

GDP per capita in selected EU countries: Economic growth factors and spatio- temporal interactions examined at the NUTS2 level

Tomáš Formánek

Department of Econometrics, University of Economics, Prague,

Czech Republic

formanek@vse.cz

ORCID 0000-0003-1638-7230

Abstract. This article evaluates the macroeconomic policy tools at which central authorities may focus in times of economic slowdown and/or to boost economic cohesion. For this purpose, a complex spatio-temporal regression model is provided, which controls for spatial and temporal dependencies in data, as well as individual effects and relevant regional macroeconomic conditions. Empirical findings firmly favors GDP-boosting policy actions oriented towards supporting knowledge-based economic activities: enhancing research and development, improving labor force structure towards specialized professional, scientific and technical activities. In contrast, transportation infrastructure (proxied by motorway density) does not enact statistically significant improvement in GDP growth. In this article, regional macroeconomic growth dynamics and its key constituent factors are analyzed using the sample of 11 EU member states (Austria, Belgium, Czechia, Denmark, Germany, Hungary, Luxembourg, the Netherlands, Poland, Slovakia, Slovenia), observed annually (from 2010 to 2016) at the NUTS2 level. Substantial effort is directed towards verification of model robustness and the corresponding validity of the conclusions drawn.

Received:
November, 2018
1st Revision:
January, 2019
Accepted:
February, 2019

DOI:
10.14254/2071-
8330.2019/12-1/8

Keywords: spatial panel data, GDP growth determinants, NUTS2 regions, spatial filtering.

JEL Classification: C23, C52, E66, F43.

1. INTRODUCTION

Over the last few years, many European countries have experienced considerable macroeconomic growth. Namely, the 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, the actual growth rates differ significantly among the regions. Figure 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 outlier, as it is clearly observable in Figure 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 (Eurostat, 2018a).

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

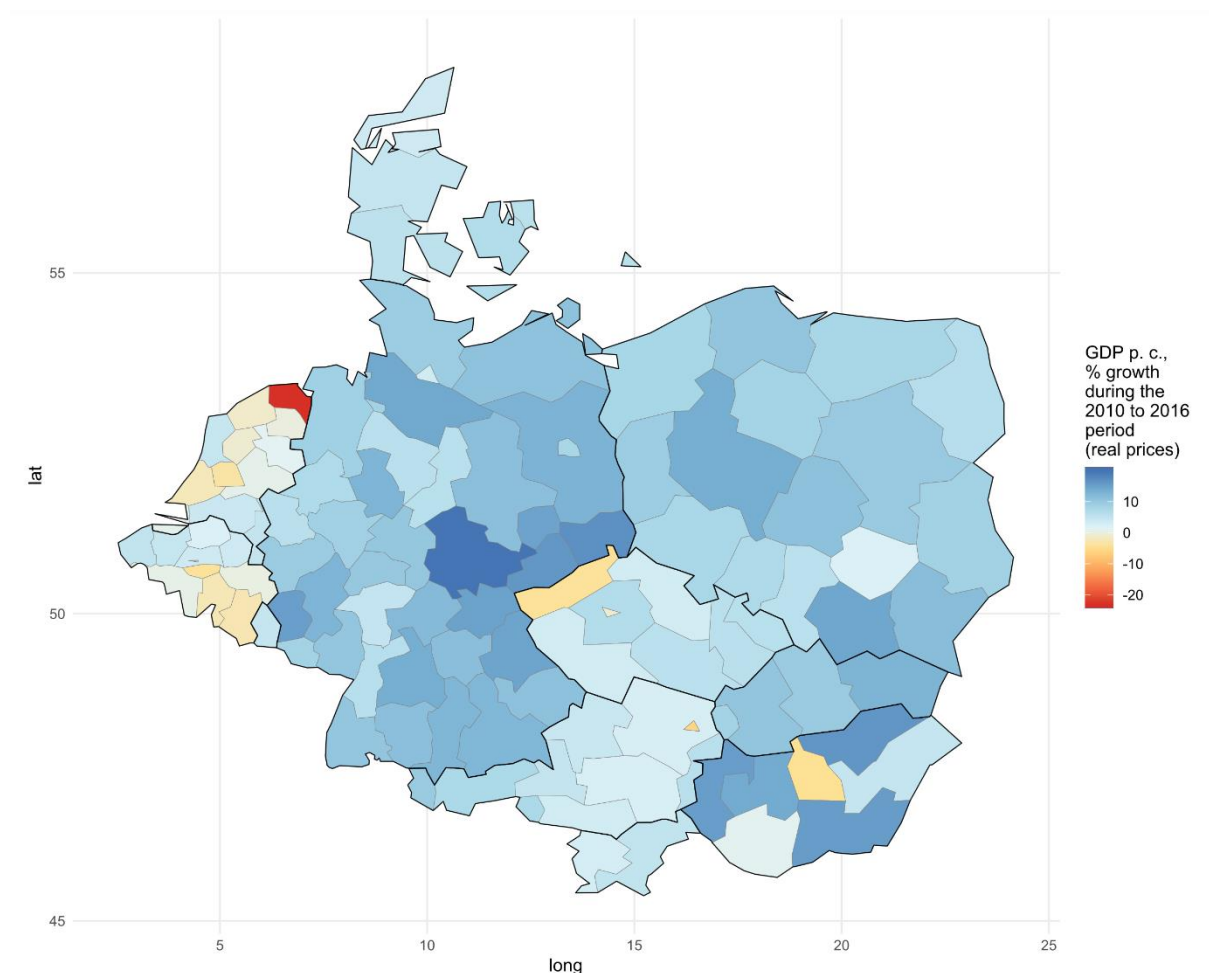


Figure 1. Choropleths showing real GDP per capita growth (2010–2016), using NUTS2 regional aggregation and fixed prices (as of 2015).

Source: Eurostat (2018b).

Spatial econometrics is used to control for the factors that would be inaccessible by other approaches to modelling economic growth and its dynamics (autocorrelation is present both in time and among spatially close observations). Although a large proportion of spatial effects (spatial dependencies) can be attributed to the omitted variables, this does not impede the use of spatial models, nor the interpretation of the estimated coefficients (LeSage & Pace, 2009). Spatial interactions are often conveniently used as proxies for many theoretically sound processes that generally evade informative and harmonized quantification and consequently incorporation into estimable models. At the regional level, it is usually empirically inaccessible to consistently measure factor mobility (labor, capital), technology and knowledge diffusion, language differences and other subtle factors that influence economic interactions. Nevertheless, quantitative methods such as spatial panel data analysis can efficiently correct for the inherent bias in classical cross-sectional or panel models describing regional macroeconomic growth and its dynamics. (For a detailed discussion see, e.g., Piras & Arbia, 2007).

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 a structured and complex *ceteris-paribus* analysis of diverse theoretically and empirically established factors influencing macroeconomic growth (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 abovementioned methodology and observed data, this study 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 3.2).

The remainder of this paper is structured as follows: the next section provides a condensed review of relevant literature, focusing on recent research papers and current discourse in the field of macroeconomic growth and regional analyses. Chapter three provides a relatively brief discussion of spatial panel data and corresponding econometric approaches applied in this paper, along with references to specialized technically oriented textbooks and papers. Detailed discussion of the dataset used for estimation is given. Chapter four outlines the main empirical findings and provides structured discussion of the results from alternative model specifications. Comprehensive model stability and robustness analysis is presented. Conclusions and the list of references finalize the paper.

2. 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, which persists even after decades of focused theoretical and empirical research (Arvanitidis et al., 2007). 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 theoretical and empirical studies and their effect has been repeatedly evaluated and tested. For focused analyses, see e.g. Bednář and Halásková (2018), Prokop and Karbowski (2018) or Romer (1990).

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. (1992) who laid groundwork to the widely used β -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 (Solow, 1956), predicts convergence in growth rates on the basis that poorer economies would grow faster than richer ones. The β -convergence paradigm has become quite popular and many research papers have been published to date in this field; Piras and Arbia (2007) provide one such contribution, along with extensive and representative references to peer papers. Despite β -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 (1957) predicts economic inequalities and imbalances as the most probable outcome of economic growth; the need for stabilizing and 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 (Fujita et al., 1999) put great emphasis on spatial clustering (including economic specialization topics) and spillover effects. Although the dispute between β -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 – namely Jutting (2003) provides a comprehensive overview and analysis of institutions, institutionalized development-supporting mechanisms, their differences and/or bottlenecks that can explain inequalities in achieved growth rates.

Besides the importance of long-term aspects of economic growth, analysis of short-term economic dynamics has many relevant implications as well. For example, Hamilton and Owyang (2012) 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 other regions. While important strong national (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 (2009) 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, Gausemann et al. (2011) 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 article, 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 (1) and (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 (2008) and to the literature listed therein.

Using a holistic approach that considers both economic growth and its environmental impacts, Nordhaus (1992, 2017) provides tools for designing macroeconomic policies aimed at general sustainability (long term economic and environmental) and towards slowing of global warming.

This article 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 article extends and improves on the pooled panel data (i.e. pseudo-panel) approach applied e.g. in Bassanini and Scarpetta (2001). In the current article, spatial, temporal and individual (unobservable) aspects and dependencies are properly addressed and incorporated into the model. This methodology allows for adequate *ceteris paribus* interpretation of the macroeconomic dynamics under scrutiny.

3. METHODOLOGY

Spatial dependency and spatial models play an ever more important role in regional studies and throughout many non-economic research fields (ecology, epidemiology, etc.). 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. As regional observations are often repeated in time, spatial panel models can be used to describe different interactions among variables: across regions and throughout time. The panel data spatial model specification and estimation methodology used in this article follow from the theoretical work by Kapoor et al. (2007). In its general form, the model may be outlined as:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \\ \mathbf{u} &= \rho(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{u} + \boldsymbol{\varepsilon}, \\ \boldsymbol{\varepsilon} &= (\mathbf{I}_T \otimes \mathbf{I}_N)\boldsymbol{\mu} + \mathbf{v}, \end{aligned} \quad (1)$$

where \mathbf{y} is a $NT \times 1$ vector (panel data with $i = 1, 2, \dots, N$ spatial units and $t = 1, 2, \dots, T$ time periods) of dependent variable observations, \mathbf{X} is a $NT \times k$ matrix of observations of exogenous regressors, \mathbf{I}_T and \mathbf{I}_N are identity matrices with dimensions corresponding to their subscripts, $\mathbf{1}_T$ is a $T \times 1$ vector of ones and \otimes is the Kronecker product operator. \mathbf{u} is a compound and spatially dependent disturbance vector of panel model (1) and the structure of $\boldsymbol{\varepsilon}$ innovations allows for the individual innovations ε_{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 \mathbf{v} varies both over cross-sectional units and over time. $\boldsymbol{\beta}$ and ρ are the parameters of interest, estimated by a maximum likelihood (ML) approach (Milo & Piras, 2012). \mathbf{W}_N is a $N \times N$ matrix of pre-defined spatial weights. For any two spatial units i and j , individual \mathbf{W}_N elements can be constructed as follows (multiple alternative approaches exist, see e.g. Formánek, 2018):

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

where h_{ij} is the distance between units and τ is an ad-hoc maximum neighbor distance threshold that maintains all sample units connected – at least by a minimum spanning tree. Units farther apart than τ are not considered neighbors and therefore do not interact (also, spatial units are not neighbors to themselves).

by definition). For detailed technical discussion on model (1) identification and stability, please refer to Kapoor et al. (2007) or Elhorst (2014). It should be stressed out that \mathbf{W}_N matrices are not estimated along with model parameters. On the contrary: \mathbf{W}_N needs to be specified before model estimation. Also, usually there is little theoretical background for choosing the “right” \mathbf{W}_N specification (in practice, different τ -values can be used for bounding spatial interactions among units). Hence, we often estimate spatial panel models like (1) using different spatial weights matrices \mathbf{W}_N by choosing a “sensible” range of alternative τ values to assess model performance and stability under alternative spatial interaction patterns.

Model (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. (1996).

3.1. Empirical model

A relatively simple yet informative and theoretically well-determined regression equation is used for estimating economic growth dynamics. Specification (2) was established iteratively (strong data availability issues are discussed separately), with the panel model in Choudhry (2009) serving as a starting point. In a simplified form – equivalent to the first row of model (1) – the regression equation may be outlined as follows:

$$\begin{aligned} \log(GDPpc_{it}) = & \beta_0 + \beta_1 ActShFY15-64_{it} + \beta_2 Unem_{it} + \beta_3 \log(R\&D_{i,t-1}) \\ & + \beta_4 \log(MWkmsq_{it}) + \beta_5 Y09GDPpc_i + \beta_6 RelEmpM\&N_{it} \\ & + \beta_7 (Y09GDPpc_i \times RelEmpM\&N_{it}) + \mathbf{NUTS0}'_i \boldsymbol{\theta} + u_{it} \end{aligned} \quad (2)$$

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 $t = 2010, \dots, 2016$. $ActShFY15-64_{it}$ is the ratio of economically active female population to total female population for the age group 15 to 64 years and it serves a proxy variable for labor market development and effectiveness situation (Scharpf & Schmidt, 2000). $Unem_{it}$ 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. The 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 variables 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 (2008) nomenclature. Although both types of activities aim at streamlining and enhancing production and productivity, activities listed under section M are designed 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. $\mathbf{NUTS0}'_i$ is a (1×10) row vector of state-level (NUTS0) dummy variables that equal 1 if the i -th region belongs to the corresponding state 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 β_j coefficients and the (10×1) vector θ are parameters to be estimated and u_{it} is the error term as defined in model (1). The presence of time invariant regressors in model (2) led to using the so-called random effects approach (for definition and testing of the assumptions involved, see Wooldridge, 2010 and Kapoor et al., 2007).

3.2. Data

All data used for quantitative analysis are retrieved from the Eurostat (2018b) 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 the sake of reproducibility, 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 (2018).

For the sake of full disclosure, it should be noted that some theoretically valid and empirically proven variables (see e.g. Choudhry, 2009) could not be used in model (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 (2) is chosen to cover all relevant and measurable constituent factors affecting GDP and its growth dynamics.

The dependent variable $\log(GDPpc_{it})$ exhibits strong positive spatial autocorrelation when tested using the Moran's I statistic (for definition and application examples, see e.g. Bednář and Halásková, 2018). However, the spatio-temporal semivariogram (STSV) as per equation (3) 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 Ma, 2005 or Pebesma, 2012), we can easily establish an empirical version of (3) and assess variability and autocorrelation (spatial and/or temporal dependency) in observed data. In general terms, STSV is declared using the following formula:

$$\gamma(\mathbf{s}; t) = \frac{1}{2} \text{var}[Z(\mathbf{s}_0 + \mathbf{s}; t_0 + t) - Z(\mathbf{s}_0; t_0)]; \quad (\mathbf{s}, t) \in \mathbb{R}^2 \times \mathbb{R}, \quad (3)$$

where \mathbf{s}_0 provides spatial identification (latitude and longitude coordinates), t_0 is a time index and $Z(\mathbf{s}_0; t_0)$ is a realization of some spatio-temporal stochastic process (i.e. observations of a given variable, located in space and time). The null subscript in \mathbf{s}_0 and t_0 identifies “origin” which is arbitrary and does not affect STSV value (common simplifying assumptions apply). \mathbf{s} and t are used to describe spatial and time distances from the “origin”. Overall, STSVs reflect variability in the data, given spatial and temporal distances among

observations: $\gamma(\mathbf{s}; t)$ is nonnegative, its value generally increases with distance (spatial and temporal) and $\gamma(\mathbf{0}; 0) = 0$. The $\frac{1}{2}$ scaling factor in (3) is a historically established technicality. For empirical analysis, observed data are grouped into convenient intervals along both spatial and temporal distance-axes and evaluated accordingly – see e.g. Journel and Huijbregts (1978).

We may observe various important spatio-temporal properties from Figure 2, which is an empirical STSV for the dependent variable in equation (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 (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 $\gamma(\mathbf{s}; t)$ increases quite rapidly with spatial distance between observations: data are more similar to each other (less varied) in closer regions as compared against observations made farther apart in space. Figure 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 (1). Please note that surface irregularities of the empirical STSV in Figure 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.

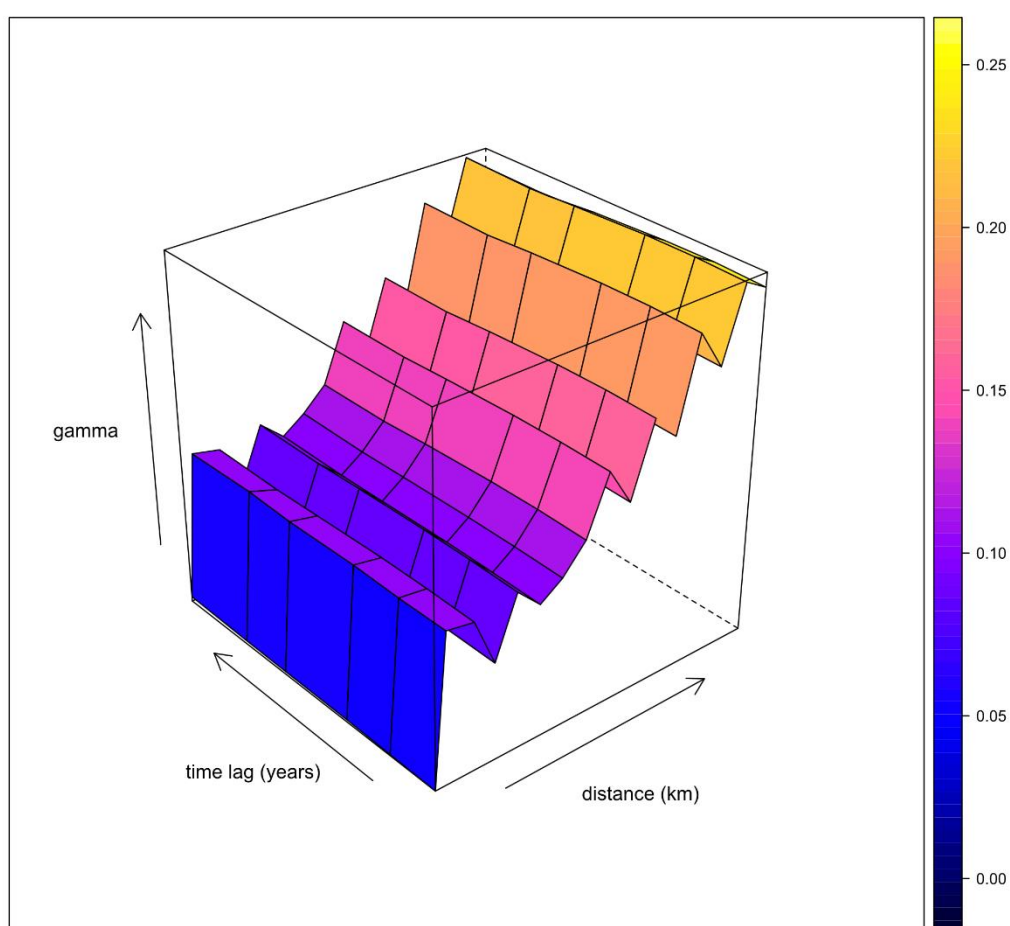


Figure 2. Spatio-temporal semivariogram of log(GDP per capita), with lags 0 to 6 years on the time axis and distances 0 to 500 km on the spatial axis.

Source: Authors' calculation based on data from Eurostat (2018b).

4. EMPIRICAL RESULTS AND THEIR ROBUSTNESS

This section provides estimation results along with interpretation and stability evaluation of a relatively compact spatial panel model, based on regressors described by equation (2).

4.1. Model estimation and interpretation

For an intuitive percentage change interpretation of the estimated coefficients, the dependent variable is log-transformed. A potential drawback of using this transformation lies in the complicated prediction of original variables – model (1) 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 1 provides coefficient estimates for three alternative model specifications – two spatial panel models with different τ values and one pooled-panel & non-spatial reference model. In the first column (a), results from the “best” specification are presented – model evaluation is performed by means of the maximized log-likelihood statistics, based on observed data and regressors as per equations (1) and (2). Middle column (b) contains estimates obtained from an alternative W_N 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 – this approach is methodologically similar to the model presented by Bassanini and Scarpetta (2001). To keep this contribution compact and to avoid printing output with marginal relevance, Table 1 only features the estimated coefficients β_1 to β_7 along with the spatial autocorrelation coefficient ρ , which are deemed relevant for this article, i.e. for analyzing the dynamics of macroeconomic growth. Hence, the intercept and θ coefficients (corresponding to dummy variables controlling state-level heterogeneities) are omitted. Nevertheless, Table 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). All the estimation outputs left-out from Table 1 are available from the author upon request, along with corresponding scripts (R language), additional tests, plots and data. Please note that given the ML estimation of model (1), the usual R^2 is not applicable. 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 1, although the distinction is not relevant for column (c).

The estimated ρ coefficients in columns (a) and (b) of Table 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 β_j coefficients in spatial models. Empirically, high ρ 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 policymaking. In column (a) of Table 1, the coefficient $\beta_1 = 0.2262$ 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 % (approx.). This regressor ($ActShFY15-64_{it}$) should not be viewed as a lever for direct macroeconomic policy. Rather it serves as a control variable addressing labor market development, overall regional competitiveness/effectiveness and welfare status in general (Scharpf & Schmidt, 2000). Similarly, for $\beta_2 = -1.2748$, 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).

Table 1

Output from alternative model specifications: coefficients and statistical inference

	Spatial panel model, $\tau =$ 288 km (a)	Spatial panel model, $\tau =$ 177 km (b)	Polled non-spatial model (OLS) (c)
ρ	0.7724 *** (0.0395) [0.0000]	0.6346 *** (0.0381) [0.0000]	-
<i>ActShFY15-64</i>	0.2262 * (0.0989) [0.0222]	0.3456 *** (0.1001) [0.0006]	0.1886 . (0.1134) [0.0967]
<i>Unem</i>	-1.2748 *** (0.1086) [0.0000]	-1.3931 *** (0.1081) [0.0000]	-1.7222 *** (0.1310) [0.0000]
<i>logR&D_{t-1}</i>	0.0174 ** (0.0059) [0.0030]	0.0235 ** (0.0057) [0.0000]	0.0202 *** (0.0038) [0.0000]
<i>lMWkmsq</i>	0.00001 (0.0002) [0.9455]	0.0001 (0.0002) [0.5452]	0.0002 (0.0001) [0.2130]
<i>Y09GDP_{pc}</i>	0.0328 *** (0.0015) [0.0000]	0.0322 *** (0.0014) [0.0000]	0.0457 *** (0.0014) [0.0000]
<i>Rel_Emp_M&N</i>	1.0502 *** (0.3089) [0.0006]	1.1723 *** (0.3011) [0.0001]	6.5554 *** (0.3427) [0.0000]
<i>(Y09GDP_{pc} × Rel_Emp_M&N)</i>	-0.0399 *** (0.0078) [0.0000]	-0.0395 *** (0.0076) [0.0000]	-0.1771 *** (0.0128) [0.0000]
Pseudo R^2	0.9815	0.9833	0.9913
Log-likelihood	1542.382	1538.925	1129.148

Estimated coefficients are accompanied by (standard errors) and [p-values].

Significance codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1. *Source:* Authors' calculations.

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 τ 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 (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(GDP_{pcit})* 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. This itself is informative: different things may happen under different settings.

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 2 shows that the dependent variable of equation (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 β_j coefficients in the model (by excluding the base GDP level, estimated coefficients of other regressors are roughly doubled in all columns of Table 1). The coefficient for $RelEmpM\&N_{it}$ 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 other supporting-services 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 ($Y09GDPpc_i \times RelEmpM\&N_{it}$), 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. It may be argued that the negative sign in the interaction element points in the direction of Solow-Swan convergence. However, the short-term data used for estimation provide only a mild supporting evidence for such inherently long-term convergence process.

The estimates in column (c) 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-based coefficients are farther from zero).

Finally – in terms of technical description of the estimated model – column (c) uses heteroskedasticity and autocorrelation robust Newey-West standard errors (see Wooldridge, 2010) and the standard errors in spatial models (a) and (b) have asymptotic validity as described e.g. by Millo and Piras (2012). Variance inflation factor (VIF) analysis does not indicate significant multicollinearity problems in the model: VIF evaluation is relevant for model in column (c) only. For estimation of models based on short panel data (condition $N \gg T$ reasonably holds for $N = 113$ and $T = 7$), the potential non-stationarity and/or distributional heterogeneity across time can be ignored without a negative effect on asymptotic properties (Wooldridge, 2010). The rather high values of $Pseudo R^2$ coefficients in Table 1 should be viewed cautiously, as they are mostly due to the presence of base (i.e. lagged) value of the dependent variable ($Y09GDPpc_i$) in the model.

4.2. Model specification robustness: stability of results

Given the diverse options available for specification of the spatial weights matrix W_N in model (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 (2), multiple estimations of the spatial panel model (1) were performed based on observed panel dataset while varying W_N . Figure 3 provides a concise robustness evaluation summary; the information provided therein can be described as follows: The estimation starts with a sparse W_N matrix constructed using $\tau = 160$ km (lower i.e. more restrictive τ thresholds would generate unconnected regions 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 W_N matrix is

generated and the model is estimated. Overall, 341 alternative spatial structures and corresponding coefficient estimates are plotted in Figure 3.

The relative instability of estimates at the lower end of the τ 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: W_N “prohibits” interactions among relatively close regions where spillovers and interactions are actually taking place. Similarly, very large distance thresholds (around 500 km or more) are not empirically justified either. Beyond 500 km (and perhaps even beyond 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” W_N 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 3. Hence, the dotted lines mark coefficient values in the (a) column of Table 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 3 (this result is not very robust against small changes in τ and W_N definition).

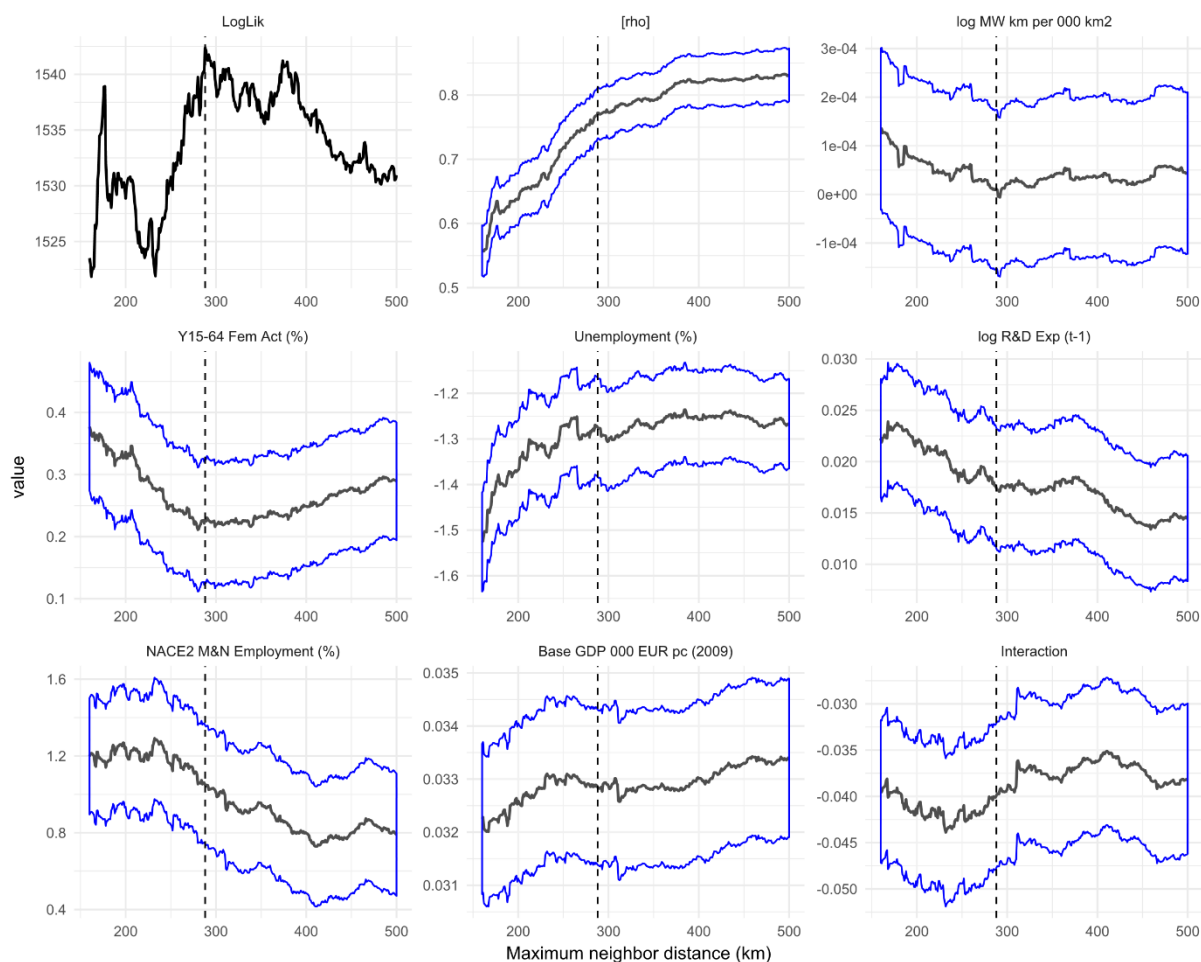


Figure 3. Stability analysis of the estimated spatial error model.

Source: Authors' calculation based on data from Eurostat (2018b).

From the log-likelihood values shown in Figure 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 τ 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 ρ 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 ρ 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.

5. CONCLUSION

The aim of this paper 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 level. 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 labor force structure, 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 controlling for 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 and possible economic policy tools aimed at boosting economic growth. Among such tools, a prominent instrument would be oriented at encouraging R&D expenditures: higher R&D activities are associated with higher GDP growth rates, even after differences in underlying regional and state-level economic environments are considered. At the state-level, additional tax incentives (transparent and easily attainable) may be employed as macroeconomic tools for R&D promotion. At the regional level, subsidies or grants directed to new and/or existing R&D-oriented startup centers may favorably affect R&D location decisions at the micro level.

Promoting the increased share of professional, scientific, organizational and similar workers within the labor force structure was identified as another important factor that central authorities may use to bolster GDP growth. In fact, most of the general policies used for enhancing R&D activities may be applied to influencing (improving) the structure of labor force towards sectors with higher skills, productivity and GDP growth potential. Besides tax incentives and grants, central authorities should support requalification programs providing specialized training/education that would be consistent with the desired changes in 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. This finding should not be viewed as diminishing the importance of transportation (and communication) in general: this type of infrastructure surely is vital and serves as economy’s backbone that enables most economic activities. However, there is no evidence of a GDP growth boosting effect in the short and mid-term horizon, that may be associated

with increased highway density (*ceteris paribus*, once individual and state-level factors and macroeconomic covariates are accounted for).

In addition to coefficient estimation, this paper also provides model stability evaluation that is used to assess the robustness of estimated coefficients with respect to varying definitions of spatial interactions. Empirical evidence gathered in this paper supports specification robustness for the model used and its strong potential towards applications in similar fields of macroeconomic research.

ACKNOWLEDGEMENT

This article was supported by grant No. IGA F4/58/2017 "Modern methods in statistical and optimization models under uncertainty and indeterminacy" of the Faculty of Informatics and Statistics, University of Economics, Prague.

REFERENCES

- Anselin, L., Bera, A.K., Florax, R., & Yoon, M.J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1), 77–104.
- Arvanitidis, P., Petrakos, G., & Pavleas, S. (2007). Determinants of economic growth: the experts' view. *Congress of European Regional Science Association*. Paris.
- Bassanini A., & Scarpetta S. (2001). *The driving forces of economic growth: panel data evidence for the OECD Countries*. OECD Economic Studies No. 33
- Beck, T. (2008). *The Econometrics of Finance and Growth*. Policy Research Working Paper No. 4608. World Bank, Washington, DC.
- Bednář, P., & Halásková, M. (2018). Innovation performance and R&D expenditures in Western European regions: Divergence or convergence? *Journal of International Studies*, 11(1), 210-224. doi: 10.14254/2071-8330.2018/11-1/16
- Choudhry, M. T. (2009). *Determinants of Labor Productivity: An Empirical Investigation of Productivity Divergence*. Working paper, University of Groningen. Retrieved from: <http://encuentros.alde.es/antiores/xiieca/trabajos/pdf/104.pdf>
- Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. Springer, New York.
- Eurostat (2018a). Regional Innovation Monitor. Retrieved from: <https://ec.europa.eu/growth/tools-databases/regional-innovation-monitor/base-profile/province-groningen>
- Eurostat (2018b). Statistics database. Retrieved from: <http://ec.europa.eu/eurostat/data/database>.
- Eurostat (2008). NACE Rev. 2. Statistical classification of economic activities in the European Community. Retrieved from: <http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF>.
- Eurostat - GISCO (2018). Geographic Information System of the COMmission. Retrieved from: <http://ec.europa.eu/eurostat/web/gisco/>.
- Formánek T. (2018) Spatially Augmented Analysis of Macroeconomic Convergence with Application to the Czech Republic and Its Neighbors. In: Silhavy R., Silhavy P., & Prokopova Z. (eds) *Applied Computational Intelligence and Mathematical Methods. CoMeSySo 2017. Advances in Intelligent Systems and Computing, vol 662*. Springer, Cham. doi: 10.1007/978-3-319-67621-0_1
- Fujita, M., Krugman, P., & Venables, A. (1999). *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge.
- Gauselmann, A., Marek, P., & Angenendt, J. P. (2011). *MNE's Regional Location Choice - A Comparative Perspective on East Germany, the Czech Republic and Poland*. IWH Discussion Paper No. 8.
- Hamilton J. D., & Owyang, M. T. (2012). The Propagation of Regional Recessions. *The Review of Economics and Statistics*, 94(4), 935-947. doi: 10.1162/REST_a_00197
- Journel, A. G., & Huijbregts, C. J. (1978). *Mining Geostatistics*. Academic Press, London.
- Jutting, J. (2003). *Institutions and Development: A Critical Review*. OECD Development Center, Working Paper 210.

- Kapoor, M., Kelejian, H. H. & Prucha, I. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1), 97–130. doi: 10.1016/j.jeconom.2006.09.004
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Press, Taylor & Francis Group, Boca Raton.
- Ma, C. (2005). Spatio-temporal variograms and covariance models. *Advances in Applied Probability* 37(3), 706–725. doi: 10.1239/aap/1127483743
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2), 407–437. doi: 10.2307/2118477
- Millo, G., & Piras, G. (2012). splm: Spatial Panel Data Models in R. *Journal of Statistical Software*, 47(1), 1–38. doi: 10.18637/jss.v047.i01
- Myrdal, G. (1957). *Economic theory and underdeveloped regions*. Hutchinson, London.
- Nordhaus, W. D. (2017). *Evolution of modeling of the economics of global warming: Changes in the dice model, 1992-2017*. Cowles Foundation Discussion Paper No. 2084, Cowles Foundation for Research in Economics, Yale University.
- Nordhaus, W. D. (1992). An optimal transition path for controlling greenhouse gases. *Science* 258(5086), 1315–1319. doi: 10.1126/science.258.5086.1315.
- Pebesma, E. (2012). spacetime: Spatio-temporal data in R. *Journal of Statistical Software*. 51(7), 1–30. doi: 10.18637/jss.v051.i07
- Piras, G., & Arbia, G. (2007). Convergence in per-capita GDP across EU-NUTS2 regions using panel data models extended to spatial autocorrelation effects. *Statistica*, 67(2), 157–172. doi: 10.6092/issn.1973-2201/3513
- Prokop, J., & Karbowski, A. (2018). R&D spillovers and cartelization of industries with differentiated products. *Journal of International Studies*, 11(3), 44–56. doi: 10.14254/2071-8330.2018/11-3/4
- Romer, P. (1990). Endogenous technological change. *Journal of Political Economy*, XCVIII, 71–102.
- Scharpf, F. W. & Schmidt, V. A. (2000). *Welfare and work in the open economy: Vol II: Diverse responses to common challenges*. Oxford University Press, Oxford.
- Solow, R.M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70, 65–94.
- Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge.