. If properly controlled for (using spatial panel models), regional heterogeneity adds variability to the set of regressors and thus it can add reliability to our estimates, therefore somewhat compensating for the limited time-span currently available from Eurostat.

The model used for $\beta$-convergence evaluation may be outlined as

$$\log\left(\frac{y_{it}}{y_{i,t-1}}\right) = \lambda\left[\sum_{j=1}^{N} w_{ij}\log\left(\frac{y_{it}}{y_{i,t-1}}\right)\right] + \beta\log\left(y_{i,t-1}\right) + \mu_i + \upsilon_{it}\,, \tag{7.1}$$

where $y_{it}$ is the GDP per capita observed in the NUTS2 region $i$ at time $t$. Eurostat's "nama_10r_2_gdp" dataset is used, with annual 2000 – 2015 GDP observations recorded

in 2010-constant prices. All the observed data exhibit strong positive spatial autocorrelation when tested using the Moran's $I$ statistic.

Logarithmic transformation provides the desired growth-rate interpretation: the LHS of equation (7.1) is the annual growth of real per capita income. The first element on the RHS is the spatial lag and it follows from (4.1). Besides observed variables, coefficient $\lambda$ and $\beta$ along with the time-invariant & region specific effects $\mu_i$ (reflecting all omitted variables that influence the growth process) constitute the functional form of our spatial panel model. The $w_{ij}$ spatial weights (elements of $\boldsymbol{W}$) are constructed using the maximum neighbor distance rule, with threshold set to 170 km. $v_{it}$ is the error term with properties corresponding to equation (4.6). Although equation (7.1) contains the first time-lag of $y_{it}$, the model doesn't have an actual dynamic specification. In fact, equation (7.1) is an empirical implementation of the RE model (4.6). In model (7.1), negative $\beta$ coefficients are consistent with the presumed inverse relationship between growth rates and the lagged GDP level values.

Equation (7.1) is estimated using a balanced panel of 82 NUTS2 regions (displayed in figure 7.1) across 16 years. Hence, a total of 1.312 individual observations of GDP per capita are collected from the following EU members: Austria (9 NUTS2 regions), Czechia (8 regions), Germany (38 regions: from those 8 (plus Berlin) are from the former East-Germany), Hungary (7 regions), Poland (16 regions) and Slovakia (4 regions). For analysis and verification purposes, diverse modifications and restrictions are applied to model (7.1): all the specifications used for estimation are described below and summarized in Table 7.1.

Table 7.1 illustrates the importance of controlling for both region specific effects and spatio-temporal dynamics in $\beta$-convergence models. Eight different specification variants of equation (7.1) are used for estimation and comparison of the $\beta$ coefficients: (a) comes from a basic pooled regression estimate, with individual and spatial effects ignored – $\lambda$ is set to zero and the $\mu_i$ intercept is identical for all units. In (b), model (a) is augmented by two dummy variables: one controls for the 2009 drop in output due to the global economic crisis, while the other dummy variable distinguishes "old EU" regions – NUTS2 regions in Austria and in the former West Germany – from their post-communist counterparts. Although statistically significant, the two dummies are not reported in Table 7.1, as they only serve to filter out two most prominent inconsistencies in the data generating process (DGP) under scrutiny, i.e. to obtain accurate $\beta$-convergence indicators with proper ceteris-paribus validity.

Table 7.1.: Estimated alternative specifications of the $\beta$-convergence model

| Model specification (classical approach) | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| $\hat{\beta}$ | -0.00418 | -0.00393 | -0.01070 | -0.01070 |
| s.e.($\hat{\beta}$) | (0.00026) | (0.00036) | (0.00128) | (0.00128) |
| $t$ value | [-16.13869] | [-11.02666] | [-8.33446] | [-8.33078] |
| Pr($>\|t\|$) | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| **Model specification (spatially augmented)** | **(e)** | **(f)** | **(g)** | **(h)** |
| Direct impacts | -0.00216 | -0.00149 | -0.00116 | -0.00084 |
| simulated s.e. | (0.00027) | (0.00031) | (0.00023) | (0.00030) |
| $z$ score | [-8.04931] | [-4.83524] | [-5.07915] | [-2.78725] |
| Pr($>|z|$) | 0.00000 | 0.00000 | 0.00000 | 0.00532 |
| $\hat{\lambda}$ | -0.13135 | -0.07829 | 0.89381 | 0.87027 |
| s.e.($\hat{\lambda}$) | (0.06732) | (0.07547) | (0.01121) | (0.01320) |
| $t$ value | [-1.95118] | [-1.03744] | [79.70808] | [65.90509] |
| Pr($>|t|$) | 0.05104 | 0.29953 | 0.00000 | 0.00000 |

Column (c) corresponds to a panel model generalization of (a): with spatial effects omitted ($\lambda = 0$), but accounting for individual effects. In column (c), model (7.1) is estimated using the FE method as in [84]. Specification (d) augments (c) by using the same two auxiliary dummies as introduced in (b) – again, the reason is to add control variables for the two distinct influences affecting GDP growth and thus filtering them out from the the pursued $\beta$-convergence dynamics estimation.

Models (a) to (d) lack spatial dependency features, yet they serve for direct comparison with their spatially augmented counterparts (e) to (h). Specification (e) amends the pooled version of (7.1) by introducing a cross-sectional spatial lag as in equation (4.10). Any individual effects are ignored here. (f) differs from (e) by featuring dummy variables as in (b) and (d). Specification (g) is the spatial panel model equation (7.1): both region-specific effects and spatial dependencies are accounted for. Again, model (h) is obtained by incorporating our two dummies into (g).

All models presented in Table 7.1 are statistically significant at the 5% significance level and were subjected to the usual model testing and verification procedures as proposed in [71] and [84]. For example, the RE assumptions for (g) and (h) were tested using the Spatial Hausman $\chi^2$ test (4.24) as in [71] as well as evaluated by generalizing both models into a CRE specification using the Mundlak-Chamberlain approach from [72]

and [84]. The residual elements of models (g) and (h) exhibit no spatial autocorrelation at $\alpha = 0.05$. The dummy variables (omitted from Table 7.1) are statistically significant in all model specifications where they are used, pointing out the important differences in the DGP of growth in GDP per capita. However, economically speaking, dummies have only a limited effect on the estimates of $\beta$-convergence parameters.

Consistently negative estimates of the $\beta$-convergence parameters provide some confidence in the stability and robustness of the underlying convergence processes of the Solow-Swan type. Yet, the signs of estimated $\beta$ coefficients need to be put into perspective: On one hand, the results provide evidence in favor of the $\beta$-convergence mechanism considered. On the other hand, when the spatio-temporal dynamics of the DGP is fully and properly accounted for, $\hat{\beta}$ values "fall" by an order of magnitude (actually, they are attenuated towards zero thus reflecting a slower convergence speed). As we compare the estimated $\beta$-convergence parameters (c) against (g) and (d) against (h), we can see that convergence dynamics change from relatively weak to practically negligible. This situation is reflected in the formatting of Table 7.1 that features 5 decimal points, as the usual 3 material points would complicate model comparison.

The estimated spatial autocorrelation parameter $\hat{\lambda}$ in Table 7.1 describes the effect of spatial interactions – it quantifies the systematic pattern in spatial distribution of the GDP growth rates. As models (e) and (f) do not control for individual effects, we can see that the spatial dependency coefficient estimates are severely biased: the negative/insignificant $\hat{\lambda}$ estimates contradict to prior theoretical beliefs, to preliminary Moran $I$ test results as well as to evidence from other published works (e.g. [28], [39] and [78]). In contrast, with both spatio-temporal and individual effects properly accounted for, $\hat{\lambda}$ coefficients in (g) and (h) confirm the presence of strong regional spillovers and provide evidence supporting the convergence mechanisms based on spatial lags (presumably through factor mobility, trade and technological relationships, etc.).

## 7.3. Results discussion and robustness evaluation

Overall, the estimated models in Table 7.1 are built from the simplest specification (a) towards more realistic setups. Given all theoretical assumptions and the data-based evidence discussed above, we may conclude that models neglecting any (or all) of the unobservable effects (regional specificities and spatio-temporal effects) lead to severely biased results: the estimated $\beta$-convergence parameters in misspecified models are roughly 5 to 10 times stronger as compared to the properly specified spatial panel models (g) or

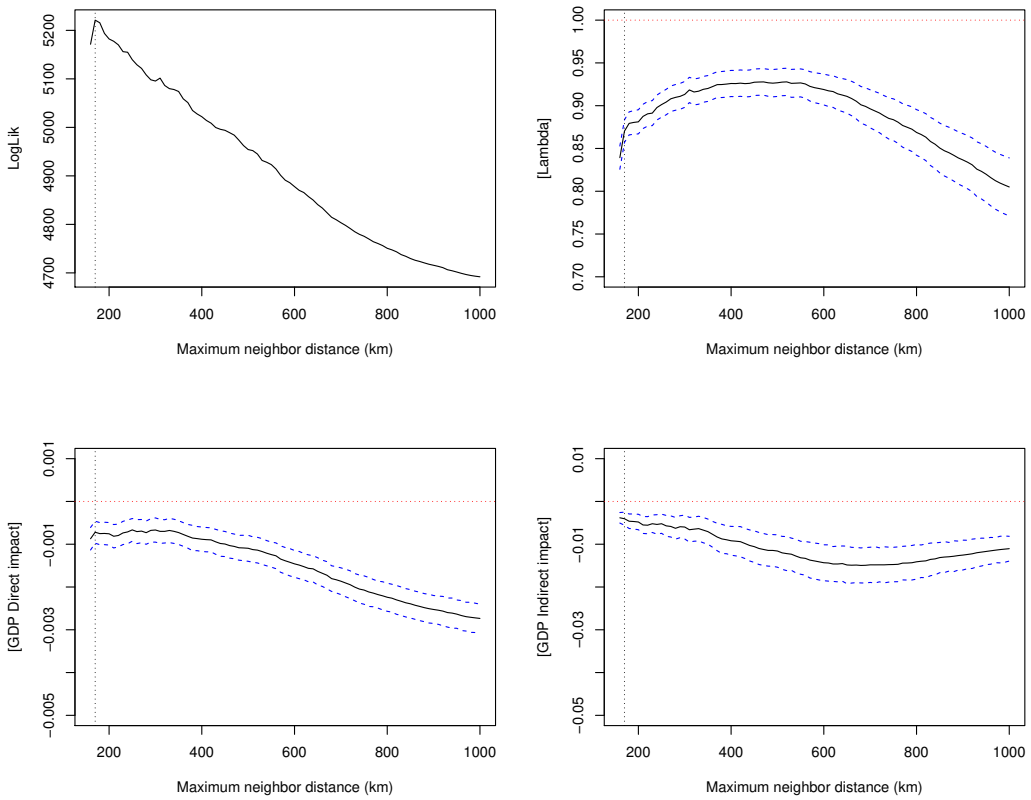


Figure 7.2.: Model stability evaluation: different $\boldsymbol{W}$ matrices considered. Source: Own calculation.

(h). The observed spatial interactions are much more prominent and influential when compared to the Solow-Swan type $\beta$-convergence dynamics.

Given the need for pre-specification of the $\boldsymbol{W}$ matrix (its $w_{ij}$ elements) in equation (7.1), as discussed in chapter 3.1, it is advisable to evaluate model stability against changes in the ad-hoc specified neighborhood definition. A simple yet effective approach is adopted and summarized in figure 7.2: model specification (h) as in Table 7.1 is estimated using alternative $\boldsymbol{W}$ matrices and results from different model setups are compared. The evaluation process starts with a relatively sparse spatial matrix constructed using a maximum neighbor distance threshold set to 160 km (lower thresholds generate disconnected units that are incompatible with the ML estimation of spatial models). Next, neighbor threshold distances are increased and new weights matrices are generated by iterations

of 10 km, up to a rather generous maximum neighbor distance of 1.000 km – beyond this threshold, the spatial properties of the model fall apart as the variance of spatial lag elements in (7.1) quickly falls to zero and spatial weak dependency assumptions are violated). At each iteration, the $\beta$-convergence model is estimated and recorded to figure 7.2: model log-likelihood values are shown, along with $\hat{\lambda}$, direct and indirect (spillover) effects and their asymptotic $\pm$ 1 s.e. bands. Finally, the maximized log-likelihood information from figure 7.2 is used to select the "best" $\boldsymbol{W}$ matrix from the 85 possibilities considered: the maximum neighbor threshold distance as used in in Table 7.1 is set to 170 km. Also, figure 7.2 provides enough confidence in overall model robustness.

## 7.4. Conclusions

This analysis pioneers the estimation and interpretation of impacts for spatial panel models in the context of macroeconomic $\beta$-convergence analysis. Compared to previous publications (e.g. [78]), this approach properly addresses the ceteris paribus effects in spatio-temporal models by focusing on the interpretation of direct impacts and spatial lag parameters instead of the $\beta$ coefficients of spatial regression models.

Considerable improvement is provided in comparison to the $\beta$ coefficients-based interpretation, which does not describe model dynamics properly under the spatial lag setup. Whenever spatial panel data are available, the framework presented here can extend the classical approach to $\beta$-convergence by controlling for both individual differences and spatial interactions.

The analysis provided in this chapter does not rule out the Solow-Swan type of macroeconomic convergence ($\beta$-convergence). However, it seems that this type of growth dynamics is more suitable for closed (large) economies. Using the appropriate spatio-temporal methodology, we can see that the regions analyzed exhibit prominent spatial convergence tendencies. The spatial part (spatial clustering) effects are much stronger than the Solow-Swan type $\beta$-convergence.

# 8. GDP Growth Factors and Spatio-temporal Interactions at the NUTS2 Level

This chapter draws from an unpublished article (2nd round of review, submitted to Journal of International Studies, eISSN 2306-3483) by Formánek.

## 8.1. Introduction and GDP growth theory

Over the last few years, many European countries have experienced considerable macroeconomic growth. Namely, Visegrad group countries and other smaller states in the vicinity of Germany have benefited greatly from cooperation with the strong and export-oriented industrial sectors of German economy and from the business opportunities originating therein. However, even during this period of economic expansion, actual growth rates differ significantly among regions. Figure 8.1 highlights the total variability of 2010 — 2016 GDP per capita growth (in 2015 real prices): while the average growth (calculated over the whole period) is 6.61 %, eight of the best performing German regions grew by 14 % or more. At the same time, 25 of the 113 NUTS2 regions grew by 3 % or less (again, calculated over the whole 2010 – 2016 period) and real GDP per capita actually decreased in 14 of these regions. Holland's region Groningen (NUTS2 code NL11) is the worst-performing spatial unit (outlier), clearly observable from figure 8.1. This unit has experienced a decline of 23.13 % in real GDP per capita, mostly due to the reduction of natural gas extraction over the past few years (see Eurostat, [31]).

Besides observed differences in macroeconomic growth during favorable times, both economic theory and historical experience suggest that tougher times may be lurking ahead. Also, crises often come suddenly and "unexpected" by mainstream economists. This paper does not search for signs of an upcoming slowdown period or crisis. Instead, differences in observed GDP per capita growth are explored and growth-driving factors are carefully examined with the aim of providing actionable suggestions that might be

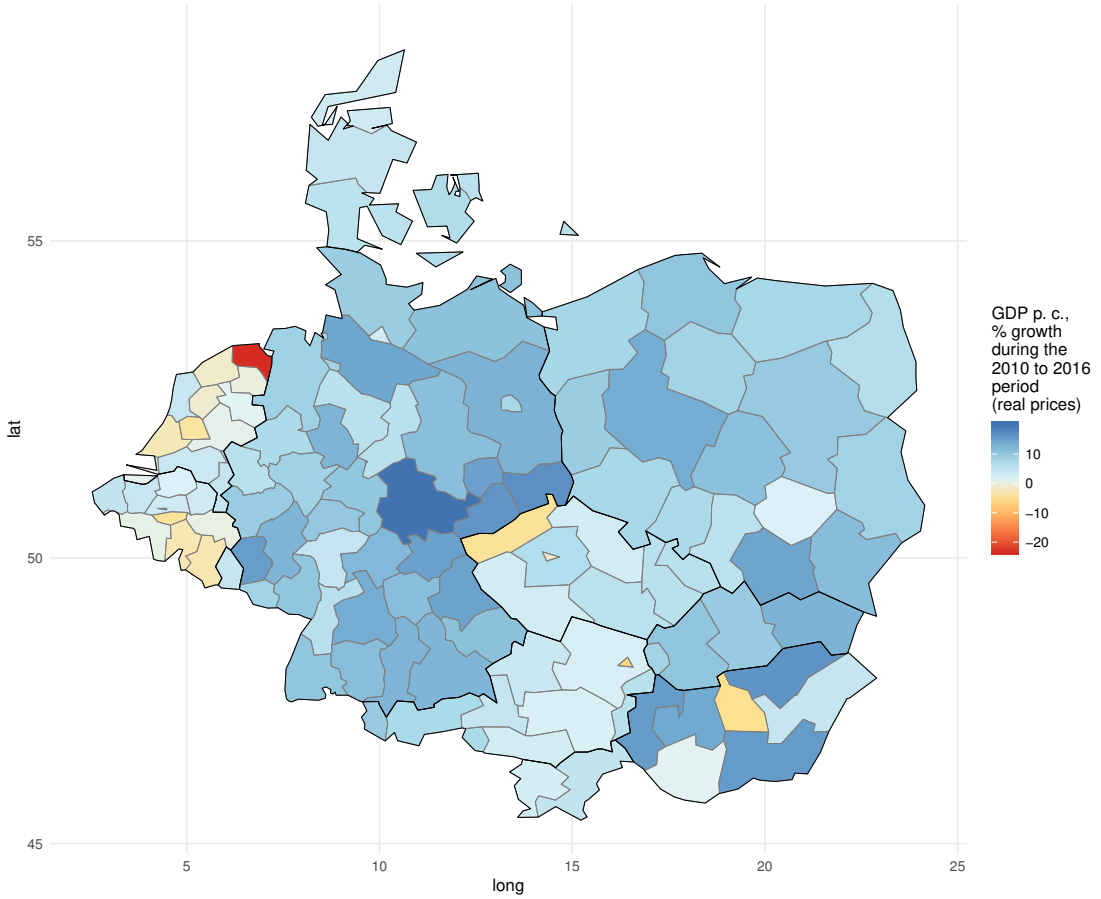


Figure 8.1.: Real GDP per capita growth (2010 – 2016), NUTS2, fixed prices (2015). Source: Own calculation using GISCO – Eurostat data.

used by policy makers in the near future. To provide such output, this analysis covers a sample of 11 relatively heterogeneous and spatially close EU member states at the NUTS2 regional level (113 regions) over a period of 7 years (2010-2016).

Geographically, historically and otherwise induced differences are a prominent feature in most regionally determined (geo-coded) data. In this paper, such differences are controlled at two levels: both individual (regional NUTS2) and state-level characteristic features (i.e. differences) are accounted for. This allows for structured and complex ceteris-paribus analysis of diverse theoretically and empirically established factors influencing macroeconomic growth (given e.g. in terms of GDP per capita changes). Indicators of labor force economic activity and structure are used, along with other variables

such as R&D expenditures, transportation infrastructure (density), etc. By combining the above-mentioned methodology and observed data, this chapter shows potential targets for economic policies aimed at boosting macroeconomic growth that are also useful for managing regional cohesion policies. Strong focus is given to evaluation of results' stability and robustness against changes in model specification and assumptions imposed. During the analysis, important data availability issues were present (see detailed discussion in section 8.3).

## Literature review

As one aims at studying macroeconomic growth and its dynamics, there are many diverse and valid approaches that can provide useful insight. Some authors even point out the lack of unifying paradigm concerning economic growth analysis, which persists even after decades of focused theoretical and empirical research (see e.g. [9]). Nevertheless, there are some basic elements that are generally deemed important for economic growth regardless of the theoretical framework used. For example, research and development (R&D) activities, measured in terms of R&D expenditures/investments are frequently considered in empirical studies and their effect has been repeatedly evaluated and tested (for focused analysis, see e.g. [14]).

To discuss economic growth theory, one would usually start with the neoclassical long-term model and data analysis approach (using highly aggregated variables), pioneered by Mankiw et al. [70] who laid ground to the widely used $\beta$-convergence approach that examines the inverse relationship between GDP per capita growth and its "base" value, measured at the beginning of some conveniently preset period. This methodology, generally based on the Solow-Swan model of long-run growth (see e.g. Solow, [80]), predicts convergence in growth rates on the basis that poorer economies would grow faster than richer ones. The $\beta$-convergence paradigm has become quite popular and many research papers have been published to date in this field; Piras and Arbia [78] provide one such contribution, along with extensive and representative references to peer papers.

Despite $\beta$-convergence's popularity, there are important contradicting theoretical approaches that predict the emergence and persistency of macroeconomic inequalities through self-reinforcing growth processes. For example, a growth theory based on cumulative causation that was first developed by Myrdal [75] predicts economic inequalities and imbalances as the most probable outcome of economic growth; the need for stabilizing (cohesion) economic policies is implied. Using a more sophisticated and formalized

methodology, the New Economic Geography (NEG) is a theoretically based approach that views initially more developed countries (regions) as benefiting from compound effects such as increasing returns to scale, manufacturing agglomeration, transportation costs, etc. Therefore, most NEG-based empirical papers (e.g. [42]) put great emphasis on spatial clustering (including economic specialization topics) and spillover effects.

Although the dispute between $\beta$-convergence and NEG approaches cannot be settled easily, the pro-convergence paradigm seems to be more prominent in current literature. Adopting a somewhat parallel perspective, various authors emphasize the role of "soft" (socio-cultural, knowledge-based and related) factors on economic growth. Jutting [58] provides a comprehensive analysis of institutions, institutionalized development-supporting mechanisms, their differences and/or bottlenecks that can explain inequalities in achieved growth rates.

Should we extend our attention beyond long-term aspects of economic growth, the analysis of short-term economic dynamics has many relevant implications as well. For example, Hamilton and Owyang [54] study macroeconomic co-movements and geographically defined differences across U.S. states. While focusing on short-term macroeconomic behavior – propagation of regional recessions -– they use Bayesian methods and analyze recession-timing differences and geographical clustering. Here, quarterly data (1956Q2 to 2007Q4) are used for modelling how regions (federal states) are entering recessions and recovering before/after others. While important strong nation-wide (common) components to most recessions are identified, individual heterogeneities turn out to play crucial roles in regional recession timing and intensity.

At the geographic scale (as opposed to short and long-term classification in time), methodological and data aspects of individual analyses may also differ significantly. Some authors use world-wide datasets to model output dynamics and its determinants: for example, Choudhry [20] uses a panel of 45 countries (highly diverse economies are followed for the period 1980 to 2005) to evaluate the effect of factors such as labor force participation, urbanization, information and communication technology (ICT) prevalence, etc. and their impact on macroeconomic growth in developed and developing countries. In contrast, Gauselmann et al. [43] provide a compelling analysis of foreign direct investments (FDI) within a relatively small area: NUTS2 regions in the Czech Republic, Poland and former East Germany. Using a proprietary "IWH FDI Micro database" of the Institute für Wirtschaftsforschung Halle organization, this analysis describes how agglomeration (i.e. clustering) influences FDI dynamics (a major factor of macroeconomic growth), while controlling for other key aspects such as production costs,

subsidies, etc.

Given the regional focus of this chapter, financial sector (individually observable at the state level, not regionally), its development and potential influences on economic growth are abstracted from. Nevertheless, the model given by equations (8.1) and (8.2) provides reasonable state-level differentiation and thus allows for implicit and separable control over both regional and state-wide effects. For specialized discussion of financial sector and its impact on economic growth, please refer e.g. to Beck [13] and to the literature listed therein.

This research attempt is not motivated in terms of searching for unification of growth theories or evaluating their validity. Instead, using a mainstream approach and generally accepted relevant assumptions, the focus is on providing empirical and actionable information concerning relevant economic growth factors in a spatially compact yet economically diverse group of countries (EU members) over a short-to-midterm time span. Methodology-wise, this application extends and improves the panel data approach – used e.g. in [12] – by accounting for spatial aspects and dependencies.

## 8.2. Methodology and data

As economic research scopes differ, spatial dependency definitions may differ accordingly. Using economic theory and statistical inference, researchers usually need to evaluate diverse spatial structure settings; both in terms of conceptual and parametric differences. The panel data spatial error model specification and estimation methodology used here follows from the theoretical work by Kapoor et al. see [59]. In its general form, the model assumes spatial correlation in both individual effects and the remaining error components. Although (8.1) may look similar to the specification (4.1), different spatial spillover mechanisms take place here:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}\,, \\ \boldsymbol{u} &= \rho\,(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{u} + \boldsymbol{\varepsilon}\,, \\ \boldsymbol{\varepsilon} &= (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N)\,\boldsymbol{\mu} + \boldsymbol{\nu}\,, \end{aligned} \tag{8.1}$$

where $\boldsymbol{u}$ is a compound and spatially dependent disturbance vector of panel model (8.1) and the structure of $\boldsymbol{\varepsilon}$ innovations allows for the individual innovations $\varepsilon_{it}$ to be correlated over time as $\boldsymbol{\mu}$ is a vector of unit-specific (time-invariant) elements, i.e. individual effects. The error component $\boldsymbol{\nu}$ varies both over cross-sectional units and over time. $\boldsymbol{\beta}$ and $\rho$ are the parameters of interest, estimated by ML approach [71]. Model

(8.1) belongs to a class of spatial error models (other specifications may involve spatial interactions in the dependent variable and/or spatial interactions among regressors) and its functional form was chosen (with respect to observed data) by means of specialized Lagrange multiplier tests for spatial dependency identification, introduced by Anselin et al. in [7].

A relatively simple yet informative and theoretically well-defined regression equation is used for estimating economic growth dynamics. Specification (8.2) was established iteratively (strong data availability issues are discussed separately), with the panel model in [20] serving as a starting point. In a simplified form – equivalent to the first row of model (8.1) – the regression equation may be outlined as follows:

$$\begin{aligned}\log(GDPpc_{it}) = \beta_0 &+ \beta_1 ActShFY15\text{-}64_{it} + \beta_2 Unem_{it} + \beta_3 \log(R\&D_{i,t-1}) \\ &+ \beta_4 \log(MWkmsq_{it}) + \beta_5 Y09GDPpc_i + \beta_6 RelEmpM\&N_{it} \\ &+ \beta_7 \left(Y09GDPpc_i \times RelEmpM\&N_{it}\right) + \boldsymbol{D}_i'\boldsymbol{\theta} + u_{it}\,,\end{aligned} \tag{8.2}$$

where $\log(GDPpc_{it})$ is the dependent variable: log-transformed GDP per capita (fixed prices, 2015) in a given NUTS2 region (113 regions, each identified by the $i$ index) observed at time $t = 2010, \ldots, 2016$. $ActShFY15\text{-}64_{it}$ is the ratio of economically active female population to total female population for the age group 15 to 64 years. $Unem_{it}$ is the unemployment rate, given as proportion (i.e. 0.03 instead of 3%) and $\log(R\&D_{i,t-1})$ describes R&D expenditures (in fixed 2015 prices) standardized to R&D per employee for consistent interpretation and log-transformed; $t-1$ lagged values are used to control for the empirically based delay between R&D expenditures and their effect on production.

Variable $\log(MWkmsq_{it})$ is calculated as the number of motorway kilometers per one thousand square kilometers or region's area (log-transformed observations) and it serves as a proxy for infrastructure quality (in terms of its relative abundance). $Y09GDPpc_i$ is the base year (pre-sample period) observation of the dependent variable (2009 GDP per capita in thousands EUR, 2015 prices) — it allows for evaluation of convergence processes as well as for controlling autocorrelation of the observed dependent variable in time. This variable changes between regions (but not across time) which is reflected in its subscript $(i)$.

$RelEmpM\&N_{it}$ is the ratio of employees in sectors $M$ (specialized professional, scientific and technical activities) and $N$ (general business support operations) as per the NACE rev. 2 Eurostat nomenclature [30]. Although both types of activities aim at streamlining and enhancing production and productivity, activities listed under section $M$ are de-

signed primarily to transfer specialized knowledge (activities in the $N$ section are not). The interaction term ($Y09GDPpc_i \times RelEmpM\&N_{it}$) allows to describe complex functional dynamics in the effects of its constituent components: the partial effect of one explanatory variable changes with the value of the other interacting regressor. $\boldsymbol{D}_i'$ is a $(1 \times 10)$ row vector of state-level (NUTS0) dummy variables that equal 1 if the $i$-th region (NUTS2) belongs to the corresponding state (NUTS0) and zero otherwise. This set of dummy regressors is used to control for country-specific differences in production (historically determined differences in macroeconomic structure, labor productivity inequalities, etc.). Germany serves as a reference country, thus it is excluded from this vector. All $\beta_j$ coefficients and the $(10 \times 1)$ vector $\boldsymbol{\theta}$ are parameters to be estimated and $u_{it}$ is the error term as defined in model (8.1). The presence of time invariant regressors in model (8.2) led to using the so-called random effects approach (for definition and testing of the assumptions involved, see [59] or [84]).

## Data

All data used for quantitative analysis are retrieved from the Eurostat database, thus ensuring consistency in observed variables. A balanced panel is used, with 113 NUTS2 regions across 11 states (Austria, Belgium, Czechia, Denmark, Germany, Hungary, Luxembourg, the Netherlands, Poland, Slovakia, Slovenia) and annual 2010 — 2016 observations. Although Eurostat has made a considerable progress in harmonization and availability of regional data (e.g. NUTS2 and NUTS3 levels), missing data are still a significant limiting factor for this type of empirical analyses. Also, regions located in unbroken (complete) study areas are necessary for spatial analysis, which limits data selection even further. Nevertheless, the dataset used covers a characteristic and diverse enough set of EU's economies over a reasonable time span, thus allowing for valid and representative statistical inference.

For reproducibility purposes, Eurostat identification codes for the data tables used are provided as follows: GDP per capita is retrieved from the "nama_10r_2gdp" dataset (including the base year observations), "lfst_r_lfp2act" is used for information on share of economically active female population (ages 15 – 64) and "lfst_r_lfu3rt" is used for unemployment rates. R&D expenditure data are based on "rd_e_gerdreg" and the corresponding standardization (R&D expenditures per employee) is performed using "lfst_r_lfe2en2". Transportation infrastructure data (motorways) are retrieved from "tran_r_net" and workforce structure data as per NACE rev. 2 comes from "lfst_r_lfe2en2". Conversion from nominal prices to 2015 real values was performed using "prc_hicp_aind"

(relevant for GDP per capita and R&D expenditure). All geographic data (shape-files, coordinates and areal information) come from Eurostat – GISCO [32].

For the sake of full disclosure, it should be noted that some theoretically valid and empirically proven variables [20] could not be used in model (8.2) because of missing data issues. Namely, the share of employees working in the ICT sector (section $J$ of the NACEr2 nomenclature), gross capital formation, railway infrastructure and other relevant datasets are not fully available at the NUTS2 level (i.e. not complete enough to make for a balanced panel dataset). Nevertheless, specification (8.2) is chosen to cover all relevant and measurable constituent factors affecting GDP and its growth dynamics.

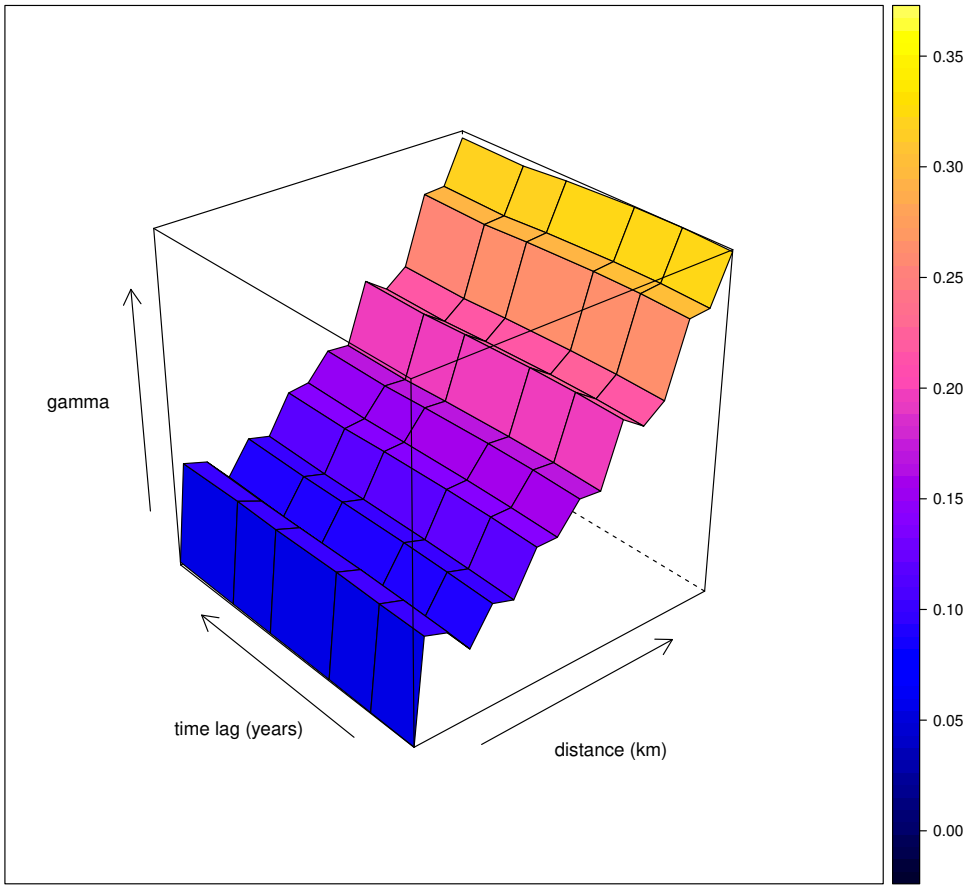


Figure 8.2.: Spatio-temporal semivariogram of log(GDP per capita), lags 0 to 6 years on the time axis and distances 0 to 500 km on the spatial axis. Source: Own calculation.

The dependent variable $\log(GDPpc_{it})$ exhibits strong positive spatial autocorrelation when tested using the Moran's $I$ statistic (4.2). However, the spatio-temporal semi-variogram (STSV) as per equation (1.17) can be used for a convenient description and visualization of both spatial and temporal variability aspects in observed data. Using some simplifying assumptions (for technical discussion and derivation of STSV, see [69] or [77]), we can easily establish an empirical version of (1.17) and assess variability and autocorrelation (spatial and/or temporal dependency) in observed data.

We may observe various important spatio-temporal properties from figure 8.2, which is an empirical STSV for the dependent variable in equation (8.2). First of all, observed data are highly persistent (autocorrelated) in time. Time lag-based increases in variability are relatively small for any fixed spatial distance. This data feature is reflected in model (8.1) specification, which accommodates temporal autocorrelation. Second, if we focus on the spatial axis, we can observe a pronounced increase in STSV values along increasing distances among observation. Next to plot's origin, the usual spatial "nugget" is present (in geo-statistics, it reflects micro-scale variations and/or measurement errors in data). We may see that $\gamma(\boldsymbol{s}, t)$ increases quite rapidly with spatial distance among observations: data are more similar to each other (less varied) in closer regions as compared against observations made farther apart in space. Figure 8.2 points towards a pronounced spatial autocorrelation (dependency) that dissipates over a relatively short spatial distance. This data property is also accommodated for in model (8.1). Please note that surface irregularities of the empirical STSV in Figure 8.2 simply reflect the stochastic and discrete nature of sampling; data grouping (along spatial and time distances) for variance calculation also plays some role here.

## 8.3. Empirical results and stability evaluation

For an intuitive percentage change interpretation of the estimated coefficients, dependent variable of model (8.2) is log-transformed. A potential drawback of using this transformation lies in the complicated prediction of original variables – model (8.2) predicts $\log(GDPpc_{it})$, not the original level values. However, this is only a minor concern as this analysis mainly focuses on evaluation of selected GDP growth driving factors.

Table 8.1 provides coefficient estimates for three alternative model specifications -– two spatial panel models with different $\tau$ values and one pooled-panel & non-spatial reference model. The first column (a) represents results from the "best" spatial model specification, as chosen by varying $\tau$ threshold (and thus $\boldsymbol{W}$ specification). Model evaluation is

performed by means of the maximized log-likelihood statistics, based on observed data and regressors as per equations (8.1) and (8.2) – see figure 8.3 and the next subsection for detailed discussion. Middle column (b) contains estimates obtained from an alternative $\boldsymbol{W}$ specification ($\tau = 177$ km instead of the 288 km in the first column). Arguably, (b) is the second-best specification (selected by comparing different spatial setups). Finally, the (c) column contains a base/reference model estimate with all spatio-temporal dynamics and individual effects ignored.

To keep this section compact and to avoid printing output with marginal relevance, Table 8.1 only features the estimated coefficients $\beta_1$ to $\beta_7$ along with the spatial autocorrelation coefficient $\rho$, which are deemed relevant for this article, i.e. for analyzing the dynamics of macroeconomic growth. Hence, the intercept and $\boldsymbol{\theta}$ coefficients (corresponding to dummy variables controlling state-level heterogeneities) are omitted. Nevertheless, table 8.1 contains all the relevant and empirically justified information necessary for discussing macroeconomic growth dynamic and its key constituent factors (while implicitly controlling for individual/NUTS2 and country-level/NUTS0 effects). Please note that given the ML estimation of model (8.1), the usual $R^2$ statistic is not applicable for model evaluation. Instead, the following statistic is used: $Pseudo\ R^2 = [corr(y_{it,observed}, y_{it,fitted}]^2$. For consistency, this applies to all columns of Table 8.1, although the distinction is not relevant for column (c).

The estimated $\rho$ coefficients in columns (a) and (b) of Table 8.1 suggest a very strong and highly statistically significant spatial dependency. From the theoretical perspective, this supports the overall validity of the methodology used (spatial panel data-based methods) and enables consistent estimates of the $\beta_j$ coefficients in spatial models. Empirically, high $\rho$ values underline the importance and prominence of spillover effects that serve as proxies for multiple minor and/or unobservable interaction mechanisms among neighboring regions and emphasize the significance and potential effectiveness of regional and cross-border cooperation in macroeconomic policy-making. In column (a) of Table 8.1, the coefficient $\hat{\beta}_1 = 0.2262$ may be interpreted as follows: given a one percentage point (pp) change in female labor-force participation, real GDP per capita would increase by 0.23 % (approximately). Similarly, for $\hat{\beta}_2 = -1.2748$, if unemployment ($Unem$) falls by 1 pp ceteris paribus, we would expect a 1.27 % rise in real GDP (and vice versa in the case of increasing unemployment rate). Lagged R&D expenditures have a positive and statistically significant effect on the expected overall GDP growth. On the other hand — given the relative sizes of both variables -– there is only a 0.02 % expected rise in GDP given a 1 % increase in R&D in the previous period (not 1 pp increase in R&D:

please note the difference in interpretation as R&D are log-transformed financial data, not ratio indicators).

As we compare the above discussed coefficients in column (a) to their counterparts in column (b), we can see that restricting neighbor interactions (by setting $\tau$ to 177 km) results in seemingly weaker spatial interactions and stronger ceteris paribus effects of individual regressors (coefficient estimates farther from zero). Nevertheless, data support $\tau = 288$ km, which may be observed by comparing log-likelihoods and *Pseudo* $R^2$ statistics. Interestingly, the ceteris paribus effect of highway infrastructure (its relative abundance as measured by $\log(MWkmsq_{it})$ is not statistically significant in any of the model specifications estimated, once other factors as in equation (8.2) are controlled for.

This contrasts with the commonly presumed boosting effects that infrastructure and corresponding investments have on GDP and its growth and also with the fact that the pairwise correlation coefficient for $\log(GDPpc_{it})$ and $\log(MWkmsq_{it})$ equals 0.63. Therefore, $\log(MWkmsq_{it})$ was not excluded from model specification (on grounds of statistical insignificance) because it provides economic insight and adds explicit control over an empirically important variable that is also a potential macroeconomic policy tool (through infrastructure investments). This particular result is somewhat unexpected, yet diverse empirical studies can often find evidence supporting opposite views.

Besides theoretical justification based on multiple economic growth concepts, the inclusion of base year GDP per capita level ( $Y09GDPpc$) has a sound technical reason as well: STSV in Figure 8.2 shows that the dependent variable of equation (8.2) is highly autocorrelated in time. Hence, the inclusion of $Y09GDPpc$ addresses temporal autocorrelation problems in model's residuals and helps with removing bias and inconsistency from the remaining $\beta_j$ coefficients in the model (by excluding the base GDP level, estimated coefficients of other regressors are roughly doubled in all columns of table 8.1). The coefficient for $RelEmpM\&N$ variable suggests a prominent positive effect of increased knowledge-based economic activities in a given economy/region: as the share of professional, scientific, organizational and similar employees increases, strong macroeconomic benefits are expected – even after controlling for regional and state-specific differences. Such result is in striking contrast with the effects of highway infrastructure. Please note that given the interaction element ( $Y09GDPpc \times RelEmpM\&N$), coefficients of the corresponding main effects (constituent variables present in the interaction) may not be interpreted on a ceteris paribus basis: their expected effects always depend on observed values of interacting regressors.

Table 8.1.: Alternative model specifications & estimates

| | Spatial panel model $\tau = 288$ km (a) | Spatial panel model $\tau = 177$ km (b) | Pooled non-spatial model (OLS) (c) |
|---|---|---|---|
| $\rho$ | 0.7724 | 0.6346 | – |
| (standard error) | (0.0395) | (0.0381) | |
| [p-value] | [0.0000] | [0.0000] | |
| $ActShFY15\text{-}64$ | 0.2262 | 0.3456 | 0.1886 |
| | (0.0989) | (0.1001) | (0.1134) |
| | [0.0222] | [0.0006] | [0. 0967] |
| $Unem$ | -1.2748 | -1.3931 | -1.7222 |
| | (0.1086) | (0.1081) | (0.1310) |
| | [0.0000] | [0.0000] | [0.0000] |
| $\log(R\&D_{t-1})$ | 0.0174 | 0.0235 | 0.0202 |
| | (0.0059) | (0.0057) | (0.0038) |
| | [0.0030] | [0.0000] | [0.0000] |
| $\log(MWkmsq)$ | 0.00001 | 0.0001 | 0.0002 |
| | (0.0002) | (0.0002) | (0.0001) |
| | [0.9455] | [0.5452] | [0.2130] |
| $Y09GDPpc$ | 0.0328 | 0.0322 | 0.0457 |
| | (0.0015) | (0.0014) | (0.0014) |
| | [0.0000] | [0.0000] | [0.0000] |
| $RelEmpM\&N$ | 1.0502 | 1.1723 | 6.5554 |
| | (0.3089) | (0.3011) | (0.3427) |
| | [0.0006] | [0.0001] | [0.0000] |
| ($Y09GDPpc\times$ | -0.0399 | -0.0395 | -0.1771 |
| $RelEmpM\&N$) | (0.0078) | (0.0076) | (0.0128) |
| | [0.0000] | [0.0000] | [0.0000] |
| $Pseudo\ R^2$ | 0.9815 | 0.9833 | 0.9913 |
| Log-likelihood | 1,542.382 | 1,538.925 | 1,129.148 |

The estimates in column (c) of table 8.1 are included mainly for reference: as spatial dependency and individual heterogeneities in the data are ignored, we can see a general tendency towards exaggeration of regressors' effects (OLS coefficients are farther from zero). Finally -– in terms of technical description of the estimated model – column (c) uses robust Newey-West standard errors [85] and the standard errors in spatial models (a) and (b) have asymptotic validity as described e.g. by Millo and Piras [71]. The rather high values of *Pseudo* $R^2$ coefficients in table 8.1 should be interpreted with caution, as they are mostly due to the presence of base (i.e. lagged) value of the dependent variable ($Y09GDPpc$) in the model.

### Model specification robustness: stability of results

Given the diverse options available for specification of the spatial weights matrix $\boldsymbol{W}$ in model (8.1), parameter estimates generally suffer from an implicit ambiguity potential and from identification problems. To address this issue, model robustness was evaluated against changes in neighborhood definitions. Using regressors from equation (8.2), multiple estimations of the spatial panel model (8.1) were performed, based on observed panel dataset while varying $\boldsymbol{W}$. Figure 8.3 provides a concise robustness evaluation summary; the information provided therein can be described as follows: The estimation starts with a sparse $\boldsymbol{W}$ matrix constructed using $\tau = 160$ km (lower i.e. more restrictive $\tau$ thresholds would generate unconnected regions – islands – that are incompatible with the ML estimation method) and then neighbor threshold distances are increased by iterations of 1 km, up to a maximum neighbor distance of 500 km. At each step, new $\boldsymbol{W}$ matrix is generated and the model is estimated. Overall, 341 alternative spatial structures and corresponding coefficient estimates are plotted in figure 8.3.

The relative instability of estimates at the lower end of the $\tau$ interval is not surprising: while 160 km is a feasible threshold (no island-regions are generated), such a short limit on neighbor interaction is too restrictive and the corresponding spatial structure is not realistic: $\boldsymbol{W}$ "prohibits" interactions among relatively close regions where spillovers and interactions are actually taking place. Similarly, very large distance thresholds (around 400 km or more) are not empirically justified either. Beyond the 400 km neighbor threshold, there is little theoretical and empirical evidence for the abundance of spatial interactions modelled and the log-likelihood statistics decrease quickly to levels that provide no improvement over non-spatial models. For reader's convenience, the "best" $\boldsymbol{W}$ specification (at $\tau = 288$ km) – as measured by the maximized log-likelihood statistics – is highlighted by a vertical dotted line in each element of figure 8.3. Hence, the dotted

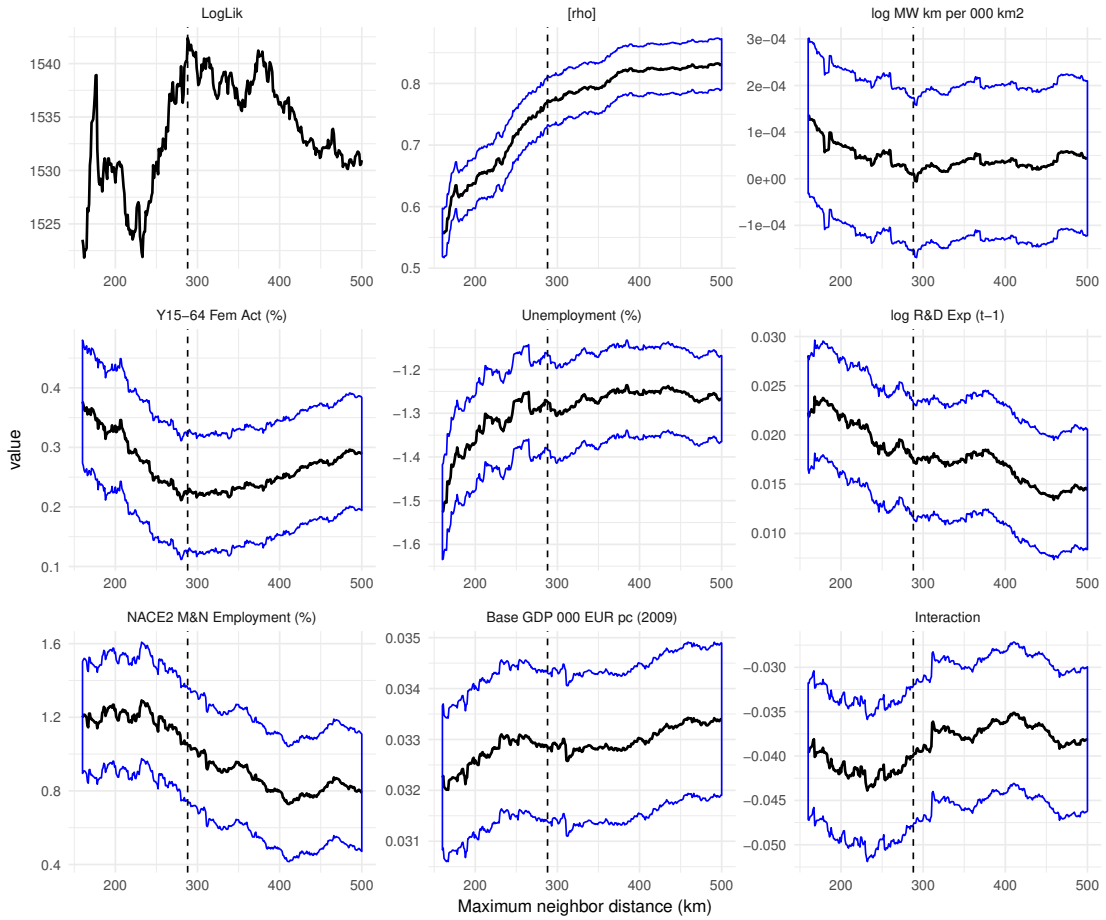


Figure 8.3.: Stability analysis of the estimated spatial error model. Source: Own calculation.

lines mark coefficient values in the (a) column of table 8.1. At $\tau = 177$, where the (b) column is generated, an isolated and unstable local maximum of the log-likelihood statistic may be observed in the top-left element of figure 8.3 (this result is not very robust against small changes in $\tau$ and $\boldsymbol{W}$ definition).

From the log-likelihood values shown in figure 8.3, one can see at least three local maxima that are associated with potentially diverging coefficient estimates and significance intervals. However, the estimated coefficients show reasonable overall stability (economically speaking) over a relatively large $\tau$ interval, roughly 250-350 km. Some coefficient estimates remain stable across even wider neighbor threshold intervals -– please refer e.g. to the *Unem* and $\log(R\&D_{i,t-1})$ variables. The spatial autoregression coefficient

$\rho$ increases along with the threshold parameter only up to values around $\tau = 375$ km. Beyond this threshold, even as we may try to involve more and more "neighbors" into the modelled spatial dynamics, we can see that $\rho$ values are rather unaffected. This – along with the decreasing log-likelihood values – clearly indicates that such extensive neighborhood definitions (interactions) are not supported by the observed data.

## 8.4. Conclusions

The aim of this chapter was to evaluate different GDP growth factors at the regional (NUTS2) level within a representative set of EU countries, in order to provide supporting and actionable material for macroeconomic decision-making processes that take place both at the regional and national levels. Despite important data availability limitations, a balanced panel dataset was gathered, covering 11 EU member states (geographically adjacent) at the NUTS2 regional level (113 regions) over a period of 7 years (2010-2016). Although the time period covered is generally considered as a period of economic growth, important differences in macroeconomic performance can be observed among the regions examined. Using a spatial panel model methodology, geographical determinations (spatial interactions) are discerned from the influences of relevant macroeconomic variables, many of which may be subject to or directly controlled by economic policy actions performed by central authorities. Factors such as female labor force participation, unemployment levels, motorway infrastructure, R&D expenditures and other macroeconomic variables are evaluated and their effect on GDP growth is estimated.

Besides the basic and theoretically conformable effects such as the inverse relationship between unemployment levels and GDP growth, this paper points out the importance of "smart" (knowledge-based) factors of economic growth such as R&D expenditures and promoting the increased share of professional, scientific, organizational and similar workers within the labor force structure. In contrast, motorway infrastructure (and the potential investments therein) exhibits no statistically significant effect on GDP growth, once other factors are controlled for. In addition to coefficient estimation, this paper also provides model stability evaluation that is used to assess robustness of the estimated coefficients with respect to varying definitions of spatial interactions. Overall, the empirical evidence gathered here supports specification robustness for the model used and its strong potential towards applications in similar fields of macroeconomic research.

# 9. Final remarks

Spatial econometric models provide an analysis framework where spatial and spatio-temporal aspects can be controlled for when estimating macroeconomic dynamics. Spatial models allow us to discern geographical determination from the influence of relevant macro-economic variables, many of which may be subject to or directly controlled by economic policy actions undertaken by the central authorities at different levels.

One potential drawback of spatial econometric analysis lies in the fact that spatial structure (spatial matrix) is not estimated along with model parameters. On the contrary: spatial structure has to be specified before model estimation. At the same time, we usually lack sufficient theoretical background (prior information) for choosing the "right" spatial setup.

For example, neighboring (mutually interacting) spatial units can be either based on contiguity evaluation or based on distances among units (plus, additional approaches are possible). Even if a specific method for spatial structure construction is chosen – e.g. based on distances – one needs to provide a controlling parameter: an ad-hoc distance value that is used for bounding spatial interactions among units.

The majority of current methodological and empirical contributions to spatial econometrics tend to downplay (or even ignore) problems and estimator instability issues involved with potentially flawed spatial structures being used in spatial models. In contrast, this contribution systematically focuses on this topic: In the methodology part, a relatively simple yet effective algorithm is provided for model robustness evaluation against changes in the underlying spatial structure. Subsequently, this algorithm is used in all three empirically based chapters (6 – 8).

This thesis is structured as follows: chapters 1 and 2 provide a brief introduction to the field of spatial analysis, basic terms are defined and described. Chapter 3 deals with geo-coded cross-sectional data, corresponding econometric models, their interpretation and estimation. Chapters 4 and 5 extend the discussion to panel data and models. Chapters 6 to 8 provide empirical applications, focused on macroeconomic dynamics of EU's NUTS2 regions, with emphasis on the Czech Republic and its neighbors (chapters

6 & 7 are based on published contributions [37, 38] and chapter 8 is based on a paper submitted to the Journal of International Studies.

# Appendix

# A. Supplementary materials

## A.1. Taxonomy of spatial models for cross sectional data

This section of the appendix contains a taxonomy of spatial models, reproduced from Elhorst [28].

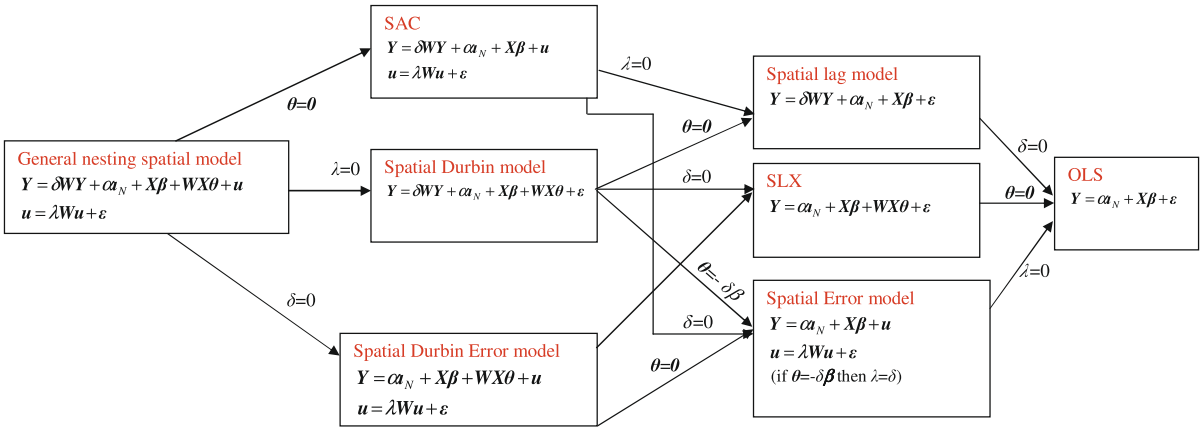


Figure A.1.: Different types of spatial dependency in models for cross-section data. Source: [28].

## A.2. Relationship between semivariogram and covariance

The (semi)variogram (1.9) is a generalization of the covariance function (1.6) and under second order stationarity the two functions are related:

$$\begin{aligned}
\gamma(\boldsymbol{h}) &= \frac{1}{2}E\left[(Z(\boldsymbol{s}+\boldsymbol{h})-Z(\boldsymbol{s}))^2\right], \\
&= \frac{1}{2}E\left[[(Z(\boldsymbol{s}+\boldsymbol{h})-\mu)-(Z(\boldsymbol{s})-\mu)]^2\right], \\
&= -E\left[(Z(\boldsymbol{s}+\boldsymbol{h})-\mu)(Z(\boldsymbol{s})-\mu)\right]+\frac{1}{2}E\left[(Z(\boldsymbol{s}+\boldsymbol{h})-\mu)^2\right] \\
&\quad +\frac{1}{2}E\left[(Z(\boldsymbol{s})-\mu)^2\right] \\
&= -C(\boldsymbol{h})+\frac{1}{2}C(\boldsymbol{0})+\frac{1}{2}C(\boldsymbol{0}) \\
&= -C(\boldsymbol{h})+C(\boldsymbol{0}) \\
&= C(\boldsymbol{0})-C(\boldsymbol{h})
\end{aligned}$$

## A.3. Specification of the $\varepsilon$ error element in spatial panel models

This section shows equivalency of the error term $\boldsymbol{\varepsilon}$ as provided in equations (4.1) and (4.6):

$$\begin{aligned}
\boldsymbol{\varepsilon} &= \rho(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{\varepsilon}+\boldsymbol{\upsilon}, \\
\boldsymbol{\varepsilon} &= (\boldsymbol{I}_T \otimes \rho\boldsymbol{W})\boldsymbol{\varepsilon}+\boldsymbol{\upsilon}, \\
\boldsymbol{\varepsilon}-(\boldsymbol{I}_T \otimes \rho\boldsymbol{W})\boldsymbol{\varepsilon} &= \boldsymbol{\upsilon}, \\
[\boldsymbol{I}_{NT}-(\boldsymbol{I}_T \otimes \rho\boldsymbol{W})]\boldsymbol{\varepsilon} &= \boldsymbol{\upsilon}, \\
[\boldsymbol{I}_T \otimes (\boldsymbol{I}_N-\rho\boldsymbol{W})]\boldsymbol{\varepsilon} &= \boldsymbol{\upsilon}, \\
(\boldsymbol{I}_T \otimes \boldsymbol{B}_N)\boldsymbol{\varepsilon} &= \boldsymbol{\upsilon}, \qquad \text{where } \boldsymbol{B}_N = (\boldsymbol{I}_N-\rho\boldsymbol{W}) \\
\boldsymbol{\varepsilon} &= (\boldsymbol{I}_T \otimes \boldsymbol{B}_N)^{-1}\boldsymbol{\upsilon}, \\
\boldsymbol{\varepsilon} &= \left(\boldsymbol{I}_T \otimes \boldsymbol{B}_N^{-1}\right)\boldsymbol{\upsilon}
\end{aligned}$$

# B. List of abbreviations

| | |
|---|---|
| AIC | Akaike information criteria |
| BIC | Bayesian information criteria |
| CRE | Correlated random effects |
| DGP | Data generating process |
| ECM | Error correction model |
| GDP | Gross domestic product |
| GLS | Generalized least squares |
| GMM | Generalized method of moments |
| GNS | Generalized nesting specification [of a spatial model] |
| ICT | Information and Communication Technology [industry-sector] |
| IVR | Instrumental variable regression |
| $k$NN | $k$-nearest neighbors |
| LAU | Local administrative unit |
| LHS | Left hand side [of an equation] |
| LL | Log likelihood [maximized value of LL function] |
| LM | Lagrange multiplier [based test] |
| MEM | Moran's eigenvector map |
| MSE | Mean squared error |
| ML | Maximum likelihood |

| | |
|---|---|
| NACE | Statistical Classification of Economic Activities in the European Community [French: Nomenclature statistique des Activités économiques dans la Communauté Européenne] |
| NEG | New Economic Geography |
| NUTS | Classification of Territorial Units for Statistics [French: Nomenclature des Unités Territoriales Statistiques] |
| OLS | Ordinary least squares |
| RHS | Right hand side [of an equation] |
| RSS | Residual sum of squares |
| SDM | Spatial Durbin model |
| SDPD | Spatial dynamic panel data [model] |
| SEM | Spatial error model |
| SLM | Spatial lag model |
| STSV | Spatio-temporal semivariogram |

# Bibliography


[1] L. Anselin. "Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity". In: *Geographical Analysis* 20.1 (1988), pp. 1–17. DOI: 10.1111/j.1538-4632.1988.tb00159.x.

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

[3] L. Anselin. *Spatial econometrics: methods and models.* Dordrecht: Kluwer, 1988.

[4] L. Anselin. *Spatial Regression Analysis in R. A Workbook.* Urbana-Champaign: University of Illinois, 2007.

[5] L. Anselin. "The Moran scatterplot as an ESDA tool to assess local instability in spatial association." In: *Spatial Analytical Perspectives on GIS.* Ed. by M. Fischer, H. Scholten, and D. Unwin. London.: Taylor and Francis, 1996.

[6] L. Anselin and S. J. Rey, eds. *Perspectives on Spatial Data Analysis.* Berlin: Springer Verlag, 2010.

[7] L. Anselin et al. "Simple diagnostic tests for spatial dependence." In: *Regional Science and Urban Economics* 26.1 (1996), pp. 77–104.

[8] M. Arellano and S. Bond. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." In: *Review of Economic Studies* 58.2 (1991), pp. 277–297.

[9] P. Arvanitidis, G. Petrakos, and S. Pavleas. "Determinants of economic growth: the experts' view." In: *Congress of European Regional Science Association.* Paris, 2007.

[10] B. H. Baltagi. *Econometric analysis of panel data.* New York: John Wiley, 2005.

[11] B. Baltagi, J. LeSage, and R.K. Pace, eds. *Spatial Econometrics: Qualitative and Limited Dependent Variables.* Vol. 37. Emerald Publishing Ltd, 2016.

[12] A. Bassanini and S. Scarpetta. *The driving forces of economic growth: panel data evidence for the OECD Countries.* OECD Economic Studies No. 33. 2001.

[13] T. Beck. *The Econometrics of Finance and Growth.* Policy Research Working Paper No. 4608. World Bank, Washington, DC. 2008.

[14] P. Bednář and M. Halásková. "Innovation performance and R&D expenditures in Western European regions: Divergence or convergence?" In: *Journal of International Studies* 11.1 (2018), pp. 210–224.

[15] A. Biswas and B. C. Si. "Model Averaging for Semivariogram Model Parameters." In: *Advances in Agrophysical Research.* Ed. by S. Grundas and A. Stepniewski. IntechOpen, 2013, pp. 81–96. DOI: 10.5772/52339.

[16] R. Bivand. "Creating Neighbours." In: *spdep R package vignette* (2016). URL: https://cran.r-project.org/web/packages/spdep/.

[17] R. Bivand. "Moran Eigenvectors." In: *spdep R package vignette* (2015). URL: https://cran.r-project.org/web/packages/spdep/.

[18] R. Bivand and G. Piras. "Comparing Implementations of Estimation Methods for Spatial Econometrics." In: *Journal of Statistical Software* 63.18 (2015), pp. 1–36. URL: http://www.jstatsoft.org/v63/i18/.

[19] M. Bun and M. Carree. "Bias-corrected estimation in dynamic panel data models." In: *Journal of Business & Economic Statistics* 3.2 (2005), pp. 200–210.

[20] M. T. Choudhry. *Determinants of Labor Productivity: An Empirical Investigation of Productivity Divergence.* Working paper, University of Groningen. 2009. URL: http://encuentros.alde.es/anteriores/xiieea/trabajos/pdf/104.pdf (visited on 07/19/2018).

[21] A. D. Cliff and J. K. Ord. *Spatial Processess: Models and Applications.* London: Pion, 1981.

[22] A. D. Cliff and J. K. Ord. "The problem of Spatial autocorrelation." In: *London Papers in Regional Science 1, Studies in Regional Science.* Ed. by A. J. Scott. London: Pion, 1969, pp. 25–55.

[23] N. A. C. Cressie. *Statistics for Spatial Data.* New York: Wiley, 1993.

[24] M. M. Deza and E. Deza. *Encyclopedia of Distances.* Berlin Heidelberg: Springer-Verlag, 2013.

[25] S. Dray, P. Legendre, and P. R. Peres-Neto. "Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM)." In: *Ecological Modelling* 196 (2006), pp. 483–493.

[26] J. P. Elhorst. "Dynamic Models in Space and Time". In: *Geographical Analysis* 33.2 (2001), pp. 119–140. DOI: 10.1111/j.1538-4632.2001.tb00440.x.

[27] J. P. Elhorst. "Serial and spatial error correlation". In: *Economics Letters* 100.3 (2008), pp. 422–424.

[28] J. P. Elhorst. *Spatial Econometrics: From Cross-sectional Data to Spatial Panels.* New York: Springer, 2014.

[29] J. P. Elhorst. "Specification and Estimation of Spatial Panel Data Models." In: *International Regional Sciences Review* 26.3 (2003), pp. 224–268.

[30] Eurostat. *NACE Rev. 2. Statistical classification of economic activities in the European Community.* 2008. URL: http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF (visited on 07/19/2018).

[31] Eurostat. *Regional Innovation Monitor.* 2018. URL: https://ec.europa.eu/growth/toolsdatabases/regional-innovation-monitor/base-profile/province-groningen (visited on 07/19/2018).

[32] Eurostat - GISCO. *Geographic Information System of the COmmission.* 2018. URL: http://ec.europa.eu/eurostat/web/gisco/.

[33] B. Fingleton. "Forecasting with dynamic spatial panel data: practical implementation methods." In: *Economics and Business Letters* 53.4 (2017), pp. 194–207. URL: https://doi.org/10.17811/ebl.3.4.2014.276-287.

[34] B. Fingleton, J. Le Gallo, and A. Pirotte. "A multidimensional spatial lag panel data model with spatial moving average nested random effects errors." In: *Empirical Economics* (2017). URL: https://doi.org/10.1007/s00181-017-1410-7.

[35] R. A. Fisher. *Design of Experiments.* Edinburgh: Oliver and Boyd, 1935.

[36] R. J. G. M. Florax and P. Nijkamp. "Misspecication in Linear Spatial Regression Models." In: *Tinbergen Institute Discussion Paper TI 2003-081/3* (2003), pp. 1–28.

[37] T. Formánek. "Spatially Augmented Analysis of Macroeconomic Convergence with Application to the Czech Republic and Its Neighbors". In: *Applied Computational Intelligence and Mathematical Methods. Part of the Advances in Intelligent Systems and Computing book series (AISC, volume 662).* Ed. by R. Silhavy, P. Silhavy, and Z. Prokopova. Cham: Springer, 2017, pp. 1–12.

[38] T. Formánek and R. Hušek. "Getis' non-parametric approach to spatial data analysis, with application to regional unemployment dynamics." In: *Mathematical Methods in Economics (MME2018), Jindřichův Hradec*. Ed. by L. Váchová and V. Kratochvíl. Praha: MatfyzPress, Publishing House of the Faculty of Mathematics and Physics Charles University, 2018, pp. 92–97.

[39] T. Formánek and R. Hušek. "On the stability of spatial econometric models: Application to the Czech Republic and its neighbors." In: *Mathematical Methods in Economics*. Ed. by A. Kocourek and M. Vavroušek. Liberec: TU Liberec, 2016, pp. 213–218.

[40] T. Formánek and S. Macková. "Spatio-Temporal Analysis of Macroeconomic Processes with Application to EU's Unemployment Dynamics". In: *Computational and Statistical Methods in Intelligent Systems*. Ed. by R. Silhavy, P. Silhavy, and Z. Prokopova. Cham: Springer International Publishing, 2019, pp. 120–131.

[41] A. S. Fotheringham, C. Brunsdon, and M. E. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Chichester: Wiley, 2002.

[42] M. Fujita, P. Krugman, and A. Venables. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge: MIT Press, 1999.

[43] A. Gauselmann, P. Marek, and J. P. Angenendt. *MNE's Regional Location Choice - A Comparative Perspective on East Germany, the Czech Republic and Poland.* IWH Discussion Paper No. 8. 2011.

[44] R. C. Geary. "The contiguity ratio and statistical mapping." In: *The Incorporated Statistician* 5 (1954), pp. 115–145.

[45] A. Getis. "Screening for spatial dependence in regression analysis." In: *Papers of the Regional Science Association* 69 (1990), pp. 69–81.

[46] A. Getis and J. K. Ord. "The Analysis of Spatial Association by Use of Distance Statistics." In: *Geographical Analysis* 24.3 (1992), pp. 189–206.

[47] T. Gneiting, M. Genton, and P. Guttorp. "Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry". In: *Statistical Methods for Spatio-temporal Systems (Monographs on Statistics and Applied Probability)* 107 (2006).

[48] B. Gräler, E. Pebesma, and G. Heuvelink. "Spatio-temporal interpolation using gstat." In: *R Journal* 8 (2016), pp. 204–218.

[49] W. Greene. *Econometric analysis. (5th ed.)* Upper Saddle River: Prentice Hall, 2002.

[50] D. A. Griffith. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization.* Berlin: Springer, 2003.

[51] D. A. Griffith and P. R. Peres-Neto. "Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses." In: *Ecology* 87.10 (2006), pp. 2603–2613.

[52] D.A. Griffith. *Spatial Autocorrelation: A primer.* Washington DC: Association of American Geographers, Resource Publications in Geography, 1987.

[53] J. Hahn and W. Newey. "Jacknife and analytical bias reduction for nonlinear panel models". In: *Econometrica* 72 (2004), pp. 1295–1319.

[54] J. D. Hamilton and M. T. Owyang. "The Propagation of Regional Recessions." In: *The Review of Economics and Statistics* 94.4 (2012), pp. 935–947.

[55] D. Harrison and D. L. Rubinfeld. "Hedonic prices and the demand for clean air." In: *Journal of Environmental Economics and Management* 5 (1978), pp. 81–102.

[56] J. A. Hausman. "Specification Tests in Econometrics." In: *Econometrica* 46 (1978), pp. 1251–1271.

[57] A. G. Journel and C. J. Huijbregts. *Mining Geostatistics.* London: Academic Press, 1978.

[58] J. Jutting. "Institutions and Development: A Critical Review." In: *OECD Development Center, Working Paper 210.* 2003.

[59] M. Kapoor, H. H. Kelejian, and I. Prucha. "Panel data models with spatially correlated error components." In: *Journal of Econometrics* 140.1 (2007), pp. 97–113.

[60] H. H. Kelejian and I. R. Prucha. "A generalized moments estimator for the autoregressive parameter in spatial model." In: *International economic review* 40.2 (1999), pp. 509–533.

[61] H. H. Kelejian and I. R. Prucha. "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances." In: *The Journal of Real Estate Finance and Economics* 17.1 (1998), pp. 99–121.

[62] H. H. Kelejian and I. R. Prucha. "On the Asymptotic Distribution of the Moran I Test Statistic with Applications." In: *Journal of Econometrics* 104.2 (2001), pp. 219–257. URL: doi:10.1016/S0304-4076(01)00064-1.

[63] L. F. Lee. "Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models." In: *Econometrica* 72.6 (2004), pp. 1899–1925.

[64] L. Lee and J. Yu. "Some recent developments in spatial panel data models." In: *Regional Science and Urban Economics* 40.5 (2010), pp. 255–271.

[65] P. Legendre. "Spatial Autocorrelation: Trouble or New Paradigm?" In: *Ecology* 74.6 (1993), pp. 1659–1673.

[66] P. Legendre and O. Gauthier. "Statistical methods for temporal and space – time analysis of community composition data." In: *Proceedings of the Royal Society B* 281.20132728 (2013).

[67] J. P. LeSage and R. K. Pace. *Introduction to spatial econometrics.* Boca Raton: CRC Press, Taylor & Francis Group, 2009.

[68] J. P. LeSage and R. K. Pace. "The biggest myth in spatial econometrics." In: *Econometrics* 2.4 (2014), pp. 217–249.

[69] C. Ma. "Spatio-temporal variograms and covariance models". In: *Advances in Applied Probability* 37.3 (2005), pp. 706–725. DOI: 10.1017/S0001867800000434.

[70] N. G. Mankiw, D. Romer, and D. N. Weil. "A contribution to the empirics of economic growth." In: *Q J Econ* 107.2 (1992), pp. 407–437. URL: doi:10.2307/2118477.

[71] G. Millo and G. Piras. "splm: Spatial panel data models in R." In: *Journal of Statistical Software* 47.1 (2012), pp. 1–38. URL: doi:10.18637/jss.v047.i01.

[72] K. Miranda, O. Martínez-Ibáñez, and M. C Manjón-Antolín. "Estimating individual effects and their spatial spillovers in linear panel data models." In: *Working Papers 2072/261028* (2015). URL: http://www.reunionesdeestudiosregionales.org/Reus2015/htdocs/pdf/p1294.pdf.

[73] P. A. P. Moran. "Notes on continuous stochastic phenomena." In: *Biometrica* 37 (1950), pp. 17–23.

[74] J. Mutl and M. Pfaffermayr. "The Hausman Test in a Cliff and Ord Panel Model." In: *Econometrics Journal* 14 (2011), pp. 48–76.

[75] G. Myrdal. *Economic theory and underdeveloped regions.* London: Hutchinson, 1957.

[76] B. Ou, X. Zhao, and M. Wang. "Power of Moran's I Test for Spatial Dependence in Panel Data Models with Time Varying Spatial Weights Matrices". In: *Journal of Systems Science and Information* 3.5 (2015), pp. 463–471. URL: https://ideas.repec.org/a/bpj/jossai/v3y2015i5p463-471n7.html.

[77] E. Pebesma. "spacetime: Spatio-Temporal Data in R." In: *Journal of Statistical Software* 51.7 (2012), pp. 1–30.

[78] G. Piras and G. Arbia. "Convergence in per-capita GDP across EU-NUTS2 regions using panel data models extended to spatial autocorrelation effects." In: *Statistica* 67.2 (2007), pp. 157–172. URL: doi:10.6092/issn.1973-2201/3513.

[79] S. J. Rey and J. Le Gallo. "Spatial analysis of economic convergence." In: *Palgrave Handbook of Econometrics: Volume 2, Applied Econometrics.* Ed. by T. C. Mills and K. Patterson. Basingstoke, United Kingdom.: Palgrave Macmillan, 2009, pp. 1251–1290.

[80] R.M. Solow. "A contribution to the theory of economic growth." In: *Quarterly Journal of Economics* 70 (1956), pp. 65–94.

[81] M. Tiefelsdorf. *Modeling spatial processes: Tthe identification and analysis of spatial relationships in regression residuals by means of Moran's I.* Berlin: Springer-Verlag, 1999.

[82] M. Tiefelsdorf and D. A. Griffith. "Semiparametric filtering of spatial autocorrelation: the eigenvector approach". In: *Environment and Planning A* 39.5 (2007), pp. 1193–1221. DOI: doi.org/10.1068/a37378.

[83] A. Ullah and D. E. A. Giles, eds. *Handbook of Applied Economic Statistics.* New York: Marcel Dekker, 1998.

[84] J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press, 2010.

[85] J. M. Wooldridge. *Introductory Econometrics: A Modern Approach, Sixth Edition.* Boston: Cengage Learning, 2016.

[86] J. Yu and L. Lee. "Estimation of unit root spatial dynamic panel data models." In: *Econometric Theory* 26.5 (2010), pp. 1332–1362.

# Subject index

# Name index