Measuring spatial aspects of variability: comparing spatial autocorrelation with regional decomposition in international unemployment research

Nosek, Vojtech; Netrdova, Pavlina

Empfohlene Zitierung / Suggested Citation:
Nosek, V., & Netrdova, P. (2014). Measuring spatial aspects of variability: comparing spatial autocorrelation with regional decomposition in international unemployment research. *Historical Social Research*, *39*(2), 292-314. https://doi.org/10.12759/hsr.39.2014.2.292-314

Measuring Spatial Aspects of Variability. Comparing Spatial Autocorrelation with Regional Decomposition in International Unemployment Research

Vojtěch Nosek & Pavlína Netrdová *

Abstract: *»Das Messen von räumlicher Variabilität. Der Vergleich von räumlicher Autokorrelation und regionaler Dekomposition in der internationalen Forschung zur Arbeitslosigkeit«.* This paper focuses on spatial aspects of variability and specifically on the relationship between regional decomposition and spatial autocorrelation. These characteristics are often supposed to be interconnected, but the subject has not yet been studied in sufficient detail and spatial methods are often neglected in regional analysis. We start with a brief discussion of a methodology suitable for identifying and quantifying spatial aspects of variability. The key part of the paper focuses on methodological reflections on measuring spatial aspects of variability and the advantages and disadvantages of our chosen methods. We use the Theil index, which is decomposable without residuum, to assess the relative importance of the regional organization of our studied phenomena. To measure spatial autocorrelation, which enables us to quantify the level of spatial concentration of the studied phenomena and reveal spatial clustering, we use Moran's I (global scale) and LISA (local scale). We explain in depth the properties of these methods, advantages/disadvantages, behaviour in different situations and the potential for them to be combined and used jointly. These methodological findings help to better understand and interpret the results of the subsequent empirical research. We apply the methods in international unemployment research with highly detailed data from Austria, Czechia, Germany, and Poland. Specifically, we are interested in the importance of socio-spatial (regional) organization in relation to unemployment rates, and we present noteworthy results concerning the spatial differentiation of unemployment in the Central European region.

Keywords: Regional decomposition, spatial autocorrelation, international research, Central Europe, unemployment, Moran's I, LISA, Theil index.

* Vojtěch Nosek, Charles University in Prague, Faculty of Science, Department of Social Geography and Regional Development, Albertov 6, 128 43, Prague 2, Czech Republic; nosek6@natur.cuni.cz.
Pavlína Netrdová, Charles University in Prague, Faculty of Science, Department of Social Geography and Regional Development, Albertov 6, 128 43, Prague 2, Czech Republic; pavlina.netrdova@natur.cuni.cz.
The authors gratefully acknowledge the financial support from the Czech Science Foundation (GAČR) under Grant No. P404-12-P340

1. Does Space Matter in the Social Sciences?

The importance of space is becoming increasingly recognized when studying socio-economic processes (Goodchild et al. 2000). However, the majority of methods commonly used in social sciences have been applied with little regard for spatially referenced data (Rey and Janikas 2005). In the social sciences, empirical data are often analysed with standard statistical methods (such as measures of dispersion, regression, and factor and cluster analysis), which do not directly reflect spatiality. This is despite the fact that in spatial analysis, which is connected predominantly with increasing the accessibility of spatial data and advances in GIS software, significant progress has been made (de Smith et al. 2013; Fischer and Getis 1997; Fotheringham and Rogerson 2009). Today the focus on developing specific methods for spatial analysis is one of the most significant trends in quantitative geography (Fotherigham et al. 2000).When analysing social data, the inclusion of spatial aspects can lead to innovative results. Thus, one of the main challenges in the social sciences is to identify and measure the concentration of various processes in space (Nosek and Netrdová 2010).

Few authors have focused their research on quantifying the spatial aspects of socio-economic variability. One of the main reasons may be strong multi-causality and the indivisibility of social processes, which makes quantifying spatial aspects problematic. This was defined by Harvey (1973, 40) as "socio-spatial confounding". On the other hand, plenty of methods suitable for quantifying spatial aspects of variability have been developed (Anselin 1995; Fotheringham et al. 2000) and applied in empirical research (see for example Nosek and Netrdová 2010; Novotný 2007; Rey 2001; Rey and Janikas 2005). Moreover, a lot of methodological inspiration can be found in the spatial econometric literature (Anselin 1988; Anselin et al. 2004a), even though its authors are often criticized for focusing more on methodological concepts than on explaining the true nature of the processes.

We acknowledge that quantifying social phenomena is inevitably simplistic to some extent, and this study is no exception. However, we believe that complex systems such as social processes in space can be represented (albeit imperfectly) by relatively simple quantitative models (Hampl et al. 1999). These simplistic quantitative representations can subsequently help to find causal explanations. In this approach, features of critical realism can be traced (Sayer 1984; Yeung 1997).

The paper builds on these challenges and its *main goals* are to:

- Introduce suitable methods of identifying and quantifying spatial aspects of variability;
- Explain in depth their properties, advantages/disadvantages, behaviour in different situations and the potential for their combination and joint utilisation; and

- Demonstrate their application in international unemployment research in a way that capitalizes on a detailed knowledge of their properties.

We use the term "*spatial aspects of variability*" to emphasize the fact that standard variability measures used to quantify differences between various (regional) units do not capture the "spatiality" in its full scope. In order to better understand the social processes and uncover possible causal mechanisms, it is important

- not only to quantify differences between regional units,
- but also to study the distribution of phenomena in space,
- to find and measure spatial clustering or spatial concentration, and
- to identify specific spatial clusters and outliers.

2. Measuring Spatial Aspects of Variability

We approach the quantification of spatial aspects of variability in two conceptually different ways, measuring regional variability (regional decomposition) and spatial autocorrelation. Although not primarily developed for quantifying spatial aspects of variability, they are well suited for this purpose, especially when used together. These two *approaches* can be measured and quantified in a variety of ways. Without loss of generality, we use two specific methods which are common in geography literature:

- the *Theil index,* from a family of generalized entropy indices, and its decomposition (Cowell and Jenkins 1995; Netrdová and Nosek 2009; Shorrocks 1984; Shorrocks and Wan 2005), and
- *Moran's I and LISA statistic* as a global and local spatial autocorrelation measure (Anselin 1988, 1995; Cliff and Ord 1973).

In this section we briefly discuss the methodological background for measuring spatial aspects of variability and we present our chosen approaches and methods.

2.1 Regional Differences: Basic Statistical Measures

The basic statistical methods used to quantify variability are *variance* and *standard variation*. However, these measures are not scale invariant, i.e. results are not independent of the choice of scale. For example, results of income variability (inequality) would be affected by the choice of currency. Variance and standard variation are thus inappropriate for comparing different variables. Scale invariance makes the *coefficient of variation* more attractive, although this does have one crucial weakness: it is strongly dependent on the mean. Unfortunately, this is very important because socio-economic data typically have more or less skewed (i.e. non-normal) distributions (Hampl et al. 1999; Novotný 2004; Novotný and Nosek 2009; Ulubasoglu and Hazari 2004). If we

use the same example and try to compare income inequality in countries with very different income distribution in society (more or less different from the normal distribution where the median equals the mean), the coefficient of variation would not reflect this disparity in distribution appropriately. The relative independence of the mean and rather low sensitivity to extreme values make the *Gini coefficient*, which is based on differences between all pairwise values, very popular in regional science (Cowell 1977; Cowell and Flachaire 2007; Lambert and Aronson 1993).

2.2 Relative Regional Variability: the Theil Index and its Decomposition

The above measures of variability enable us to quantify regional variability. At this point, we should distinguish between regional variability and relative regional variability. While *regional variability* quantifies differences between regional means, the concept of *relative regional variability* quantifies spatial aspects of variability according to the proportion of overall variability, which can be attributed to different regional levels (for example NUTS3 or LAU1 level). It is possible to quantify the relative regional variability by decomposing variability into its between-group (between-region in this case) and within-group (within-region in this case) components. Overall variability can be understood as the sum of between-region and within-region components, and it ideally represents inter-personal inequality, which for practical reasons is substituted by inter-municipality variability. The relative regional variability could be defined as the *share of the between-region component of variability in overall variability*.

This decomposition can be calculated for the Gini coefficient or the *Theil index* from the family of generalized entropy indices, and for the Theil index without residuum (Cowell and Jenkins 1995; Netrdová and Nosek 2009; Novotný 2007; Shorrocks and Wan 2005).

2.3 Spatial (Regional) Patterns: Visualizing Variables in Maps

The most straightforward way to assess *spatial (regional) patterns* is to *visualize variables in a map* (absolute or relative values, index of localization, etc.). Yet when using a very detailed spatial structure, as is the case with municipalities, the final map is often very fragmented and therefore difficult to interpret. On the other hand, when using larger regional units, some local anomalies may remain hidden in the regional means due to aggregation, and some interpretations may be skewed by the *Modifiable Area Unit Problem (MAUP)* (Openshaw 1984; Wong 2009). MAUP can be viewed as a type of ecological fallacy in which results of analysis differ depending on the aggregation of data to different regional units. For example, by using different regional levels (scale effect) or another delimitation of regions (zoning effect) one can obtain very distinctive

results. There is no scientific agreement on the solution to MAUP yet, although one approach is to use highly detailed spatial data.

2.4 Spatial Autocorrelation: Moran's I and LISA

By mapping a spatial pattern visually we can assess spatial aspects of variability according to a *concentration of similar values in space*. This approach is based on the assumption that "everything is related to everything else, but near things are more related than distant things" (Tobler 1970 in Sui 2004, 269). This could be measured by *spatial autocorrelation*, i.e. correlation of a variable with itself in space (Anselin 1988; Cliff and Ord 1973). For instance, if we have data for average incomes in regions, we can study whether the income levels in different regions are more similar in closer (neighbouring) regions. If they are, we can conclude that there is a positive spatial autocorrelation present, which can be mapped in the form of spatial clusters. If the values are randomly distributed in space and no significant spatial clusters are formed, spatial aspects of this variability can be considered unimportant or non-existent. On the other hand, large clusters of values autocorrelated in space imply that the spatial aspects of variability of a studied variable are important.

In the case of spatial autocorrelation, one must distinguish between global and local statistics. In *global statistics*, spatial clustering in the whole studied area is quantified using one value, while *local statistics* can reveal local specifics and spatial clusters and outliers through mapping. To measure global spatial autocorrelation we use the popular *Moran's I*, (Anselin 1988; Cliff and Ord 1973), which has many similarities with Pearson's correlation coefficient. The spatial weighting scheme defines which units are considered geographically close for the calculation of spatial autocorrelation. The choice of spatial weighting scheme is important in a spatial autocorrelation methodology and depends on the spatial structure of the studied area. The values of Moran's I range from +1 (maximum positive spatial autocorrelation) to -1 (maximum negative spatial autocorrelation). Values close to 0[1] indicate a random pattern (Fotheringham et al. 2000). To reveal local spatial autocorrelation and identify spatial clusters we can use *LISA analysis (local indicator of spatial association)*, the local equivalent of Moran's I, (Anselin 1995). The results of LISA analysis in the form of a cluster map (i.e. a map with locally specific values which may form visual clusters) and a significance map answer important questions, such as where the clusters can be found, what they look like, and whether they are random or statistically significant.

[1] More precisely, if the value of Moran's I is close to the expected value $I = -1/(n - 1)$. However, by analysing large datasets the practical error is insignificant.

3. Theoretical and Methodological Properties of the Selected Methods

Both the concepts of regional variability decomposition and spatial autocorrelation, although methodologically different, can lead to similar and complementary results (Netrdová and Nosek 2009; Rey 2001). *Spatial autocorrelation (Moran's I and LISA)* helps to find spatial patterns independently of the administrative definition of regions, while the *Theil index decomposition* quantifies the relative significance of predefined regions. By applying both methods, we can approach spatial aspects of variability in a more comprehensive way and this enables us to come to innovative interpretations. Since we are using several specific terms, a short list of the most important ones, including a brief explanation and the methods used, is presented in Figure 1.

Figure 1: Terminology and Selected Methods

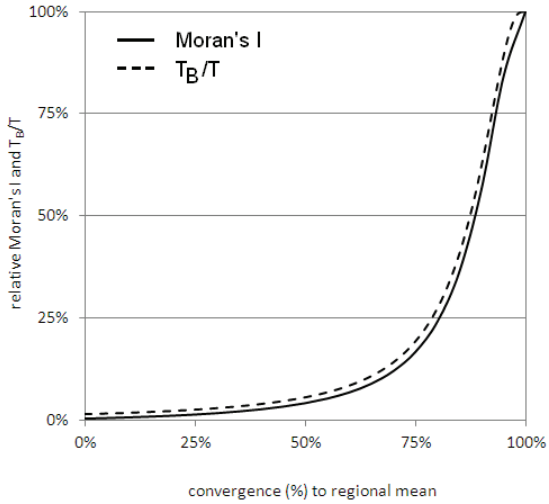
Term	Explanation
Overall variability	Variability measures between sub-regional units (the smallest possible; ideally between individuals). - Theil index (T)
Regional variability	
(i) Simple regional variability	Variability between regional units, which enables us to identify the differences in phenomena between regions. However, this may still be a relative form of variability when the data are weighted (normalized) by various regional population characteristics. - the between-group component of Theil index (T_B)
(ii) Relative regional variability	The ratio of variability between regional means to the overall variability, which enables us to assess the relative importance of the spatial organization of a phenomenon. - the share of the between-group component in overall variability (T_B/T)
Spatial autocorrelation	
(i) Global spatial autocorrelation	Spatial autocorrelation expressed by a single value for the entire system studied, which enables us to measure the extent of spatial clustering of various phenomena in space. - Moran's I
(ii) Local spatial autocorrelation	Spatial autocorrelation mapped for each geographical unit under analysis in order to uncover various local specificities of the studied system and identify spatial clusters and outliers. - LISA analysis

We now focus on general theoretical and methodological properties of the selected methods, with special attention to the quantification of spatial aspects of variability. First, we describe the methodological properties of the selected methods in the theoretical models (through theoretical simulations). Second, we use the Central European region and unemployment data as an example to show how these methods "behave" and can be applied in empirical research.

3.1 Nonlinearity

The fundamental part of any empirical study is the final interpretation. To interpret results correctly, one has to understand the methods which were applied. Essential to methods measuring (not only) spatial aspects of variability is the course of their functions, which indicates how to interpret the numeric results of computed coefficients and their changes over time. Figure 2 captures the course of the functions of values of Moran's I and Theil index decomposition (T_B/T) with regard to their "meaning", which in this case was represented by convergence to regional means in a modelled log-normal distribution.

Figure 2: Nonlinearity of Theil index decomposition and Moran's I



Note: For this demonstration the regional structure of Czechia was used and data with modelled log-normal distribution in municipalities (LAU2) were gradually changed to their regional means (LAU1). One horizontal point on the curve represents 1% convergence to the municipality's regional mean. Moran's I is in relative form (maximal value of Moran's I in the case of 100% convergence is 100%) in order to match the scale of the Theil index.

Figure 2 confirms that both methods behave non-linearly. Changes in equal percentage values of T_B/T and Moran's I (on the vertical axis) mean different changes in the similarity of real values in regions (on the horizontal axis). For example a change of T_B/T from 5% to 15% is of much greater magnitude (representing approximately 24% convergence to regional means) than a change from 65% to 75% (representing convergence to regional means of approximately 1%). An incorrect interpretation of Moran's I would be, for example, to regard a value of 0.4 as twice the spatial clustering of a Moran's I of 0.2. Even small changes in the studied indices can mean a significant change in convergence to regional means, and vice versa. This observation is especially im-

portant when studying the evolution of spatial aspects of variability over time and interpreting the changes in the computed indices.

3.2 Statistical Inference

Another often underestimated yet very important technique when measuring spatial aspects of variability is statistical inference. This is typically used to infer properties of the population based on the characteristics of a sample. Since we often have data for the whole population, statistical inference can also be used to find the probability that the measured results are random. In the context of spatial analysis, we can compare the observed results with a situation where the data are distributed randomly. If we find the results statistically significant, it basically means that the observed results are not random (where random implies "no ontological meaning") but affected by some contextual factors. Though not new, this kind of test is still rather rare in regional variability research (Mills and Zandvakili 1997; Stine 1989). Basic asymptotic tests are not suitable in this case; one has to use non-parametric ones, which are usually based on re-sampling. The confidence interval is constructed from the simulated values of the tested characteristics, which are calculated from data repeatedly generated from the original data set. These simulated values are called the null model.[2] It is inevitable that even in the null model some regional variability will be found. Therefore, regional variability can be understood as the sum of two components:

1) the stochastic component (regional variability of the null model); and
2) the contextual component (regional variability exceeding the null model, i.e. measured regional variability minus the regional variability of the null model).

Isolating the contextual component of regional variability is useful especially when comparing different geographical systems (such as countries) since each system has different stochastic variability embedded in the results. This technique is described in detail in Novotný and Nosek (2012).

The inference in the case of spatial autocorrelation is based on the same principle. To assess the significance of Moran's I against a null hypothesis (no spatial autocorrelation), we use a permutation procedure, specifically the conditional permutation procedure embedded in the GeoDa 1.4.0 software. We use 9,999 permutations, which in most cases are sufficient to obtain stable results (Anselin 2003).

[2] In this paper we use 1,000 permutations to calculate the null model. Visual Basic scripts in MS Excel were used to perform the permutations.

3.3 Mutual Interaction of Spatial Autocorrelation and Regional Decomposition

When using the Theil index and its decomposition, Moran's I and LISA analysis jointly in empirical analysis, it is worthwhile to consider their mutual interaction. An understanding of their relationship may help in interpreting the results. This relationship was studied through theoretical simulations using model data: 10,000 log-normally distributed pseudo-random data.[3] These data were randomly distributed in a regular 100x100 grid. A regular grid was used in order to minimize the bias caused by different regional delimitations. To measure the spatial autocorrelation, a spatial weighting scheme queen 1st order has been chosen.

At this point we distinguish between overall variability (T), regional variability (T_B) and relative regional variability (T_B/T); see the terminology above (Figure 1). There is *no statistical relationship between overall variability and spatial autocorrelation* even though a contextual relationship is supported by empirical data (Rey 2001). This can be demonstrated by a rather simple model exercise where all values in a studied area are modified in the same way, but their spatial arrangement remains the same.

Analysis of the relationship between *simple regional variability and spatial autocorrelation* through random and specific arrangements of values in a regular grid produces a trend, albeit a rather weak one. With increasing regional variability (T_B), the values of Moran's I increase in a majority of cases. However, this relationship differs significantly with different values of overall variability (T).

Figure 3: General Typology of Phenomena Due to the Relationship between Relative Regional Variability and Global Spatial Autocorrelation

	Relative Regional Variability HIGH	Relative Regional Variability LOW
Spatial Autocorrelation HIGH	SPATIALLY dependent and bounded in REGIONS (concentrations in regions)	SPATIALLY dependent with no relation to REGIONS (concentrations across regional borders)
Spatial Autocorrelation LOW		Both SPATIALLY and REGIONALLY independent (no concentrations)

Source: Nosek and Netrdová 2010 (modified).

The relationship between *relative regional variability (T_B/T) and spatial autocorrelation* is slightly more straightforward, but still rather complex. The relationship can be categorized into several types, which are presented in Figure 3. When a high spatial autocorrelation is observed, both high and low values of

[3] Log-normal distribution is often considered to represent socio-geographical data the most accurately (Novotný and Nosek 2009).

regional variability can be present. On the other hand, with very low spatial autocorrelation, it is theoretically impossible to have low regional variability.

In several ways this typology can help in the interpretation of empirical results. For instance, the low values of relative regional variability do not necessarily imply that there are no spatial aspects in the studied phenomenon. This result might be caused by inappropriate regional delimitation. Spatial autocorrelation would thus provide significant added value in this case. In general, if the values of relative regional variability and spatial autocorrelation differ significantly, it is obvious that the regional delimitation (definition of regions) is not suitable or not detailed enough for the phenomenon under analysis. Empirical demonstrations of the three defined types of phenomena according to their spatial aspects of variability are documented using Czechia as an example in Nosek and Netrdová (2010).

4. Testing the Methods Using Empirical Data from International Unemployment Research

After presenting and discussing some important issues concerning the methodological characteristics of the selected methods, including their mutual relationship, we test the methodology on empirical data – the unemployment rate in four countries in the Central European region (CER). The main goal of the empirical research is to demonstrate a suitable strategy for analysing spatial aspects of variability, and we also discuss limitations of the data and methods common in this kind of analysis.

4.1 Data and Regional Structure

We are using empirical data on four Central European countries (Austria, Czechia, Germany, and Poland), which are studied in the regional structure of European Union statistical units – NUTS (*Nomenclature of Units for Territorial Statistics)* and LAU (*Local Administrative Units)*. There is one important prerequisite for this type of analysis – spatially highly detailed data – therefore, for the current study, we use municipalities (LAU2 units) as the basic regional structure. The comparability of regions in the respective countries is problematic despite the fact that we are using standard EU regional units. This is obvious from the average area of the respective regions in the four countries presented in Table 1.

For example, NUTS3 in Czechia and in Poland are much larger than NUTS3 in Austria and Germany. For this reason, the NUTS3 regional level in Czechia and Poland was substituted by the LAU1 level in the subsequent analyses. Moreover, the administrative regions often do not represent the real functional organization of social (socio-economic) processes in space and results can thus

be misleading. To overcome this *spatial mismatch*, *functional delimitation of regions* can be adopted.

Table 1: Number and Average Area Size of Administrative Units

Country	LAU2	LAU1	NUTS3	NUTS2	NUTS1
Austria	2379 35 km^2	-	35 2401 km^2	9 9336 km^2	3 28009 km^2
Czechia	6251 12 km^2	77 1024 km^2	14 5633 km^2	8 9858 km^2	1 78867 km^2
Germany	11516 31 km^2	4625 75 km^2	412 869 km^2	33 10855 km^2	16 22389 km^2
Poland	2478 126 km^2	379 824 km^2	66 4743 km^2	16 19567 km^2	6 52178 km^2

Note: data as of 1st January 2011; examples of administrative units in Germany: NUTS1 = states (Bavaria), NUTS2 = government regions (Middle Franconia), NUTS3 = districts (Weißenburg-Gunzenhausen), LAU1 = collective municipalities (Hahnenkamm), LAU2 = municipalities (Westheim).

However, this regionalization is rather complicated and specific detailed data are needed (see for example Hampl et al. 1999). It is less accurate yet much easier to adjust existing administrative regions in the most obvious cases, usually by *merging regions with complementary socio-economic functions*. Merging Prague with its hinterland (together forming one functional region) would be a perfect example. Yet since our main goal is methodological, we have not proceeded with these adjustments, with the exception of using LAU1 regional level in Czechia and Poland in the NUTS3 analyses. This regional level, as the most detailed one possible, has been chosen for the analyses of relative regional variability.

Table 2: Basic Characteristics of the Unemployment Rate (UR) in 2010

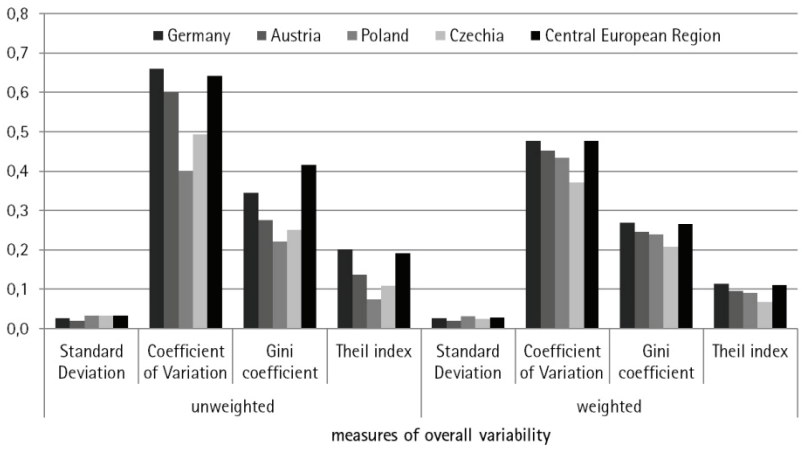
Characteristic	Austria	Czechia	Germany	Poland	CER
Number of registered unemployed people	244 923	495 160	2 941 146	1 942 756	5 623 985
Unemployment rate (UR)	4.32%	6.71%	5.45%	7.10%	5.96%
Number of LAU2 regions with data about the unemployment rate (UR)	2 379	6 250	11 222	2 478	22 329
Mean of UR in LAU2 units	3.41%	6.73%	4.00%	8.28%	5.18%
Median of UR in LAU2 units	2.98%	6.29%	3.32%	7.88%	4.42%
Range of UR in LAU2 units	25.78%	100%	23.93%	22.85%	100%

Note: The unemployment rate = number of registered unemployed people in October 2010 (from official statistical sources in studied countries) normalized by the working-age population (15-64) on 31st December 2010.
Source: Statistik Austria (<http://www.statistik.at>), Czech Ministry of Labour and Social Affairs (<http://www.mpsv.cz>), Czech Statistical Office (<http://www.czso.cz>), Statistik der Bundesagentur für Arbeit (<http://statistik.arbeitsagentur.de>), Statistische Ämter des Bundes und der Länder (https://www.regionalstatistik.de/genesis/online/data>), Central Statistical Office of Poland (<http://www.stat.gov.pl>).

The data present another issue. When undertaking international research, data collection and adjustment necessary for comparability are very important. Based on the expected strong spatial aspects of its variability (Nosek and Netrdová 2010), we have chosen the *unemployment rate* in 2010. In addition, the administrative regional delimitation should correspond very well with its functional regions. The unemployment rate is usually defined as the number of unemployed normalized by the sum of the economically active population. Due to data accessibility in all four countries, we have normalized the number of registered unemployed by the working-age population (15-64). The unemployment rate values are therefore lower compared to other studies. All data document the unemployment situation in October 2010. The highest rates are in Poland and Czechia, the lowest in Austria and Germany (see Table 2).

4.2 Regional Variability

Figure 4: Measures of Overall Unemployment Rate Variability in 2010 in LAU2 Units



Note: Measures of overall variability are weighted by the working-age population (15-64).
Source: see Table 2.

The variability measures of the unemployment rate are very different across the studied countries (see Figure 4). Despite being rather well known, the difference between *weighted and un-weighted measures of variability* needs to be stressed. Un-weighted measures do not take into account the population in the respective regions and all regions are treated equally. In other words, each region has the same weight no matter how big it is. This is important especially if there is large variance between populations in the studied units. Compare for example the values of weighted and un-weighted measures in Czechia with the

same values in other countries. The variability is highest in Germany, probably due to the still visible (socio-)economic polarity between the western and eastern parts of the country. On the contrary, the lowest variability can be observed in Czechia and Poland.

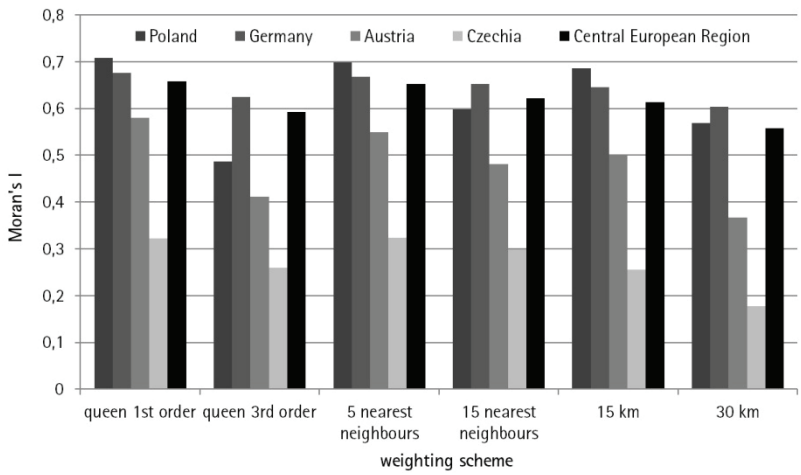
4.3 Choice of Spatial Weighting Scheme

As described above, one of the most suitable ways of measuring and visualizing the clustering of a socio-economic phenomenon is *spatial autocorrelation*. One methodological issue associated with this concept is the choice of spatial weighting scheme, which operationalizes the position of geographical units and is an important element of all spatial autocorrelation measures; Moran's I and LISA analysis are no exception to this (Anselin 1988; Cliff and Ord 1973; Getis and Aldstadt 2004). Since the choice of spatial weighting scheme is to a large extent arbitrary and the influence on the final results could be significant, we calculate Moran's I for several different weighting schemes. To demonstrate, we use contiguity weights (rook 1st and 3rd order) and distance weights based on the x- and y-coordinates of municipalities (5 and 15 nearest neighbours, the distance threshold with a cut-off of 15 and 30 km).

Although the selection of a particular spatial weighting scheme is subjective and often considered crucial (Unwin and Unwin 1998), the results in Figure 5 indicate the contrary. In relative terms, the results are very similar. However, there are a few exceptions. The spatial weighting scheme is, for example, influenced by the *number of neighbours* that each unit has. This strongly depends on the regional structure and areas of the units under analysis. The more neighbours a unit has on average, the higher Moran's I is. According to Table 3, however, the selection of a distance-based spatial weight matrix does not alter the results (in relative terms) even though the number of regions varies significantly. The results confirm the intuitive prediction that the value of Moran's I decreases as the neighbourhood defined within the spatial weighting scheme grows. Nevertheless, when comparing different regional systems, the utilization of a spatial weighting scheme based on queen contiguity (with small differences in the average number of neighbours) seems to be more suitable.

In general, the choice of spatial weighting scheme can alter the final results in absolute terms, but it rarely influences its interpretation (the cluster can be found in the same areas, but their sizes vary). When comparing several variables in the same regional structure the choice of spatial weighting scheme should not alter the outcomes (see Spurná 2008). However, when comparing different regional systems (different countries), non-distance-based spatial weight matrices (type queen or rook) should be preferred. Based on these outcomes, we used spatial weights based on queen contiguity (2nd order) in the empirical analyses.

Figure 5: Spatial Autocorrelation of the Unemployment Rate in 2010 Measured by Moran's I, Different Weighting Schemes



Source: see Table 2.

Table 3: Average Number of Neighbours for Different Weighting Schemes

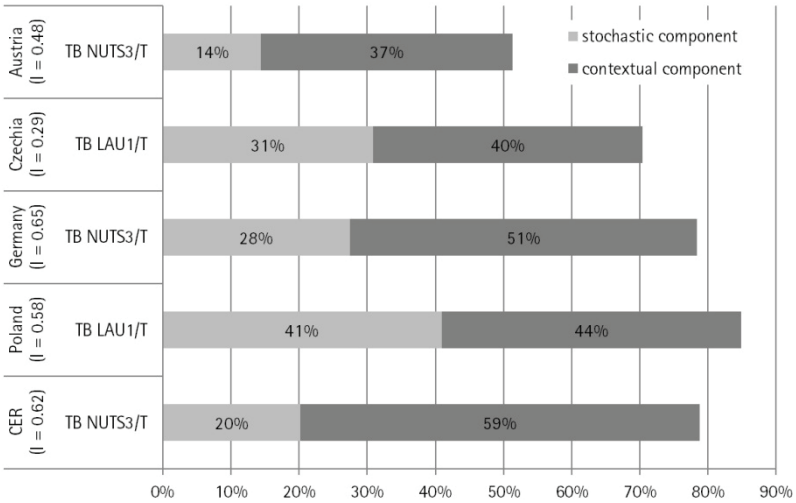
Weighting Scheme		Austria	Czechia	Germany	Poland	CER
Queen Contiguity	1st order	5.80	5.90	5.95	5.68	5.89
	2nd order	19.33	21.88	21.29	18.80	20.98
	3rd order	41.36	50.11	47.95	39.87	46.99
Rook Contiguity	1st order	5.67	5.86	5.86	5.58	5.81
Threshold Distance	15 km	24.31	63.64	36.74	5.39	39.89
	20 km	41.78	110.61	62.97	9.90	69.10
	25 km	62.81	169.47	94.99	15.75	105.36
	30 km	87.10	239.65	132.42	22.80	148.38

4.4 Regional Decomposition and Global Spatial Autocorrelation

The results of T_B/T and Moran's I, which represent relative regional variability and global spatial autocorrelation of the unemployment rate in the studied countries, are presented in Figure 6. For all four countries the values of Moran's I are significant at the 1% level, documenting a strong positive spatial autocorrelation. Due to the different regional structure and number of LAU2 units in the countries, the values of Moran's I are not directly comparable and the differences between countries should not be interpreted. The un-adjusted value of T_B/T for NUTS3 level (LAU1 in Czechia and Poland) is the highest in Poland (88%), followed by Germany (79%) and Czechia (71%); with the lowest values by far observed in Austria (51%). However, if we apply the adjustment mentioned above, the results are significantly different. Now the highest value

(contextual value) is Germany (51%), followed by Poland (44%), Czechia (40%), and Austria (37%) – all three with rather similar values. The change in the results is cause by a strong stochastic component of variability in Poland and a weak stochastic component of variability in Austria. This exercise proves the importance of statistical significance and distinguishing between the stochastic and contextual components of variability in relative regional variability research.

Figure 6: Moran's I and T_B/T for the Unemployment Rate in 2010



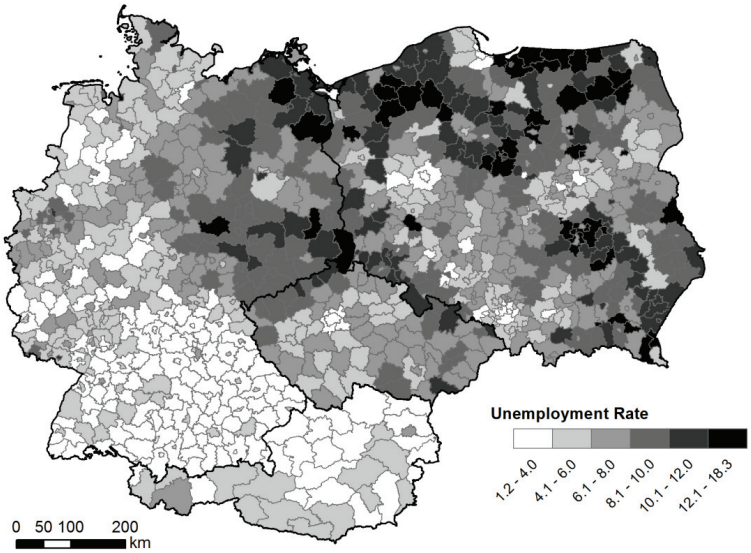
Notes: In calculating Moran's I, spatial weights based on queen contiguity (2nd order of contiguity) were used. All values (both Moran's I and Theil index decomposition) are statistically significant at the 1% level.
Source: see Table 2.

4.5 Visualization and Local Spatial Autocorrelation

So far, we have shown the results of quantifying regional and spatial differentiation of the unemployment rate through values of T_B/T (relative regional variability) and Moran's I (spatial autocorrelation). These methods are suitable for identifying regional levels to which the majority of the observed variability can be attributed, and for studying the intensity of spatial clustering. When combined, these methods are well suited to studying spatial aspects of a phenomenon under analysis. However, they are global measures and thus we cannot use them to answer some important questions such as: *What is the nature of spatial clustering? Can we identify axes, nodes, areas of peripheries? In what localities does statistically significant clustering occur?* Answering these questions is the first step for contextual understanding of the studied processes, which is more important than mere quantifications.

The most obvious local representation of variability is the *visualization of values on a map*. Graphical representations have many advantages. First of all, they are usually easy to construct and they are useful for uncovering the basic regional (local) patterns of the studied phenomenon. In many cases this may be sufficient. Empirical examples from the studied region can be found in Figure 7. The map depicts the unemployment rate at the NUTS3 regional level of detail and reveals a basic spatial pattern – a significant difference between the western (former West Germany and Austria) and eastern part of the region. On the other hand, this simple visualization has several disadvantages. The NUTS3 regional level does not allow the study of local (sub-regional) differences which may be crucial for analysing spatial patterns. When visualising data at the LAU2 regional level, the map would be too fragmented and difficult to interpret. In addition, the categories to scale the variable under analysis are often set arbitrarily or on a linear basis (such as quantiles) and may produce significant bias.

Figure 7: The Unemployment Rate in NUTS3 regions (LAU1 in Czechia and Poland) in 2010

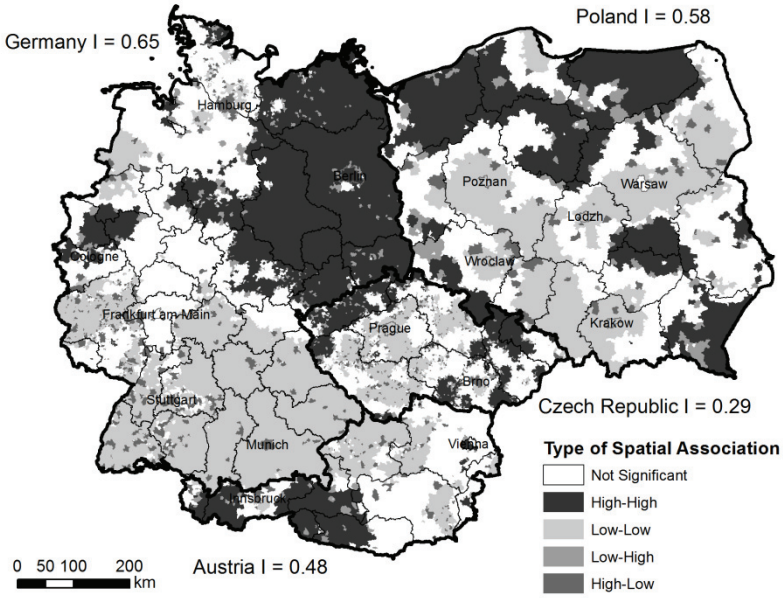


Source: see Table 2.

The best way to support graphical analysis is to use local statistics of spatial autocorrelation (for example LISA), which enables us to identify and test spatial clusters. Compared to simple visualization and global statistics, local statistics have several advantages. They eliminate problems of analysing spatially

aggregated data and help to discover deviations from global statistics. Thus they help to better understand and interpret spatial processes (Fotheringham 1997; Unwin and Unwin 1998). An empirical demonstration from the Central European region is presented in Figures 8 and 9.

Figure 8: LISA Cluster Maps for the Unemployment Rate in 2010 in the Studied Countries (Separately), Weighting Scheme Queen 2nd Order

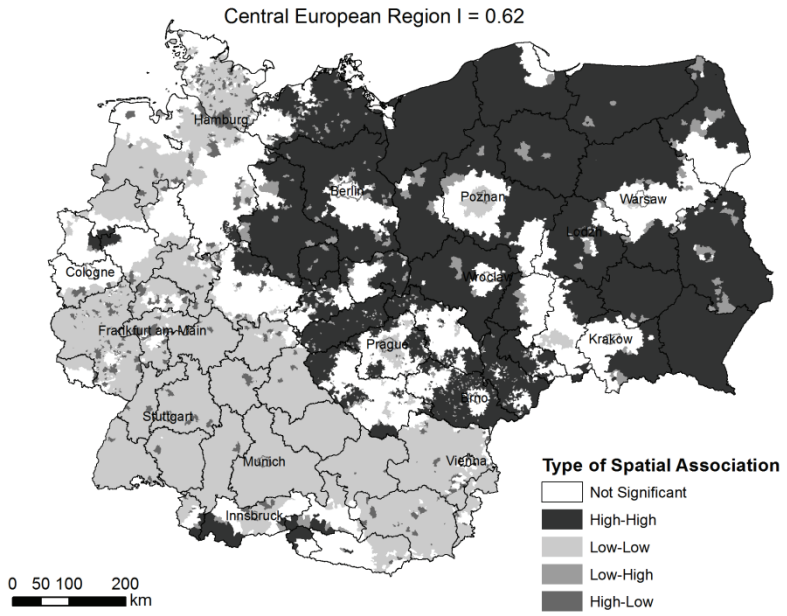


Note: The type of spatial association High-Low means that the LAU2 unit with an unemployment rate above the national mean is surrounded by LAU2 units with unemployment rates below the national mean and vice versa. A significance cut off value of 0.05 is used (after carrying out 9,999 permutations). The permutation procedure was performed using GeoDa 1.4.0. A LISA cluster map was constructed for each country separately (using the national mean) and later merged into one map.
Source: see Table 2.

First, we constructed LISA cluster maps separately for each country and then combined the results from these countries into one map (Figure 8). The strongest spatial autocorrelation is observed in Germany, where the polarity between the western and eastern parts of the country is still visible. A High-High cluster can also be found in the Ruhr region. In Poland, there are several High-High clusters. To some extent these boundaries copy historical boundaries – the former West-East Germany border in the first example and the 1938-Poland boundary in the latter. A cluster of low unemployment can be found in larger cities, from where it stretches radially in the shape of an axis. In Czechia, the

pattern is slightly more complicated. The most significant are High-High clusters in the north-western part of the country, with Low-Low clusters in Prague and its wider hinterland. Some axes of a low unemployment rate centred in Prague can be also identified (Blažek and Netrdová 2009). In Austria, there is a relatively strong west-east gradient, disrupted only by the Innsbruck region.

Figure 9: LISA Cluster Map for the Unemployment Rate in 2010 in Central European Region (Together), Weighting Scheme Queen 2nd Order



Note: The High-Low type of spatial association means that the LAU2 unit with an unemployment rate above the CER mean is surrounded by LAU2 units with unemployment rates below the CER mean, etc. A significance cut off value of 0.05 is used (after carrying out 9,999 permutations). The permutation procedure was performed using GeoDa 1.4.0. A LISA cluster map was constructed for the mean of the Central European Region as a whole.
Source: see Table 2.

An analysis that treats the entire region as a single system, and which uses a single mean, offers different information (Figure 9). This approach enables us to study, for example, the effects of borders. In general, this is suitable to understand the entire region as a whole. The west-east polarity (former Eastern Bloc) of the region is clearly visible. It may be hypothesized that historical boundaries have strong inertia and spatial patterns are visible long after they were established. However, some recent changes might be observed. First, the High-High clusters are disrupted by Low-Low clusters (and insignificant ones) around bigger cities, and development axes start to appear. The most evident is

the axis starting in Berlin, going through Poznan and Silesia to Warsaw. The west-east boundary is still very sharp, dissolving partly in Germany and on the Czech-German border.

5. Conclusions

In this paper, we have focused on spatial aspects of variability, how it can be measured and methodological issues connected with this type of analysis. We have introduced two concepts: spatial autocorrelation and relative regional variability. Despite using specific measures for their quantification (Moran's I, LISA analysis and Theil index decomposition T_B/T), we believe that the conclusions are of a general nature. *This paper highlights that the concepts of spatial autocorrelation and relative regional variability are at their most powerful when used jointly.*

Therefore, explaining the properties, advantages and disadvantages, behaviour in different situations and possibilities of combining both concepts and using them jointly were the main goals of this paper. These goals are achieved both through a theoretical-methodological discussion, as well as through an example of empirical application in international unemployment research. These findings should help to better understand the concepts and selected methods themselves, which is important for more accurate interpretation of results. The most important methodological outcomes of this study can be summarized in the following points:

- The utilized methods behave *non-linearly.* In other words, changes in different parts of their distribution can have very different contextual meaning. For example a change of T_B/T from 5% to 15% is much more important than a change from 65% to 75%. Therefore it would be wrong to interpret the same percentage changes in T_B/T and Moran's I in the same way. This is not reflected in current research, even though it is important for proper interpretation of results, especially in empirical studies.
- *Statistical inference* proved to be important especially in empirical analyses comparing different systems. When calculating the Theil index and its decomposition (and other variability measures), inference is usually not applied. In this paper, we distinguish between the stochastic and contextual components of variability. Isolating the contextual part of variability helps in comparing different geographical systems, and thus can be viewed as a form of geographical standardization. In the empirical part this proved to have significant importance. Thanks to this standardization, the order of the respective countries changed, as did the absolute differences (see the results for Austria).
- In the empirical part of the paper, we tested the importance of the *spatial weighting scheme*. Although the choice of spatial weights is often mentioned

as very important, the empirical analysis using several different spatial weighting schemes suggests the contrary, though only in relative terms. However, in the case of international research when different regional systems are compared, a spatial weighting scheme based on queen (preferably 2nd order) or rook contiguity seems to be more suitable.

- An interesting question is also the *mutual relationship between the utilized methods*. A strong and clearly positive relationship can be observed between spatial autocorrelation and relative regional variability. However, different combinations are still possible and can be categorized as follows: spatially dependent and bounded in regions (high spatial autocorrelation and high relative regional variability), spatially dependent with no relation to regions (high spatial autocorrelation and low relative regional variability), and both spatially and regionally independent (low spatial autocorrelation and low relative regional variability). These categories match empirical observations very well (see for instance Nosek and Netrdová 2010). The results in our empirical research document that the unemployment rate is a phenomenon which has typically high both relative regional variability and spatial autocorrelation. The NUTS3 regional level represents contextual values from 37% to 51% of the overall variability – a very significant share. The high and significant Moran's I values prove the importance of spatial aspects of variability irrespective of regional levels.

In addition to the methodological conclusions, there are several innovations which might be employed in empirical research. Thanks to the presented methods, international comparison can become more accurate. However, some methodological problems still remain:

- It is important to have *comparable data*. In the case of the unemployment rate this is not such a big problem, though this is unfortunately not true for other socio-economic variables. Data adjustments need to be applied.
- All data have to be *geo-referenced at the same time at a very detailed level*, which makes analyses of some socio-economic variables almost impossible.
- Differences in regional structure might complicate the process of quantification, especially the choice of a spatial weighting scheme and the results of relative regional variability. However, we show methods and strategies for how to deal with these problems.

Despite interesting outcomes and some innovative results on the spatial aspects of variability in unemployment in the Central European region (with very detailed spatial data), many opportunities for future research remain, both from a methodological and empirical perspective. From a methodological point of a view, for example, LISA maps can be used in a different way, combining different spatial weighting schemes in order to minimize the effect of its arbitrary selection. Different measures should be used in order to test these results and support our belief that these outcomes are of a general nature. More space

should be also devoted to testing the effect of different properties of the studied systems, such as number of units, number of regions, distribution of values of the variable under analysis, etc. Some simple testing of this type can be found in Novotný and Nosek (2012). These findings can be applied in empirical research. The development of unemployment rates over time would be an interesting area of study. In addition, the role of functional, historical or administrative boundaries could be assessed. Theories of spatial (cross-border) spill-overs may also be tested. In addition to unemployment, other social and socio-economic variables should be studied in order to approach spatial aspects of socio-economic variability more comprehensively. The quantification of spatial aspects of variability as presented in this paper can surely help to study these topics more accurately and effectively.

References

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Anselin, Luc. 1995. Local Indicators of Spatial Association – LISA. *Geographical Analysis* 27: 93-115.

Anselin, Luc. 2003. *An introduction to spatial autocorrelation analysis with Geoda* <https://geodacenter.asu.edu/system/files/spauto_0.pdf> (accessed September 18, 2006).

Anselin, Luc, Raymond J. G. M. Florax, and Sergio J. Rey, eds. 2004a. *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Berlin: Springer.

Anselin, Luc, Ibnu Syabri, and Youngihn Kho. 2004b. *GeoDa: An Introduction to Spatial Data Analysis* <http://www.csiss.org/events/workshops/geodaGA.pdf> (accessed September 18, 2006).

Blažek, Jiří, and Pavlína Netrdová. 2009. Can development axes be identified by socio-economic variables? The case study of Czechia. *Geografie* 114 (4): 245-62.

Cliff, Andrew D., and J. Keith Ord. 1973. *Spatial Autocorrelation*. London: Pion.

Cowell, Frank A. 1977. *Measuring Inequality*. Oxford: Phillip Allan.

Cowell, Frank A., and Emmanuel Flachaire. 2007. Income Distribution and Inequality Measurement: The Problem of Extreme Values. *Journal of Econometrics* 141 (2): 1044-72.

Cowell, Frank A., and Stephen P. Jenkins. 1995. How Much Inequality Can We Explain? A Methodology and an Application to the United States. *The Economic Journal* 105 (429): 421-30.

Fischer, Manfred M., and Arthur Getis, eds. 1997. *Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling, and Computational Intelligence*. Berlin: Springer.

Fotheringham, A. Stewart. 1997. Trends in quantitative methods I: stressing the local. *Progress in Human Geography* 21 (1): 88-96.

Fotheringham, A. Stewart, Chris Brunsdon, and Martin Charlton. 2000. *Quantitative Geography – Perspectives on Spatial Data Analysis*. London: SAGE.

Fotheringham, A. Stewart, and Peter A. Rogerson, eds. 2009. *The SAGE Handbook of Spatial Analysis*. London: SAGE.
Getis, Arthur, and Jared Aldstadt. 2004. Constructing the Spatial Weights Matrix Using a Local Statistic. *Geographical Analysis* 36 (2): 90-104.
Goodchild, Michael F., Luc Anselin, Richard P. Appelbaum, and Barbara Herr Harthorn. 2000. Towards a Spatially Integrated Social Science. *International Regional Science Review* 23 (2): 139-59.
Hampl, Martin, ed. 1999. *Geography of societal transformation in the Czech Republic.* Prague: Department of Social Geography and Regional Development, Charles University of Prague, Faculty of Science.
Harvey, David. 1973. *Social Justice and the City*. London: Edward Arnold.
Lambert, Peter J., and J. Richard Aronson. 1993. Inequality Decomposition Analysis and the Gini Coefficient Revisited. *The Economic Journal* 103 (420): 1221-7.
Mills, Jeffrey A., and Sourushe Zandvakili. 1997. Statistical Inference Via Bootstrapping for Measures of Inequality. *Journal of Applied Econometrics* 12 (2): 133-50.
Netrdová, Pavlína, and Vojtěch Nosek. 2009. Přístupy k měření významu geografického rozměru nerovnoměrností [Approaches to Measuring the Significance of Geographical Dimension of Societal Inequality]. *Geografie* 114 (1): 52-65.
Nosek, Vojtěch, and Pavlína Netrdová. 2010. Regional and Spatial Concentration of Socio-economic Phenomena: Empirical Evidence from the Czech Republic. *Journal of Economics* 58 (4): 344-59.
Novotný, Josef. 2004. Decomposition of Global and European Socio-economic Inequalities with Attention to their Regional Dimensions. *Regional and Sectoral Economic Studies* 4 (2): 59-78.
Novotný, Josef. 2007. On the Measurement of Regional Inequality: Does Spatial Dimension of Income Inequality Matter? *The Annals of Regional Science* 41 (3): 563-80.
Novotný, Josef, and Vojtěch Nosek. 2009. Nomothetic Geography Revisited: Statistical Distributions, their Underlying Principles, and Inequality Measures. *Geografie* 114 (4): 282-97.
Novotný, Josef, and Vojtěch Nosek. 2012. Comparison of regional inequality in unemployment among four Central European countries: an inferential approach. *Letters in Spatial and Resource Sciences* 5 (2): 95-101.
Openshaw, Stan. 1984. *The modifiable areal unit problem* (CATMOG 37). Norwich: Geo Books.
Rey, Sergio J. 2001. Spatial Analysis of Regional Income Inequality. *Urban/Regional 0110002, EconWPA* <http://128.118.178.162/eps/urb/papers/0110/0110002.pdf> (accessed May 10, 2007).
Rey, Sergio J., and Mark V. Janikas. 2005. Regional Convergence, Inequality and Space. *Journal of Economic Geography* 5 (2): 155-76.
Sayer, Andrew. 1984. *Method in Social Science: A Realist Approach*. London: Hutchinson.
Shorrocks, Anthony F. 1984. Inequality Decomposition by Population Subgroups. *Econometrica* 52: 1369-85.
Shorrocks, Anthony F., and Guanghua H. Wan. 2005. Spatial Decomposition of Inequality. *Journal of Economic Geography* 5 (2): 59-81.

Smith, Michael J. de, Michael F. Goodchild, and Paul A. Longley. 2013. *Geospatial Analysis – A Comprehensive Guide to Principles, Techniques and Software Tools*, 4th ed. Winchelsea: The Winchelsea Press <http://www.spatialanalysis online.com/HTML/index.html> [accessed July 15, 2013].

Spurná, Pavlína. 2008. Prostorová autokorelace – všudypřítomný jev při analýze prostorových dat? [Spatial Autocorrelation – A Pervasive Phenomenon in the Analysis of Spatial Data?]. *Sociologický časopis / Czech Sociological Review* 44 (4): 767-87.

Stine, Robert. 1989. An Introduction to Bootstrap Methods: Examples and Ideas. *Sociological Methods and Research* 18 (2-3): 243-91.

Sui, Daniel Z. 2004. Tobler's First Law of Geography: A Big Idea for a Small World? *Annals of the Association of American Geographers* 94 (2): 269-77.

Ulubasoglu, Mehmet A., and Bharat R. Hazari. 2004. Zipf's Law Strikes Again: The Case of Tourism. *Journal of Economic Geography* 4 (4): 459-72.

Unwin, Antony, and David Unwin. 1998. Exploratory Spatial Data Analysis with Local Statistics. *The Statistician* 47 (3): 415-21.

Wong, David W. S. 2009. The Modifiable Areal Unit Problem (MAUP). In *The SAGE Handbook of Spatial Analysis*, ed. A. Stewart Fotheringham and Peter A. Rogerson, 105-23. London: SAGE.

Yeung, Henry Wai-chung. 1997. Critical realism and realist research in human geography: a method or a philosophy in search of a method? *Progress in Human Geography* 21 (1): 51-74.