The importance of aggregation in regional household income estimates: A case study from Hungary, 2019

Tibor Bareith
Institute of Economics
Centre for Economic and Regional Studies
Budapest, Hungary
E-mail: bareith.tibor@krtk.hun-ren.hu

Adrián Csizmadia
Hungarian University of Agriculture and Life Sciences,
Doctoral School of Business and Management, Institute of Rural Development and Sustainable Economy,
Kaposvár, Hungary
E-mail: csizmadia.adrian@gmail.com

Do the results of analyses with spatial data depend on the level of aggregation? The literature refers to the problem of spatial aggregation as the „modifiable areal unit problem" (MAUP). The main research question is whether spatial analysis using different estimation techniques (OLS, SEM, SAR, SDM) is affected by the MAUP problem. Our spatial analysis focuses on household incomes. For Hungary, spatial data are available at the municipal, district and county levels to explore the problem, and income inequality is average at the European level.

The results suggest that the MAUP problem exists in Hungary. Increasing the aggregation significantly reduces the proportion of significant explanatory variables for all models. This implies that spatial analyses should be performed at the smallest possible spatial scale to obtain the most accurate model estimates.

The spatial autocorrelation of the income indicator is also affected by aggregation: there is no difference in the global autocorrelation between the municipal and district level, but the global indicator is much lower at the county level. The degree of local autocorrelation decreases significantly with the level of aggregation.

Another finding is that household incomes are mainly influenced by the working-age population, the presence of entrepreneurs, the number of jobseekers and schooling, while occupational classification also has a significant impact on incomes.

Keywords:
income,
aggregation,
MAUP,
spatial models,
spatial autocorrelation

Introduction

The issue of aggregation is important to address because the results of spatial studies are used in administrative, spatial development, and economic policy decision-making, and studies conducted at different levels of aggregation may lead to different results, which may influence the allocation of spatial development or other subsidies. In addition, spatial aggregation has other effects, which Dusek (2004: p. 2) lists as follows: effect on the results of descriptive statistics; effect on the interpretation of results; practical-policy effect; effect on the applicability of inferential statistics; effect on theories and models.

The literature refers to the problem of spatial aggregation as the „modifiable areal unit problem" (MAUP). This includes, among others, the problem of area delimitation, zoning and spatial aggregation. The MAUP thus has two related but distinct components: the scale problem and the zone problem. Scaling is the process of aggregating the same spatial data into increasingly larger sets of spatial units of analysis. Zonation, on the other hand, is the spatialization of a constant number of spatial units into alternative analysis zones of different locations (Openshaw–Taylor 1979, Openshaw 1984, cited in Jelinski–Wu 1996). According to Dusek (2003), a priori given delimitations may be justified not only because of the necessity of data availability but also because the spatial units with which we are already familiar are the easiest to process and interpret the results.

This paper examines the issue of scale effects and whether social and economic variables aggregated into different territorial units lead to different results for different econometric models. Based on the literature (e.g., Tagashira–Okabe 2002, Dusek 2003, Wang–Di 2020), for spatial analyses, it is appropriate to consider the data available at the settlement level as the smallest possible spatial base unit. More detailed spatial disaggregation provides more information, as some of the information is lost in spatial aggregations. However, it should also be borne in mind that it is advisable to choose the spatial division that is most advantageous from the point of view of the problem under investigation. For example, in the case of Hungary, it is more appropriate to use a regional or county division to identify broader trends (Dusek 2001, 2003). It is appropriate to examine certain social phenomena in the spatial framework at the territorial level, where they are actually manifested and organized (Nemes Nagy 1990).

Literature review

The first recognition of the MAUP dates back many decades (Gehlke–Biehl 1934), but the problem is still often overlooked by researchers, with serious consequences for the reliability of the results of analyses using spatially aggregated data. Indeed, the basic assumption of the MAUP is that the levels of aggregation and the arbitrary and modifiable sizes, shapes and layouts of zones affect the validity of the results obtained from the analysis of spatially aggregated data. In turn, the conclusions drawn from these results have implications for public policy and the allocation of economic, health and other resources (Nelson–Brewer 2015), as each allocation results in different values of aggregate statistics (Pawitan–Steel 2009).

Researchers investigating the MAUP have found that scale effects and zone effects cause a significant loss of information when aggregating data in large geographic units (Nthiwa 2011, Bell et al. 2013 cited in Prouse et al. 2014). Dusek (2003) investigates the relationship of spatial aggregation with certain statistical indicators such as correlation, the ratio of minimum to maximum values, and weighted relative standard deviation. He gives examples to show that as the spatial level increases, the correlation becomes closer because the variance of the variables decreases; for example, the higher the level of aggregation of the average potato and wheat yields in British provinces, the higher the correlation (Yule–Kendall 1964, cited in Dusek 2003). The ratio of maximum and minimum values decreases with aggregation because the averaging reduces the maximum values and increases the minimum values. The weighted relative standard deviation decreases with aggregation (telephone line per capita, personal income tax per capita, car per capita). For example, when counties are aggregated into regions, the differences between counties aggregated into the same regions disappear. However, this finding is not always true for the standard deviation calculated from unweighted data when merging territorial units (income tax base per capita) (Dusek 2003).

The reason for the general increase in correlation coefficients is that, regardless of the method of aggregating spatial data (e.g., by averaging or summing), the process has a „smoothing effect", so that as aggregation increases, the variance of the variable decreases. Researchers have demonstrated this both through mathematical and nonmathematical methods (Robinson [1950] and Fotheringham–Wong [1991]). The work of Lee et al. (2015) on the determinants of housing prices in South Korea provides evidence for the above claim in a concrete case study, as did the study of Zhou et al. (2022) for Shanghai, in which they investigate the effects of job, housing and social demographic characteristics on commuting in a MAUP-focused manner. However, the change in correlation coefficients in this direction is not automatic for all aggregations. In bivariate regression models, the correlation coefficient and the estimated coefficient do not necessarily increase monotonically with increasing levels of aggregation (Blalock 1964, Clark–Avery 1976 cited in Fotheringham–Wong 1991).

Most studies on the impact of the MAUP on multivariate statistical analyses have concluded that the impact of the modifiable spatial unit problem is unpredictable. As a consequence, multivariate analyses with spatial data may lead to unreliable results (Fotheringham–Wong 1991, Tagashira–Okabe 2002). The effect of aggregation in these analyses is not predictable (Openshaw 1984); therefore, empirical work has a major role to play in demonstrating the importance of aggregation (Dusek 2001, 2004).

Nelson–Brewer (2015) also list research that has suggested that the effect of data aggregation is minimal or nonexistent (Amrhein–Flowerdew [1992]: Poisson regression, Flowerdew [2011]: correlation of demographic variables). In addition to presenting the results of previous research, the two researchers also present their own study of spatial autocorrelation analysis of median income and cancer incidence in Pennsylvania and New York states. They find that the level of aggregation affects the results, with similar results at the census tract and block group level but different conclusions from the aggregated data at the county level. From this, the authors conclude that *„Spatial phenomena operate and interact at different scales. These operations and interactions are not constrained to a single scale but happen and respond differently across scales. Characterizing relationships between data aggregation and spatial scale using an exploratory statistical–visual approach can greatly enhance our understanding of place"* (Nelson–Brewer 2015: p. 14).

According to Fotheringham–Wong (1991), it is not clear to what extent the results from univariate and bivariate analysis can be transferred to multivariate analysis. While Openshaw (1978) shows that in simple linear regression, the magnitude of the estimated coefficient increases with increasing aggregation of the data, it is not clear whether the same phenomenon can be observed in multivariate regression.

The value of the indices measuring spatial autocorrelation is also affected by spatial aggregation. Both Jelinski–Wu (1996) and Dusek (2004) find from empirical studies that spatial aggregation reduces the value of Moran's I. This can be explained by the fact that when analyzed at smaller spatial units, autocorrelations within spatial units can be detected, unlike at a higher level of aggregation where these autocorrelations disappear. Openshaw–Taylor (1979) show that aggregation increases the value of the correlation coefficient, but no clear relationship between MAUP and spatial autocorrelation can be detected (cited in Fotheringham–Wong 1991). Lee et al. (2015) find that the neighborhood effect is limited and depends on the level of aggregation of the variables.

Qi–Wu (1996), in their empirical work on plant community analysis, find that changes in spatial scale also significantly affect the values of three autocorrelation indices (Moran coefficient, Geary ratio and Cliff-Ord statistic). The degree of spatial autocorrelation tends to decrease as the spatial scale increases. Thus, the author argues that conducting spatial analyses at a single spatial scale tends to yield little useful, or even misleading, information.

Fotheringham–Wong (1991) find that multivariate models are highly sensitive to aggregation levels and zoning (see also Dusek 2001). According to the authors, it is almost impossible to analytically predict the effect of scale change or zonal change on parameter estimates in multivariate analysis. They investigated the sensitivity of the calibration results of two multivariate models, a multivariate linear regression model and a multivariate logit regression model, to aggregation. They conclude that by aggregating data in different ways, almost any desired result can be obtained. Jelinski–

Wu (1996) also find that the impact of MAUP on multivariate analyses is complex and largely unpredictable.

Pietrzak (2014) examines the causal relationship between the level of business investment expenditure per capita and the number of economic units per capita and the relationship between the registered unemployment rate and the level of investment expenditure per capita in Poland. Data are aggregated at the Nomenclature of territorial units for statistics (NUTS) 4 (district) and NUTS 3 (microregion) levels. In general, his calculations suggest that correlation analysis and regression analysis may lead to different conclusions depending on the level of aggregation. However, the aggregation of variables did not affect the appearance of spatial dependence as measured by Moran's I test. Looking at the results in more detail, the simulation analysis shows that for variables expressed in relative quantities, the mean does not change as a result of the aggregation process, and the variance decreases. For variables expressed in absolute quantities, the mean and the variance increase. Pietrzak (2014) argues that these results provide an important argument for the use of relative quantities in spatial economic analysis. For the variables expressed in relative quantities, the mean of the correlation between variables does not change during aggregation, but the standard deviation of the correlation increases significantly. Thus, as a result of the aggregation process, the values of the Pearson correlation coefficient may change significantly.

For the regression analysis, only variables expressed in relative quantities were used by Pietrzak (2014). The aggregation process also increases the standard deviation values for the regression parameter and the coefficient of determination. This also significantly increases the standard error of the regression parameter.

In addition to a detailed review of theoretical work on the impact of the MAUP on multivariate studies, which finds that some studies have shown dramatic differences in the results of regression estimates due to differences in the aggregation scale used for geographical units (e.g., Amrhein–Flowerdew 1992, Wong et al. 1999, Manley 2006, Pawitan–Steel 2009, Shah et al. 2014), Prouse et al. (2014) also report empirical results.

Using a multivariate regression model, Prouse et al. (2014) investigate the factors that determine the percentage of residents classified as low income in the city of Halifax (Canada) at two levels of aggregation (census tracts [CT]; dissemination areas [DA]; the latter being the smaller scale unit). This is measured by the low income cut off (LICO). Their results show that MAUP is present in the analysis but does not change the overall conclusions on the main determinants of socioeconomic outcomes. The direction of the relationships between variables does not change at different geographical scales, but the magnitude of the relationships does: the CTs provide a clearer picture of the proportion of low-income earners, better model fit, and fewer statistically significant variables. DAs show a poorer model fit, but the results show that at this level, several factors significantly affect the LICO indicator.

Thus, at a theoretical level, it is difficult to decide whether the CT or DA level is more appropriate for the study of socioeconomic inequalities. An analysis at the CT level is more helpful in understanding general trends, while an analysis at the DA level is more suitable for identifying trends in individual neighborhoods (Prouse et al. 2014).

Arbia–Petrarca (2011) investigate the impact of MAUP on the accuracy of parameter estimation in spatial econometric models (spatial lag and SARAR model). Their findings show that the loss of efficiency due to aggregation is generally mitigated by the presence of a positive spatial correlation parameter and exacerbated by the presence of a negative spatial correlation parameter. This is because in the case of positive spatial correlation, we aggregate similar values, thus preserving variability. Conversely, negative spatial correlation implies aggregation between very different values, which implies a severe loss of variability.

In summary, the MAUP effect in multivariate methods can only be identified individually through empirical studies, and it is advisable to perform the analysis at multiple levels of aggregation for comparison. For this reason, we investigate the determinants of per capita income in Hungary for spatial econometric models run at different levels of aggregation. The stability of each variable can be inferred by computing the results with different cutoffs (Fotheringham–Wong 1991). The data are analyzed at the level of municipalities, districts and counties. Below, we describe the main differences between the models and the reasons for these differences.

Material and methods

Income is an important indicator of regional development, which provides a basis for comparing the well-being of the population of different territorial units. This is because income data are widely available and detailed sources are accessible on them (Pénzes 2011). The advantage of income indicators (as opposed to gross domestic product (GDP), for example) is that they are geographically included in statistical accounts based on place of residence rather than place of work (McCann 2020). Thus, in our research, we also used income indicators as a measure of economic development.

The data used in the analysis are from the National Regional Development and Spatial Planning Information System (TeIR) (land information system – LIS) and refer to the year 2019. The outcome variable is the total disposable domestic income per working-age resident (aged 15–64). This income category excludes social and other income and only includes income subject to social security contributions. Our explanatory variables are composed of important economic and social variables (see Table A1 in the Annex), which, in addition to previous experience, were chosen to be available at the municipal and district levels. The main objective of our research was to compare the results of econometric models run at different levels of aggregation. The presentation and descriptive statistics of the variables can be found in Annex Table A2 and A3.

Among the global autocorrelation tests, Moran's I-test is used, while the local Moran test and the local Geary C index are used to identify spatial autocorrelation trends and spatial clustering (Moran 1948, Cliff–Ord 1981, Anselin 1995, cited in Varga 2002, Tóth 2003, Geary 1954).

Studies analyzing the determinants of development, income status or other social phenomena have extensively used econometric models that take into account spatial correlations and neighborhood relations (e.g., Szendi 2017, Chasco et al. 2008, Saib et al. 2014). Case studies show that these models have better fit at all levels of aggregation compared to standard Ordinary Least Squares (OLS) models (Lee et al. 2015).

In spatial regression models, the value of a variable at a certain point is related to the values of the same variable at other points in space using weight matrices. In our analysis, we incorporated 3 spatial econometric models. We used spatial lag (SAR), spatial error (SEM) and spatial Durbin (SDM) models (Fitriani et al. 2021, Duran–Karahasan 2022, Aritenang 2022) to examine how our explanatory variables affect the outcome variable at the municipality and district levels.

In the SAR (spatial lag) model, the weighted average of the values measured in the observation units belonging to a given neighborhood degree is used (Varga 2002). General form of the SAR model (LeSage–Pace 2009):

$$Y = \rho W Y + X\beta + \varepsilon$$

The SEM model also filters out the spatial autocorrelation effects of the explanatory variables and the independent variable and serves to correct for spatial autocorrelation between error terms (Varga 2002). General form of the model (LeSage–Pace 2009):

$$Y_{(N*1)} = X\beta + \text{u, and } u = \lambda W u + \varepsilon$$

In the SAR model, ϱ (rho) is a scalar parameter that indicates the effect of the dependent variable of adjacent areas on the value of the dependent variable of the area under study (Drukker et al. 2013). In the SEM, λ (lambda) indicates the spatial correlation between error terms. When ϱ and/or λ are not 0, models that include spatial autocorrelation can provide a more accurate estimate than the OLS model (Tsionas 2019). The more appropriate of the two models can be selected by using Lagrange Multiplier (LM) diagnostics, i.e., running LM-error and LM-lag tests (Anselin 2005, cited in Szendi 2017).

The SDM model is an extension of the SAR model. SDM includes the effect of spatial lags for both independent and dependent variables (unlike the SAR model, which only considers the effect of the spatial lag on the independent variables). The general form of the model is as follows (LeSage–Pace 2009):

$$Y = \rho W Y + \alpha \iota_n + X\beta + \mathrm{WX}\theta + \varepsilon$$

Results

Regional autocorrelation of domestic income per working-age population

The spatial autocorrelation between income indicators is widely documented in the literature (e.g., Rey–Montouri 1999, Kalogirou–Hatzichristos 2007). However, the effect of aggregation (scaling) on spatial autocorrelation can be verified by conducting empirical studies, whether Moran's I or the Geary c indicator (Dusek 2004). In this study, both indicators are used to measure spatial autocorrelation at three different aggregation levels: municipality, district and county. Spatial autocorrelation for total domestic income per capita is examined using several weighting matrices (queen, rook and k-nearest neighbors). (*The calculations are also performed with the contiguity values of the queen and rook matrices 3 and 5, but in these cases only the Moran's I values are reported, not the map results. This is because the aim of our research is only to analyze the effect of aggregation, and for this purpose, we have chosen the value 1 among the different contiguity values. The comparison of the weight matrices with different contiguity values could be the subject of a separate study.)*

The global value of Moran's I (Table 1) decreases in four of the six cases examined with different weight matrices and increases minimally in two cases when aggregating income data at the district level relative to the municipal level. This phenomenon is consistent with the finding in much of the literature that at higher scaling levels, the value of spatial autocorrelation tends to decrease.

At both the municipal and district levels, the degree of autocorrelation is moderately strong. However, when we aggregated domestic income per working age population at even higher levels, we obtained results consistent with the literature, as Moran's I value showed a significant decrease: in some cases, the positive autocorrelation turned negative.

The reason for the reduction in the spatial autocorrelation is that this is because the higher level of aggregation masks differences within territorial units, the size effect causes smaller territorial units to merge into larger ones and the whole unit to appear as a homogeneous unit (Brunsdon et al. 2002, cited in Dusek 2004, Jelinski–Wu 1996).

Table 1

Moran's I values at different aggregation levels

Level of aggregation	Weight matrix					
	queen, order of contiguity		rook, order of contiguity		k-nearest	
	1	3	1	3	4	8
Municipality	0.611	0.447	0.611	0.478	0.636	0.609
District	0.612	0.335	0.610	0.334	0.646	0.593
County	0.449	–0.285	0.449	–0.285	0.435	0.187

Source: Own editing based on GeoDa.

Local autocorrelation tests can be used to describe the spatial pattern structure (clustering) (Figures A1 and A2 in the Appendix). The local Moran's I and Geary C can be used to detect whether neighboring observations of the same phenomenon are correlated; in other words, spatial autocorrelation describes the degree of spatial clustering, which allows us to see spatial units where the values measured at one location are partly determined by values measured at neighboring locations.

In Figures A1 and A2 (queen neighborhood weight matrix, order of contiguity 1), the two indicators show very similar clustering at the municipality and district level, with significant differences at the county level. Studies using both indicators show very similar clustering at the municipality and district levels, with significant differences at the county level. At the municipality level, we can see that the municipalities located in the Budapest agglomeration area and, in the North and West Transdanubian regions are those where the high income of the population of a municipality has a positive effect on the income of its neighboring population. This is indicated in the darkest color on the maps. It is clear that these are regions of the country where the income of the population has been high since the change of regime. The only difference between the two maps of municipalities is that the Geary C indicator shows that high-income clusters are also found in the central and southern parts of the Great Plain, while the local Moran's I indicator shows that this phenomenon is not observed in this region. Of the 3155 municipalities, the local Geary C classifies 794 in the high-high cluster, while the local Moran's I classifies only 635. Similarly, for the clustering of low-income settlements, Geary C indicates more settlements (645) compared to the other spatial autocorrelation indicator (553). These spatial units are indicated by the lightest gray color. Maps show clustering of poorer settlements in the South Transdanubian region and along the eastern and northern borders of the country, but the local Geary C also shows smaller low-low clusters in almost all areas of the country, even in the more inland regions, in contrast to the local Moran's I.

If we examine Moran's I or Geary C indicators with a rook (order of contiguity 1) or k-nearest (4 closest neighbors) weight matrix, we obtain similar results both in the number of settlements in each cluster and, in the pattern of the maps. (Figures A3–A6 in the Appendix show the number of spatial units in each cluster.) Our basic findings remain unchanged: the richest clusters are found around the capital and, in the northwest of the country, while the poorest clusters are located in the South Transdanubian region and, in the eastern and northeastern parts of the country. (For rook and queen neighborhoods, the patterns are very similar, since the boundaries of territorial units usually touch each other not only at one but at several points, so the neighborhood relationships are almost the same for both types of weight matrix.)

Aggregation also reduces the size of high-high and low-low clusters. Only those areas where the autocorrelation between neighboring municipalities is truly strong are retained on the maps at the district and county level compared to the municipal level.

At the district level, for both indicators, the clusters of the western border and the central parts of the country are the ones that disappear most in comparison with the maps of the municipalities.

However, county-level maps show a large difference between local Moran’s I and Geary C values. In the case of the queen neighborhood, the former shows high-high clustering in Komárom-Esztergom and Pest counties, while the latter clusters Vas, Győr-Moson-Sopron and Komárom-Esztergom, where the favorable income situation of the county's population improves the income position of some of its neighbors. This difference between the two indicators is also observed in the low-low clusters. According to the local Moran’s I, the poor income situation of Borsod-Abaúj and Hajdú-Bihar counties has a mutually negative impact on the status of the other. In the local Geary C map, on the other hand, only Csongrád-Csanád County is included in this category. For the rook neighborhood matrix, exactly the same difference can be observed between Moran's I and Geary's C patterns. For the k-nearest neighborhood, the clustering of counties in a favorable income position is roughly the same for Local Moran's I and Geary's C, with only Győr-Moson-Sopron and Pest counties being the difference between the two maps. The low-income counties are the same as in the queen weight matrix: Borsod-Abaúj Zemplén and Hajdú-Bihar in one case and Csongrád-Csanád in the other.

Specification of regression models and descriptive statistics

The outcome variable in our regression models is the total domestic income per working-age population (15–64 years). Several limiting factors must be taken into account when choosing the explanatory variables. An important criterion is to choose variables that have been shown in similar studies to be able to explain the spatial distribution of per capita income (Csizmadia–Bareith 2022) and that do not have the problem of endogeneity, which in our case means that the causal direction is clear between the explanatory and the outcome variable. In country-level studies related to the testing of convergence theories, the following explanatory variables are the most commonly included in the models as factors affecting income/GDP: technological differences, employment structure, human capital stock, migration, unemployment, government expenditure, public debt, education, savings rate, and population growth (Barro–Sala-i-Martin 1992, Caselli et al. 1996, Glaeser et al. 1995, Islam 1995, Mankiw et al. 1992).

Furthermore, an important criterion is that the selected explanatory variables should also be available at the smallest spatial scale of the study, i.e., at the municipality level. Along these selection criteria, a total of 19 explanatory variables are used to build our econometric models. These can be grouped into three categories.

The social characteristics include the number of working-age permanent population (15–64 years), the number of registered jobseekers and the number of people with different levels of education (vocational diploma, without high school graduation/certificate, with high school diploma/university, college graduation).

The relationship between population size and income level leads to diverging results in some studies: some researchers find a strong relationship between population size and income level (Molnár–Ilk 2010), while others refute (Lőcsei 2004) or nuance (Turczak–Zwiech 2014) the finding that there is a positive relationship between population size and per capita income. Migration also particularly affects the working-age population in Hungary. Support programmes are key to return migration, especially for less skilled workers (Lados–Hegedűs 2019).

The relationship between the number of jobseekers and per capita income is much clearer in the research, with the negative impact of unemployment on income being confirmed by several studies (e.g., Kalogirou–Hatzichristos 2007, Chasco et al. 2008, Kilgarriff–Charlton 2020). Hajdú–Koncz (2022) points out that the primary labor market employment rate is steadily declining despite state-funded vocational training. Another significant impact can be attributed to public employment.

Educational attainment is a hard barrier to occupational positions, and through this, it has a significant impact on income levels (Vastagh 2015). Empirical studies show that in households where the head of household has a university degree, the equivalent household income of individuals is 60–80% higher than the average income. The household income of university graduates is also significantly higher than the household income of those with a high school diploma (Tóth 2006, cited in Medgyesi 2006). Based on these results, we also expect educational attainment indicators to have a positive impact on income. Alhendi et al. (2021) note that primary education plays a more decisive role than secondary education in the development of a region.

The economic characteristics variables include the number of individual and joint enterprises as a share of the population, assuming that entrepreneurship can provide a higher than average income for the entrepreneur and his or her family and can also increase the average income in the municipality through job creation. Previous research on Hungary has found that the concentration of entrepreneurship and businesses has a positive impact on income (Szendi 2017, Kolber et al. 2019). The consideration of the balance sheet total of businesses as a share of the population is that the larger the size of businesses in a municipality, the more competitive they are, the higher the income they provide to the municipality's population, and the more likely they are to create jobs (Gál et al. 2014).

The European Union (EU) subsidies explanatory variable includes subsidies received from the EU between 2015 and 2019 under operational programmes by place of implementation, including both local government and private sector funding. In the TeIR, only the Rural Development Programme grants are not available. The literature (e.g., Fertő–Varga 2015) suggests that EU subsidies are unlikely to affect household incomes. This phenomenon is probably because some of the subsidies that are used in municipalities increase the local economy and, presumably, local wages, but the other part of the funds that are targeted at municipalities and the competitive sector are not (necessarily) reflected in local labor incomes. For example, a business operating in a given municipality may not employ only local workers, or a municipality may not necessarily hire local businesses to carry out development. At the same time, if cohesion policy is to be taken seriously, lagging settlements and areas should also be supported (Atkinson 2019), taking into account that larger cities within a given territorial unit act as power centers and engines of development in Hungary (Rechnitzer et al. 2019).

Variables on the structure of employment form are the third group of explanatory variables in our econometric models. For each territorial unit, we identify the proportion of the total number of employees living there that belong to each occupational class of the Hungarian Standard Classification of Occupations (FEOR-08). The structure of employment is expected to affect the income situation of the population, as there are clear wage differentials between occupational classes, with unskilled labor market status (unskilled and unskilled unskilled) reducing income (Sik 2014) while holding positions requiring higher education is associated with higher income.

The natural logarithm of the variables was used in the regression models, and the table of descriptive statistics (Annex Table A2) shows the natural units of the variables before logarithmicization.

Results of the regression models

The presence of spatial autocorrelation for the outcome variable has been demonstrated above. The autocorrelation effect between variables justifies the inclusion of neighborhood effects in our regression analysis. Four different regression models (OLS, SAR, SEM, SDM) are used for the analyses at two different aggregation levels (municipal, district). (The county level may lead to biased estimates due to the small number of observations, so we omit this level of aggregation from our analysis.). GeoDa, GeodaSpace and STATA 17.0 software were used to perform the analyses.

The statistical software used allows us to incorporate the neighborhood matrix (control neighborhood) into the OLS model so that the regression analysis result indicates the presence or absence of spatial dependence. We can then also use the Lagrange multiplier diagnostic to decide whether the use of a SAR (spatial lag) or SEM (spatial error model) leads to a more accurate result. The results in Annex Table A4 show that the SEM (spatial error model) gives the most accurate results at both the municipal and district levels. We interpret the results of all four models at both spatial levels. (It is important to note that the model suggested by Lagrange multiplier diagnostics is the SEM at both the municipal and district level, as opposed to the SAR model. Our calculations were also performed using the SDM model.) The models are presented in Tables 2–5 by the estimation procedure at the municipal and district levels. At the municipal level, all variables are used; at the district level, one variable (*UCgrad*) is excluded due to multicollinearity. Employment variables were included in separate models, and their combined effect was tested by the F test.

For the OLS model at the municipal level (Table 2), the variables *Wapop, Ienterp, Jseeker, EUsubs, Vocdip, Hsdip* and *UCgrad* are significant, while *EUsubs* is not significant after controlling for employment data, but the individual occupational categories together have an impact on income. These are inversely related for *jobseekers* (*Jseeker*) and *EUsubs*. The magnitude of the coefficients on education decreases after the inclusion of the data on occupation.

The model produces far fewer significant variables at the district level. Only 4/8 variables affect the evolution of total domestic income per capita compared to the municipal level, where the l level and 7/9 variables are significant. At the district level, *the number of jobseekers* (*Jseeker*) is not significant, and among the variables on educational attainment, only *Hsdip*.

The SAR model (Table 3) shows a high degree of agreement with the OLS model results at the municipal level. At the district level, the use of a SAR model is proposed based on the Lagrange multiplier diagnostics. There are 5/8 significant variables in the model. *Unemployment (Jseeker)* and *working-age population (Wapop)* reduce earnings, while *companies with larger balance sheets (Tota) and a higher proportion of persons with a high school diploma (Hsdip)* result in higher pay.

EUsubs are significant at the municipal and district levels, and the models suggest that EU subsidies reduce incomes. For the municipal and district models, there is a sign shift in the *Wapop* variable; at the district level, an increase in the *working-age population* reduces income at the district level.

The SEM model (Table 4) gives the most accurate result at the municipal and district aggregation levels based on Lagrange multiplier diagnostics. All but two of the other explanatory variables (*Jenterp, Tota*) are significant. *Population (Wapop), the number of cooperative enterprises (Ienterp)* and *the educational variables (Vocdip, Hsdip, UCgrad)* increase incomes. *Unemployment (Jseeker) and EU subsidies (EUsubs)*, on the other hand, cause a decrease in income according to the model.

At the district level, the model produces 3/8 and 4/8 significant variables. *Unemployment (Jseeker)* is negatively related to income, while the variables for *the size of the balance sheet as a share of the population (Tota)* are positively related.

The results for the SDM model (Table 5) are very similar to the SEM (Table 4), with the *Ienterp* variable not significant in the model with employment variables compared to the SEM. At the district level, 3/8 variables are significant for both models. Interestingly, the variable *Tota* is significant in the model without employment variables, while the variable *Jseeker* is significant in the model with employment data.

Overall, the different estimation procedures for the different levels of aggregation give very similar results, and the estimates are consistent. For all models, the spatial effect is significant, and the employment structure has an impact on incomes.

Table 2

OLS model

Y – Total domestic income per working age population, HUF	Municipal		District	
	(X1–X9)	(X1–X19)	(X1–X8)	(X1–X19) (without X9)
Constant	11.888*** (0.175)	14.583*** (0.212)	11.251*** (0.719)	17.872*** (0.836)
X1 – Wapop	0.029*** (0.004)	0.016*** (0.003)	–0.002 (0.015)	–0.019 (0.012)
X2 – Ienterp	0.037*** (0.010)	0.015* (0.009)	0.032 (0.061)	0.066 (0.059)
X3 – Jenterp	0.007 (0.007)	0.003 (0.007)	0.047 (0.040)	0.009 (0.038)
X4 – Tota	–0.002 (0.001)	–0.001 (0.001)	0.037*** (0.013)	0.016** (0.008)
X5 – Jseeker	–0.096*** (0.007)	–0.050*** (0.007)	–0.096*** (0.026)	–0.023 (0.023)
X6 – EUsubs	–0.001* (0.001)	–0.000 (0.000)	–0.016 (0.010)	–0.013* (0.007)
X7 – Vocdip	0.172*** (0.028)	0.066** (0.030)	0.081 (0.092)	–0.003 (0.082)
X8 – Hsdip	0.265*** (0.020)	0.119*** (0.018)	0.483*** (0.075)	0.246*** (0.085)
X9 – UCgrad	0.058*** (0.010)	0.038*** (0.010)	–	–
Employment category variables	X	✓	X	✓
F-statistics for employment category variables	not relevant	***	not relevant	***
R^2	*0.6358*	*0.7358*	*0.7344*	*0.8817*
Adjusted R^2	*0.6347*	*0.7352*	*0.7216*	*0.8680*
Log likelihood	*820.576*	*1,326.708*	*151.213*	*221.988*
Number of observations	*3,155*	*3,155*	*175*	*175*

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$) (in parentheses the White-standard error).
Source: Own editing based on GeoDa and Stata.

Table 3

Spatial lag (SAR) model

Y – Total domestic income per working age population, HUF	Municipal		District	
	(X1–X9)	(X1–X19)	(X1–X8)	(X1–X19) (without X9)
Constant	11.887*** (0.172)	14.574*** (0.210)	11.546*** (0.696)	18.025*** (0.799)
X1 – Wapop	0.026*** (0.004)	0.013*** (0.004)	–0.028* (0.017)	–0.033** (0.013)
X2 – Ienterp	0.036*** (0.010)	0.015* (0.009)	–0.080 (0.054)	0.023 (0.052)
X3 – Jenterp	0.006 (0.007)	0.003 (0.007)	0.034 (0.038)	–0.009 (0.037)
X4 – Tota	–0.002 (0.001)	–0.001 (0.001)	0.034*** (0.011)	0.015** (0.007)
X5 – Jseeker	–0.096*** (0.007)	–0.051*** (0.007)	–0.106*** (0.023)	–0.029 (0.020)
X6 – EUsubs	–0.001* (0.001)	–0.000 (0.000)	–0.019** (0.010)	–0.016** (0.007)
X7 – Vocdip	0.170*** (0.028)	0.065** (0.029)	0.047 (0.086)	0.015 (0.075)
X8 – Hsdip	0.267*** (0.020)	0.119*** (0.018)	0.572*** (0.075)	0.284*** (0.078)
X9 – UCgrad	0.057*** (0.010)	0.037*** (0.010)	–	–
ϱ (rho)	0.000** (0.000)	0.000** (0.000)	0.001*** (0.000)	0.001*** (0.000)
σ (sigma)	0.186*** (0.004)	0.159*** (0.004)	0.098*** (0.006)	0.066*** (0.005)
Employment category variables	X	✓	X	✓
F-statistics for employment category variables	not relevant	***	not relevant	***
R^2	*0.4940*	*0.6526*	*0.7554*	*0.9314*
Adjusted R^2	*0.4925*	*0.6505*	*0.7436*	*0.9235*
Log likelihood	*823.572*	*1,329.005*	*158.183*	*226.475*
Number of observations	*3,155*	*3,155*	*175*	*175*

Notes: (*** p < 0.01; ** p < 0.05; * p < 0.1) (in parentheses robust standard error).
Source: Own editing based on GeoDa.

Table 4

Spatial error (SEM) model

Y – Total domestic income per working age population (HUF)	Municipal		District	
	(X1–X9)	(X1–X19)	(X1–X8)	(X1–X19) (without X9)
Constant	12.889*** (0.087)	14.578*** (0.137)	11.553*** (0.651)	18.046*** (0.878)
X1 – Wapop	0.026*** (0.004)	0.013*** (0.003)	–0.027 (0.017)	–0.033*** (0.012)
X2 – Ienterp	0.036*** (0.007)	0.015** (0.006)	–0.080 (0.063)	0.024 (0.054)
X3 – Jenterp	0.007 (0.006)	0.03 (0.005)	0.034 (0.039)	–0.008 (0.039)
X4 – Tota	–0.002 (0.001)	–0.001 (0.001)	0.034*** (0.012)	0.015* (0.009)
X5 – Jseeker	–0.097*** (0.004)	–0.051*** (0.004)	–0.107*** (0.021)	–0.029** (0.017)
X6 – EUsubs	–0.001* (0.000)	–0.000 (0.000)	–0.019* (0.010)	–0.016** (0.007)
X7 – Vocdip	0.171*** (0.014)	0.066*** (0.015)	0.049 (0.076)	0.015 (0.085)
X8 – Hsdip	0.267*** (0.013)	0.119*** (0.013)	0.572*** (0.082)	0.283*** (0.084)
X9 – UCgrad	0.057*** (0.007)	0.037*** (0.007)	–	–
λ (lambda)	0.000** (0.000)	0.000* (0.000)	0.002*** (0.000)	0.001*** (0.000)
σ (sigma)	0.186*** (0.002)	0.159*** (0.002)	0.098*** (0.005)	0.066*** (0.004)
Employment category variables	X	✓	X	✓
F-statistics for employment category variables	not relevant	***	not relevant	***
R^2	*0.6309*	*0.7335*	*0.5045*	*0.8061*
Adjusted R^2	*0.6299*	*0.7319*	*0.4806*	*0.7837*
Log likelihood	*823.369*	*1,328.450*	*158.018*	*226.171*
Number of observations	*3,155*	*3,155*	*175*	*175*

Source: Own editing based on GeoDa.

Table 5

Spatial Durbin (SDM) model

Y - Total domestic income per working age population, HUF	Municipal		District	
	(X1–X9)	(X1–X19)	(X1–X8)	(X1–X19) (without X9)
Constant	11.506*** (0.196)	13.769*** (0.224)	10.315*** (0.746)	17.029*** (0.913)
X1 – Wapop	0.019*** (0.006)	0.005 (0.005)	–0.042*** (0.016)	–0.045*** (0.013)
X2 – Ienterp	0.035*** (0.010)	0.019** (0.008)	–0.022 (0.060)	0.013 (0.050)
X3 – Jenterp	0.001 (0.008)	–0.003 (0.006)	0.028 (0.040)	–0.035 (0.037)
X4 – Tota	0.000 (0.001)	–0.000 (0.001)	0.022** (0.010)	0.010 (0.006)
X5 – Jseeker	–0.060*** (0.008)	–0.044*** (0.008)	–0.020 (0.029)	–0.029** (0.017)
X6 – EUsubs	–0.001* (0.001)	–0.000 (0.000)	–0.010 (0.010)	–0.011 (0.022)
X7 – Vocdip	0.212*** (0.034)	0.090*** (0.028)	0.125 (0.088)	0.108 (0.074)
X8 – Hsdip	0.276*** (0.020)	0.137*** (0.018)	0.630*** (0.081)	0.302*** (0.074)
X9 – UCgrad	0.060*** (0.010)	0.041*** (0.009)	–	–
w1x_X1 – Wapop	–0.000 (0.001)	0.002** (0.001)	–0.005 (0.005)	0.006 (0.004)
w1x_X2 – Ienterp	–0.001 (0.003)	0.000 (0.003)	–0.020 (0.018)	0.037** (0.016)
w1x_X3 – Jenterp	0.004* (0.002)	0.008*** (0.002)	–0.012 (0.012)	–0.021 (0.013)
w1x_X4 – Tota	–0.002*** (0.000)	–0.002*** (0.000)	0.008 (0.005)	–0.001 (0.003)
w1x_X5 – Jseeker	–0.013*** (0.001)	0.004*** (0.002)	–0.020*** (0.006)	0.015** (0.006)
w1x_X6 – EUsubs	0.001*** (0.000)	0.000*** (0.000)	–0.004 (0.004)	–0.004 (0.003)
w1x_X7 – Vocdip	–0.041*** (0.006)	–0.021*** (0.006)	–0.062*** (0.017)	–0.084*** (0.026)
w1x_X8 – Hsdip	–0.014*** (0.005)	–0.024*** (0.004)	–0.067** (0.028)	–0.047 (0.035)
w1x_X9 – UCgrad	–0.006** (0.003)	–0.003 (0.003)	–	–
ϱ (rho)	0.027*** (0.03)	0.039*** (0.003)	0.061*** (0.012)	0.072*** (0.012)
σ (sigma)	0.177*** (0.004)	0.149*** (0.004)	0.086*** (0.005)	0.053*** (0.004)
Employment category variables	X	✓	X	✓
F-statistics for employment category variables	not relevant	***	not relevant	***
F-statistics for employment category lagged variables	not relevant	***	not relevant	***
R^2	*0.5569*	*0.7447*	*0.8385*	*0.9652*
Adjusted R^2	*0.5564*	*0.7416*	*0.8222*	*0.9562*
Log likelihood	*972.223*	*1,603.704*	*178.096*	*226.171*
Number of observations	*3,155*	*3,155*	*175*	*175*

Notes: *** p < 0.01; ** p < 0.05; * p < 0.1 (in parentheses robust standard error).
Source: Own editing based on GeoDa.

Interpreting the results of regression models

The more precise results of the spatial models are confirmed by the fact that the spatial autoregressive coefficients (ϱ and λ) at the municipal and district aggregation levels are all positive and significant (Table 2–5).

From the results presented in the previous chapter, we can conclude in two directions. On the one hand, we obtain an idea of the explanatory variables that, according to the different econometric models, truly affect the total domestic income per working-age population. On the other hand, we can also infer how the results of each model change with aggregation, in other words, at what level it is worth aggregating the data in similar regression studies.

In all four types of models (OLS, SAR, SEM, SDM), *the working-age population (Wapop)* is a significant factor at the municipal and district levels as well, where it is positively related to income at the municipal level and has a negative effect at the district level. At the municipal level, this confirms the common belief, also observed in scientific studies, that people live in larger settlements with a higher standard of living. At the district level, this effect is reversed, probably due to the large number of small settlements with limited opportunities, and aggregation will make their combined numbers significant.

The number of registered individual and joint enterprises (Ienterp and Jenterp) is not always significant at the municipality level, but when they are, they have an increasing effect on the income of the population of the area unit. At the district level, the presence of entrepreneurs is not significant in either case. This suggests that the positive impact of the number of enterprises at the local and municipal levels does not extend to districts. The variable *balance sheet total as a share of population (Tota)* is the opposite of *the number of registered individual and joint enterprises (Ienterp and Jenterp*). *Tota* is significant only at the district level. However, they also indicate that the size of the firm has a spatial effect, with the establishment of a large firm increasing district incomes. It is not necessarily that a large company has to establish more district incomes; many smaller companies can have such an effect, but it is more likely that a large company will have such an effect.

The number of jobseekers (Jseeker) was expected to have a negative impact on incomes, i.e., higher unemployment leads to lower incomes. This effect is significant in almost all cases.

The variable *for EU subsidies (EUsubs)* from 2015–2019 is significant and negative in most cases in the municipal- and district-level models. Our results are in line with those of Dedák (2015) and Madár (2016). The reason for the lack of impact is probably that EU subsidies do not necessarily translate into income for the businesses or people living in the municipality.

The effect of the three variables on educational attainment (*Vocdip, Hsdip, UCgrad*) is significant and positive for all models. *UCgrad* is not included in the district models due to multicollinearity. The results confirm that education has a positive effect on income (Tóth 2006, cited in Medgyesi 2006, Vastagh 2015).

The results of the employment category model are summarized in Table 6.

Table 6

Effect of occupational categories on income

Occupatiomal category[a)]	OLS_m	OLS_d	SAR_m	SAR_d	SEM_m	SEM_d	SDM_m	SDM_d
FEOR-08-0	+	+	+	+	+	+	+	+
FEOR-08-1	0	0	0	0	–	0	0	–
FEOR-08-2	0	0	0	0	0	0	0	0
FEOR-08-3	0	–	0	–	0	–	0	0
FEOR-08-4	+	+	+	+	+	+	+	+
FEOR-08-5	–	–	–	–	–	–	0	–
FEOR-08-6	–	0	–	0	–	0	0	0
FEOR-08-7	+	0	+	0	+	0	+	0
FEOR-08-8	+	0	+	0	+	0	+	0
FEOR-08-9	–	–	–	–	–	–	–	–

a) https://www.ksh.hu/feor_08_struktura_eng.pdf

Notes: +: positive significant impact; –: negative significant impact; 0: not significant impact.

Source: Own editing based on Stata.

For all estimates, the categories Armed forces occupations (FEOR-08-0) and Office and management (customer services) occupations (FEOR-08-04) have a significant and positive effect on both municipal and district incomes. (Elementary) occupations not requiring qualifications (FEOR-08-9) and Commercial and services occupations (FEOR-08-5) reduce incomes. Industry and construction industry occupations (FEOR-08-7) and Machine operators, assembly workers, drivers of vehicles (FEOR-08-8) increase earnings at the municipal level, while there is no statistically verifiable effect at the district level.

The results of the econometric models can also be compared in terms of how the number of significant explanatory variables and their coefficient signs vary at different levels of aggregation.

Table 7 shows that in all models, the number of significant explanatory variables decreases at the district level relative to the municipality level. Aggregation of the data leads to a kind of smoothing, which removes the differences observed for smaller spatial units and thus affects the results of the regression models. It is also observed that the higher the level of aggregation, the higher the standard error of the models. Higher standard errors introduce uncertainty into the models, which is one reason why the number of significant variables decreases. This is in line with the literature (e.g., Pietrzak 2014). These results indicate that spatial analyses should be conducted at the smallest spatial scale possible to obtain the most accurate model estimates.

If we compare model by model the explanatory variables that are significant at both the municipality and district level, we can also see that the sign of the variables changes with aggregation in only 4 cases (in all cases, *the number of working-age population* (*Wapop*) variable), so the sign of the coefficients is not significantly affected by aggregation.

Table 7

Count of significant explanatory variables

Model	Level of aggregation	Significant/all explanatory variables	% of significant variables
OLS	Municipal (X1–X9)	7/9	77.8
	Municipal (X1–X19)	13/19	68.4
	District (X1–X8)	3/8	37.5
	District (X1–X19) (without X9)	8/18	44.4
Spatial lag (SAR)	Municipal (X1–X9)	7/9	77.8
	Municipal (X1–X19)	13/19	68.4
	District (X1–X8)	5/8	62.5
	District (X1–X19) (without X9)	9/18	50.0
Spatial error (SEM)	Municipal (X1–X9)	7/9	77.8
	Municipal (X1–X19)	14/19	73.7
	District (X1–X8)	4/8	50.0
	District (X1–X19) (without X9)	10/18	55.6
Spatial Durbin (SDM)	Municipal (X1–X9)	7/9	77.8
	Municipal (X1–X19)	10/19	52.6
	District (X1–X8)	3/8	37.5
	District (X1–X19) (without X9)	7/18	38.9

Source: Own editing based on Stata.

The results are consistent with the phenomena described in the literature. Aggregating the data increases the goodness of fit of multivariate models, coefficients of determination and correlation coefficients due to the so-called smoothing effect, which is due to the reduction of variance within a single variable (Fotheringham–Wong 1991, Arbia 2012, cited in Zhou et al. 2022). When sociodemographic characteristics are aggregated at larger scales, a significant part of the within-area variation may be lost (Zhou et al. 2022). The standard error of the regression parameter also increases in models as a result of aggregation, so that the number of significant explanatory variables decreases at higher levels of aggregation (Pietrzak 2014). This is because the standard error depends in part on the number of observations used in the models (Fotheringham–Wong 1991).

Conclusions

In socioeconomic surveys, there are characteristics for which data are not available at the individual level (Fotheringham–Wong 1991), so regression coefficients have to be estimated using aggregate models. This, in turn, raises the problem of the modifiable area unit (MAUP), as evidenced by the results of our study above. The level of aggregation affects the results obtained from the analysis of spatially aggregated data (Nelson–Brewer 2015).

On the one hand, our study identifies important explanatory variables that affect the income situation of the population based on econometric models, and on the other hand, it highlights how aggregation affects the results of the former models.

The results of the analysis of the determinants of domestic income per working-age population show that the number of significant variables decreases as the level of aggregation increases. Higher levels of aggregation mask the variability and detail found in lower-level spatial units (Prouse et al. 2014).

Spatial autocorrelation typically decreases as aggregation increases (Qi–Wu 1996), and our studies used global and local Moran's I and Geary C to confirm this for domestic income per working-age resident, with a striking reduction in autocorrelation, especially at the county aggregation level.

Among the most important results of our econometric models, we find that *the working-age population (Wapop*) has an inverse effect on income at the municipal and district levels. The effect is positive at the municipal level but negative at the district level, probably because of the large number of small settlements with limited opportunities in each district and the aggregation of the population of these settlements that makes certain districts populous.

At the district level, the *presence of businesses* (Ienterp and Jenterp) is not significant in either case. Nor is it significant at the municipal level in all cases, but when it is, it increases income. Therefore, the effect at the municipal level does not extend to the districts.

The variable *balance sheet total as a share of population (Tota)* is significant only at the district level. This indicates that the size of firms has a spatial impact that goes beyond the municipal level, as the creation of a large firm increases incomes in the district. The presence of one large firm is not necessary to increase district incomes; the combined effect of several smaller firms can also lead to an increase in incomes.

Our results also show the unsurprising result that higher *educational attainment* (*Vocdip, Hsdip, UCgrad*) leads to higher income. This relationship is also observed at the municipality and district levels.

The main lesson of our study is that the municipality level is the most important level for the analysis of factors affecting income and the most useful level of aggregation for conducting spatial studies. This is in line with the literature that the smaller the geographical unit, the more homogeneous the characteristics of the population, so smaller geographical units provide better measures and stronger evidence of neighborhood effects (Prouse et al. 2014). The effect of district or county aggregation is that the poorest settlements may 'disappear' from the map, and it may appear that there are no poor settlements in the districts or counties where the richer settlements are found. This, in turn, can have implications for policy-making (Minot–Baulch 2005, cited in Pawitan–Steel 2009). To produce more accurate econometric models and scientific results from settlement-level studies, more data at the municipal level should be made available in the TeIR.

The importance of the municipal level has already been recognized for indicators that are usually available at the national or regional level. In the case of the HDI (Human Development Index), for example, several attempts have been made to calculate local human development and human resources at the municipal level in Hungary, for which researchers have adapted the original HDI indicator system (Malatyinszki 2008, Lipták 2017, Horváthné Kovács et al. 2017).

In our opinion, the district level is the highest acceptable level of aggregation for income analysis in Hungary, and aggregation above this level is not recommended due to the loss of information and inaccuracy of model estimates. It should be noted, however, that in the case of a district, we are talking about a delimitation where the boundaries are artificially drawn and do not necessarily define a homogeneous economic environment.

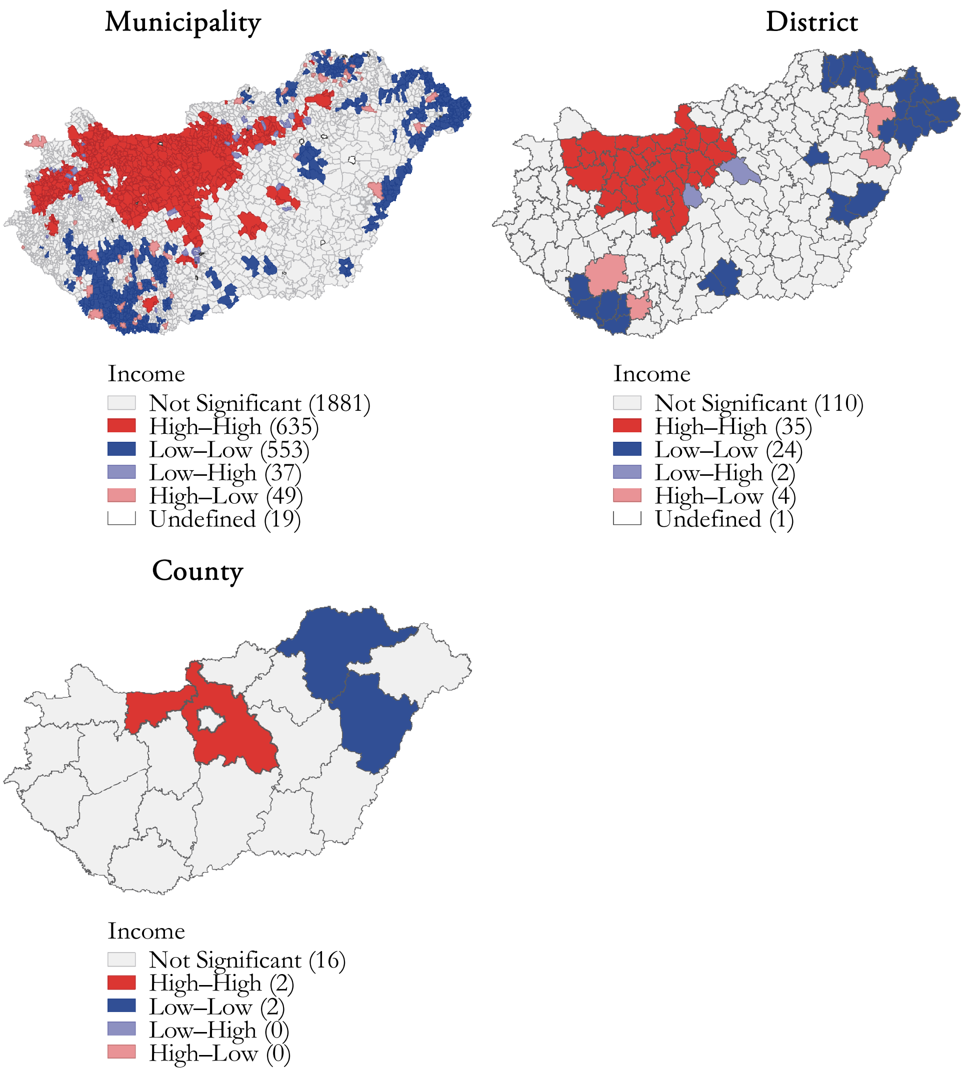
Acknowledgments

Supported by the ÚNKP-21-3-I-(II)-MATE-20 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

Appendix

Figure A1

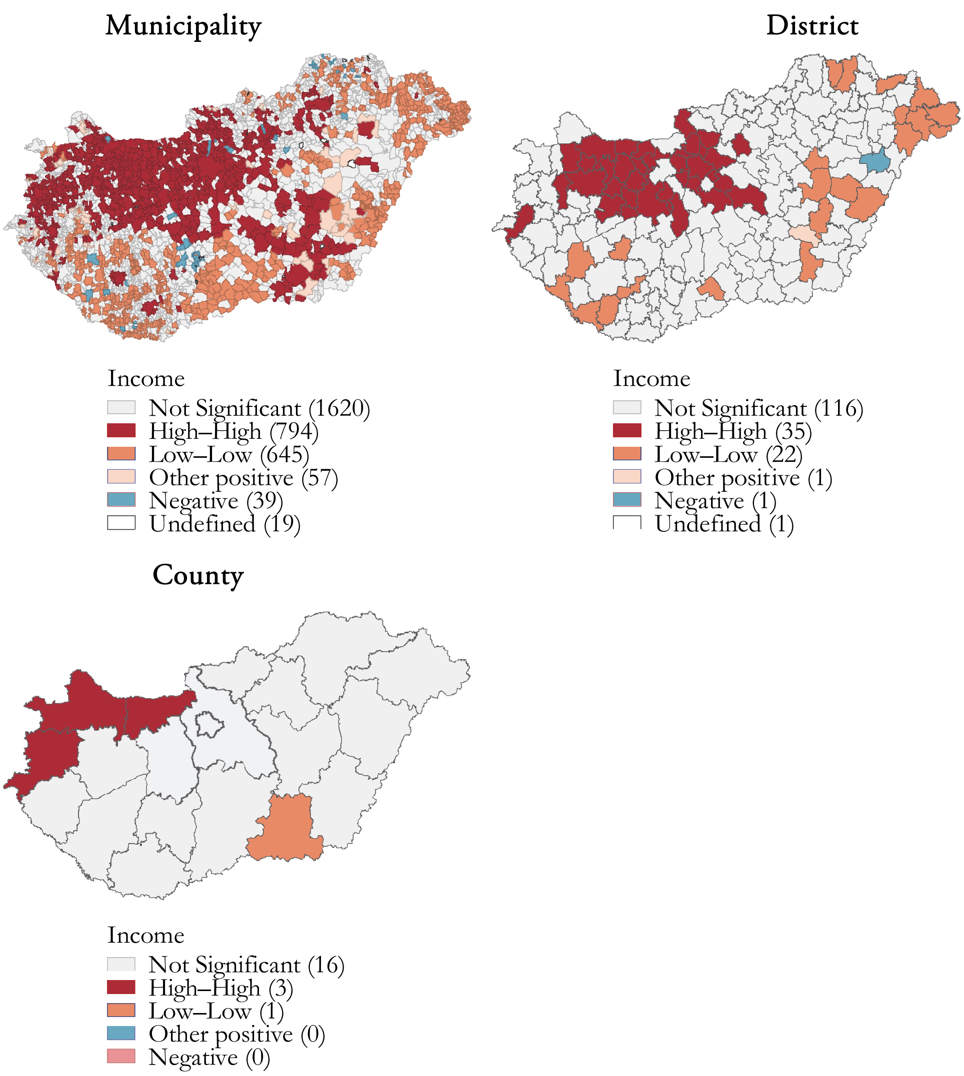
LISA cluster maps for Local Moran's I
(queen neighborhood, order of contiguity 1), 2019



Source: Own editing based on GeoDa.

Figure A2

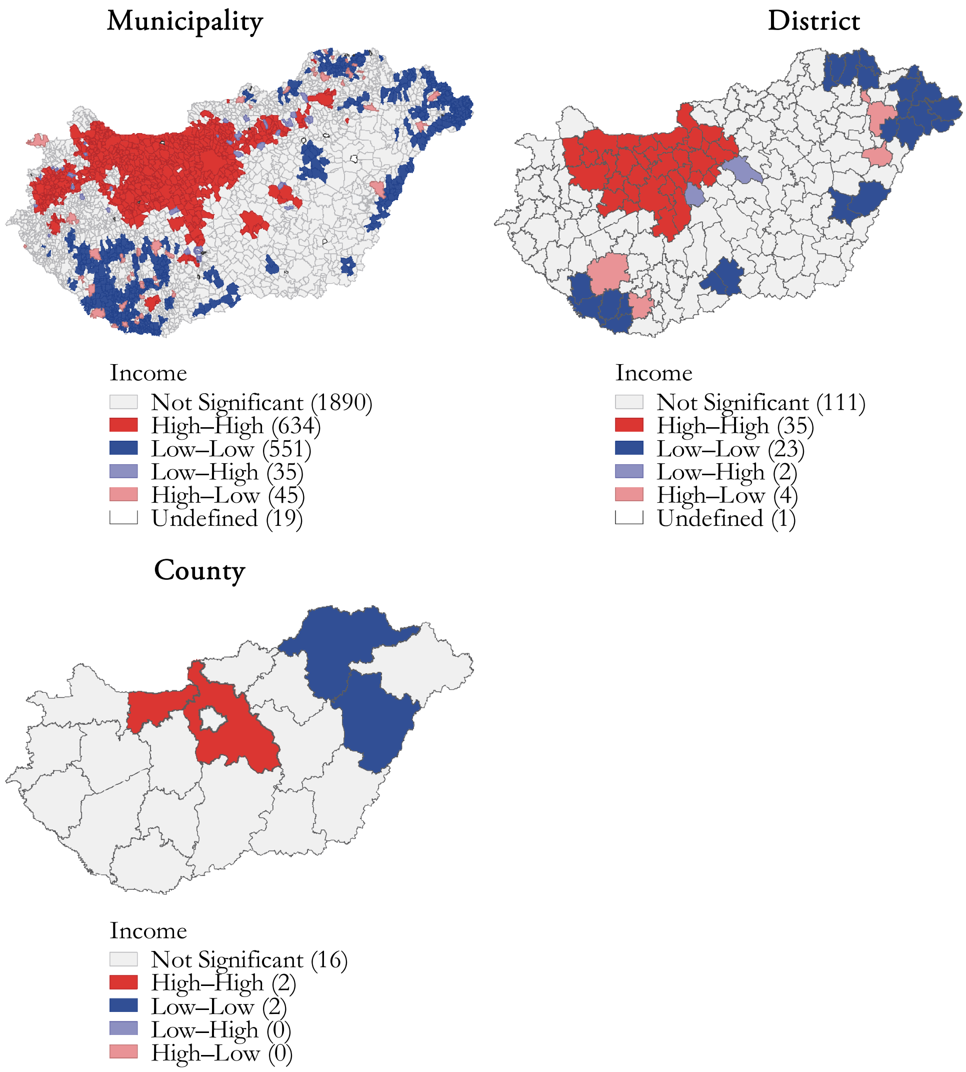
Cluster maps for Local Geary C (queen neighborhood, order of contiguity 1)



Source: Own editing based on GeoDa.

Figure A3

LISA cluster maps for Local Moran's I (rook neighborhood, order of contiguity 1), 2019

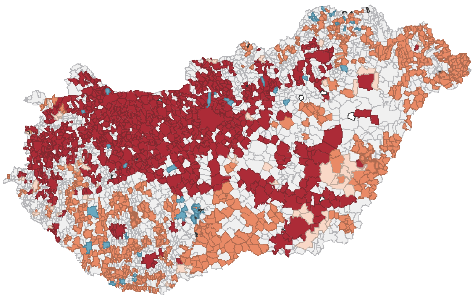


Source: Own editing based on GeoDa.

Figure A4

Cluster maps for Local Geary C (rook neighborhood, order of contiguity 1), 2019

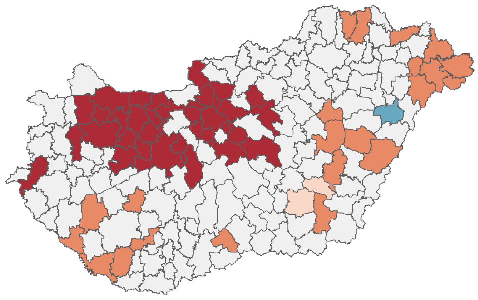
Municipality



Income

- Not Significant (1635)
- High–High (788)
- Low–Low (636)
- Other positive (52)
- Negative (44)
- Undefined (19)

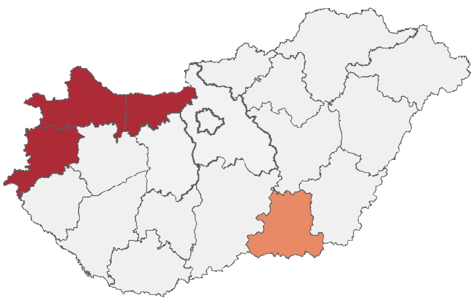
District



Income

- Not Significant (114)
- High–High (36)
- Low–Low (22)
- Other positive (2)
- Negative (1)
- Undefined (1)

County



Income

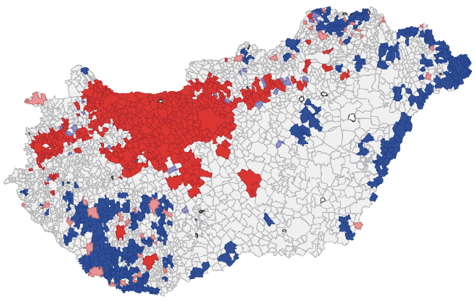
- Not Significant (16)
- High–High (3)
- Low–Low (1)
- Other positive (0)
- Negative (0)

Source: Own editing based on GeoDa.

Figure A5

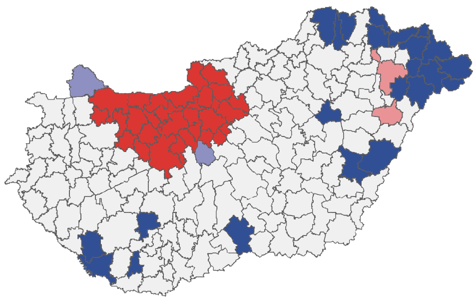
LISA cluster maps for Local Moran's I (k-nearest neighbors, k=4), 2019

Municipality



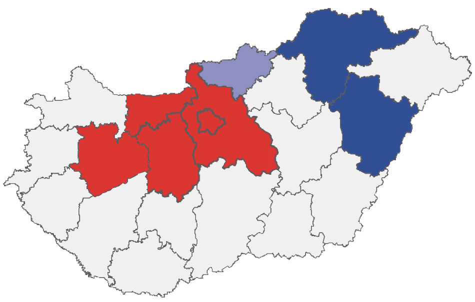
Income
Not Significant (2043)
High–High (530)
Low–Low (515)
Low–High (24)
High–Low (43)
Undefined (19)

District



Income
Not Significant (118)
High–High (30)
Low–Low (23)
Low–High (2)
High–Low (2)
Undefined (1)

County

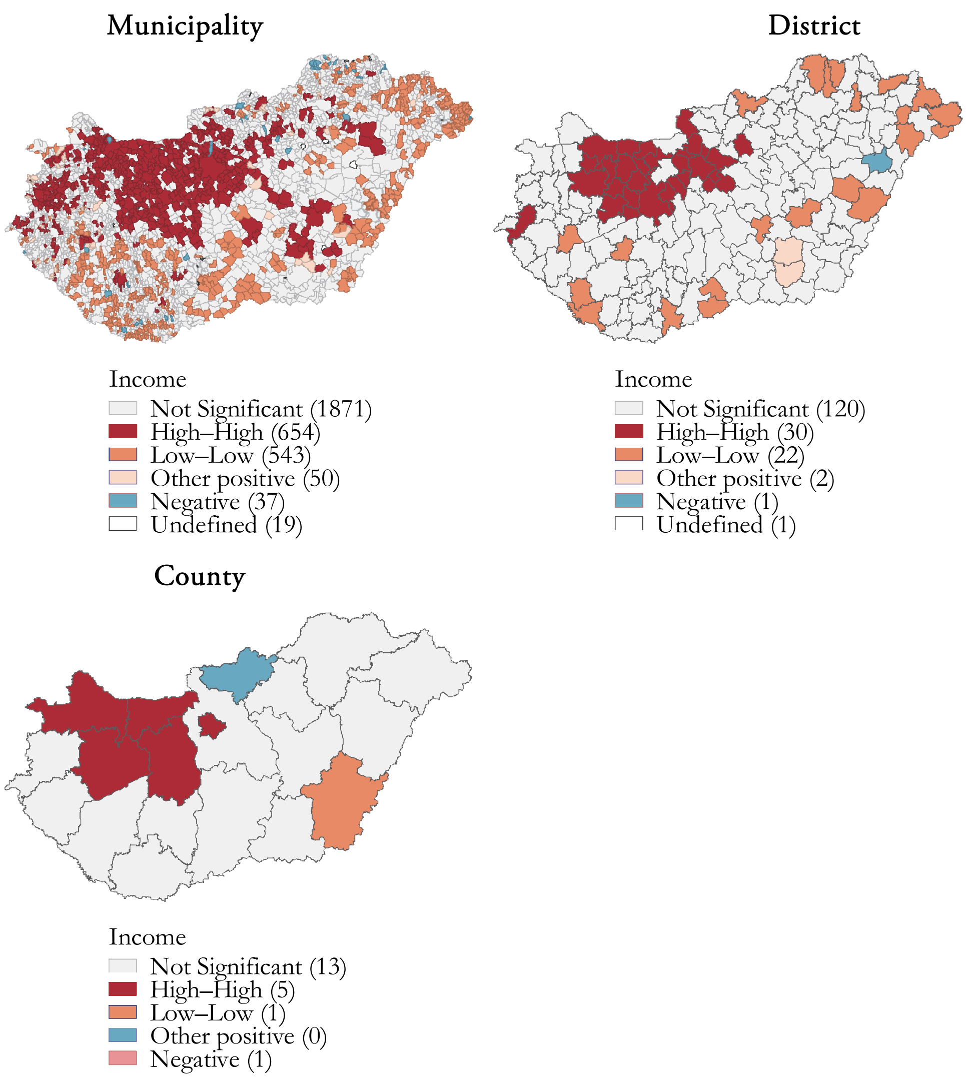


Income
Not Significant (12)
High–High (5)
Low–Low (2)
Low–High (1)
High–Low (0)

Source: Own editing based on GeoDa.

Figure A6

Cluster maps for Local Geary C (k-nearest neighbors, k=4), 2019



Source: Own editing based on GeoDa.

Table A1

List of variables

	Denomination
Y	Total domestic income per working age population (HUF/persons)
X1	Working-age permanent population (15–64 years) (31 December) (persons) *(Wapop)*
X2	Registered individual enterprises per 1,000 working age population (number/1,000 persons) *(Ienterp)*
X3	Registered joint enterprises per 1,000 working age population (number of enterprises per 1,000 persons) *(Jenterp)*
X4	Balance sheet total per 1,000 working age population (HUF per 1,000 persons) *(Tota)*
X5	Number of jobseekers per 1,000 working age population (persons/1,000 persons) *(Jseeker)*
X6	EU subsidy per 1,000 working age population 2015–2019 (in 2019 values) (HUF/1,000 persons) *(EUsubs)*
X7	Number of persons per 1,000 inhabitants aged 7 to X with vocational diploma, without high school graduation certificate (persons/1,000 persons) (year 2011) *(Vocdip)*
X8	Number of persons per 1,000 inhabitants aged 7 to X with high school diploma (persons/1,000 persons) (year 2011) *(Hsdip)*
X9	Number of persons per 1,000 inhabitants aged 7 to X with university, college, etc., graduation (persons/1,000 persons) (year 2011) *(UCgrad)*
X10	Employed in FEOR-08-0 (Armed forces occupations) per 1,000 persons (persons/1,000 persons)
X11	Employed in FEOR-08-1 (Managers) per 1,000 employees (persons/1,000 persons)
X12	Employed in FEOR-08-2 (Professionals) per 1000 persons employed (persons/1,000 persons)
X13	Employed in FEOR-08-3 (Technicians and associate professionals) per 1,000 persons (persons/1,000 persons)
X14	Employed in FEOR-08-4 (Office and management [customer services] occupations) per 1,000 employees (persons/1,000 persons)
X15	Employed in FEOR-08-5 (Commercial and services occupations) per 1,000 employees (persons/1,000 persons)
X16	Employed in FEOR-08-6 (Agricultural and forestry occupations) per 1,000 employees (persons/1,000 persons)
X17	Employed in FEOR-08-7 (Industry and construction industry occupations) per 1,000 persons employed (persons/1,000 persons)
X18	Employed in FEOR-08-8 (Machine operators, assembly workers, drivers of vehicles) per 1,000 persons employed (persons/1,000 persons)
X19	Employed in FEOR-08-9 ([Elementary] occupations not requiring qualifications) per 1,000 persons (persons/1,000 persons)

Table A2

Descriptive statistics of variables

Variables	Mean	Standard deviation	Minimum	Maximum
	Municiality level			
Y	1,846,385.481	528,204.900	298,881.102	4,482,484.952
X1	2,078.426	20,170.737	9	1,085,390
X2	60.725	30.141	0	442.857
X3	43.450	149.157	0	7,342.857
X4	6,331,764,145.018	51,710,631,683.503	0	2,684,989,724,050.630
X5	56.004	44.443	0	369.718
X6	701,478,728.332	5,646,400,799.856	0	269,354,744,014.028
X7	237.315	47.452	0	545.455
X8	194.327	64.460	0	416.732
X9	69.915	47.503	0	414.964
X10	6.174	7.371	0	111.111
X11	44.519	32.037	0	375
X12	73.422	45.602	0	340.909
X13	116.260	41.709	0	285.714
X14	54.748	24.658	0	250
X15	107.828	35.826	0	312.500
X16	13.473	17.502	0	272.727
X17	104.157	37.734	0	305.556
X18	145.297	62.390	0	453.586
X19	334.122	136.200	53,570	969.510

(Table continues on the next page.)

(Continued.)

Variables	Mean	Standard deviation	Minimum	Maximum
District level				
Y	2,061,620.446	416,440.549	1,358,791.735	3,260,350.877
X1	37,471.057	84,167.073	5,585.000	1,085,390.000
X2	70.446	17.951	27.231	130.595
X3	45.762	26.007	13.843	193.008
X4	9,889,128,201.612	16,810,015,166.889	693,996,083.789	167,028,077,859.571
X5	48.181	28.567	6.680	133.783
X6	809,055,032.935	921,110,208.597	39,361,361.369	9,076,885,859.161
X7	222.354	28.125	115.740	294.888
X8	236.694	43.547	130.102	348.342
X9	103.143	44.981	43.481	288.759
X10	7.259	4.149	1.580	27.950
X11	50.913	20.568	17.326	139.693
X12	102.884	38.251	43.793	247.986
X13	135.984	23.256	84.196	198.909
X14	64.861	14.218	38.827	102.502
X15	116.398	17.994	65.894	194.217
X16	9.425	6.246	1.591	45.873
X17	99.300	20.024	45.295	142.604
X18	131.131	44.762	40.747	295.329
X19	281.845	80.068	130.250	508.590

Table A3

Pairwise correlations of variables

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
	Municiality level																		
X1	1.000																		
X2	0.249	1.000																	
X3	0.193	0.503	1.000																
X4	0.346	0.345	0.680	1.000															
X5	–0.060	–0.405	–0.341	–0.173	1.000														
X6	0.531	0.149	0.181	0.311	0.041	1.000													
X7	–0.051	0.203	0.027	0.089	–0.168	–0.021	1.000												
X8	0.422	0.577	0.439	0.368	–0.466	0.260	0.248	1.000											
X9	0.453	0.563	0.478	0.352	–0.437	0.279	0.056	0.736	1.000										
X10	0.332	0.179	0.116	0.137	–0.017	0.238	0.025	0.277	0.268	1.000									
X11	0.193	0.429	0.458	0.262	–0.354	0.110	–0.010	0.460	0.493	0.125	1.000								
X12	0.435	0.511	0.387	0.307	–0.326	0.281	0.105	0.655	0.720	0.242	0.400	1.000							
X13	0.351	0.450	0.345	0.329	–0.281	0.240	0.254	0.599	0.550	0.217	0.262	0.496	1.000						
X14	0.403	0.361	0.332	0.346	–0.209	0.286	0.150	0.537	0.444	0.239	0.316	0.416	0.406	1.000					
X15	0.302	0.271	0.169	0.206	–0.154	0.192	0.215	0.427	0.363	0.140	0.208	0.366	0.309	0.337	1.000				
X16	–0.069	–0.004	–0.079	–0.004	–0.020	–0.018	0.121	–0.051	–0.048	–0.048	–0.093	–0.048	–0.029	–0.051	0.029	1.000			
X17	0.149	0.138	–0.028	0.104	–0.023	0.097	0.451	0.190	0.064	0.052	–0.086	0.122	0.156	0.175	0.133	0.023	1.000		
X18	0.011	0.077	–0.058	0.086	–0.157	–0.016	0.426	0.158	–0.016	–0.048	–0.101	0.044	0.095	0.058	0.028	0.078	0.251	1.000	
X19	–0.312	–0.579	–0.470	–0.288	0.635	–0.127	–0.180	–0.741	–0.693	–0.199	–0.536	–0.628	–0.539	–0.472	–0.361	0.053	–0.170	–0.219	1.000

(Table continues on the next page.)

(Continued.)

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19
	District level																		
X1	1.000																		
X2	0.471	1.000																	
X3	0.618	0.801	1.000																
X4	0.507	0.544	0.673	1.000															
X5	–0.408	–0.738	–0.661	–0.472	1.000														
X6	0.016	–0.040	–0.195	–0.016	0.283	1.000													
X7	–0.553	–0.161	–0.435	–0.238	0.043	–0.014	1.000												
X8	0.662	0.763	0.783	0.603	–0.670	–0.070	–0.373	1.000											
X9	0.717	0.813	0.872	0.629	–0.626	–0.011	–0.549	0.894	1.000										
X10	0.206	0.020	0.104	–0.040	0.115	0.113	–0.235	0.148	0.217	1.000									
X11	0.575	0.855	0.922	0.627	–0.656	–0.113	–0.362	0.829	0.900	0.085	1.000								
X12	0.699	0.793	0.838	0.592	–0.611	–0.017	–0.558	0.876	0.975	0.195	0.887	1.000							
X13	0.594	0.747	0.772	0.536	–0.574	–0.084	–0.419	0.870	0.874	0.130	0.830	0.886	1.000						
X14	0.598	0.766	0.880	0.601	–0.696	–0.198	–0.334	0.817	0.831	0.138	0.866	0.813	0.782	1.000					
X15	0.232	0.569	0.454	0.155	–0.406	–0.010	–0.050	0.451	0.411	0.071	0.527	0.429	0.466	0.525	1.000				
X16	–0.592	–0.275	–0.500	–0.331	0.288	0.158	0.489	–0.604	–0.575	–0.140	–0.510	–0.575	–0.526	–0.477	–0.075	1.000			
X17	–0.430	–0.314	–0.529	–0.290	0.176	–0.029	0.652	–0.359	–0.517	–0.199	–0.475	–0.502	–0.430	–0.402	–0.356	0.297	1.000		
X18	–0.318	–0.243	–0.394	–0.102	0.022	–0.023	0.647	–0.235	–0.458	–0.214	–0.422	–0.475	–0.424	–0.371	–0.366	0.183	0.540	1.000	
X19	–0.591	–0.840	–0.824	–0.624	0.807	0.142	0.254	–0.879	–0.866	–0.018	–0.853	–0.852	–0.805	–0.809	–0.398	0.512	0.288	0.092	1.000

Table A4

Lagrange multiplier diagnostics

Test	Moran's I/DF	Value
Municipal level (N = 3155) (X1–X9)		
Moran's I (error)	0.4189	39.297***
Lagrange Multiplier (lag)	1	1,061.431***
Robust LM (lag)	1	84.311***
Lagrange Multiplier (error)	**1**	**1,528.403*** **
Robust LM (error)	**1**	**551.284*** **
Lagrange Multiplier (SARMA)	2	1,612.715***
District level (N = 175) (X1–X8)		
Moran's I (error)	0.4283	9.959***
Lagrange Multiplier (lag)	1	77.754***
Robust LM (lag)	1	14.660***
Lagrange Multiplier (error)	**1**	**82.407*** **
Robust LM (error)	**1**	**19.313*** **
Lagrange Multiplier (SARMA)	2	97.066***
Municipal level (N = 3155) (X1–X19)		
Moran's I (error)	0.3360	31.616***
Lagrange Multiplier (lag)	1	612.248***
Robust LM (lag)	1	62.988***
Lagrange Multiplier (error)	**1**	**983.168*** **
Robust LM (error)	**1**	**433.908*** **
Lagrange Multiplier (SARMA)	2	1,046.155***
District level (N = 175) (X1–X19, without X9)		
Moran's I (error)	0.3184	8,139***
Lagrange Multiplier (lag)	1	43,363***
Robust LM (lag)	1	13,175***
Lagrange Multiplier (error)	**1**	**45,550*** **
Robust LM (error)	**1**	**15,362*** **
Lagrange Multiplier (SARMA)	2	58,724***

Notes: *** p<0.01; ** p<0.05; * p<0.1.
Source: Own editing based on GeoDa.

REFERENCES

ALHENDI, O.–DÁVID, L. D.–FODOR, G.–ADOL, G. F. C.–BALOGH, P. (2021): The impact of language and quality education on regional and economic development: A study of 99 countries *Regional Statistics* 11 (1): 42–57. https://doi.org/10.15196/RS110101

AMRHEIN, C. G.–FLOWERDEW, R. (1992): The effect of data aggregation on a poisson regression model of Canadian migration *Environment and Planning A: Economy and Space* 24 (10): 1381–1391. https://doi.org/10.1068/a241381

ANSELIN, L. (1995): Local indicators of spatial association – LISA *Geographical Analysis* 27 (2): 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

ANSELIN, L. (2005): *Exploring spatial data with GeoDa: A workbook* University of Illinois, Urbana-Champaign, Urbana.

ARBIA, G. (2012): *Spatial data configuration in statistical analysis of regional economic and related problems* 14. Springer Science & Business Media.

ARBIA, G.–PETRARCA, F. (2011): Effects of MAUP on spatial econometric models *Letters in Spatial and Resource Sciences* 4 (3): 173–185. https://doi.org/10.1007/s12076-011-0065-9

ARITENANG, A. F. (2022): The impact of urban characteristics on the spread of Covid-19 in 2020: The case of Java Island cities, Indonesia *Regional Statistics* 12 (3): 3–17. https://doi.org/10.15196/RS120301

ATKINSON, R. (2019): The small towns conundrum: What do we do about them? *Regional Statistics* 9 (2): 3–19. https://doi.org/10.15196/RS090201

BARRO, R. J.–SALA-I-MARTIN, X. (1992): Convergence *Journal of Political Economy* 100 (2): 223–251. https://doi.org/10.1086/261816

BELL, S.–WILSON, K.–BISSONETTE, L.–SHAH, T. (2013): Access to primary health care: Does neighborhood of residence matter? *Annals of the Association of American Geographers* 103 (1): 85–105. https://doi.org/10.1080/00045608.2012.685050

BLALOCK, H. M. (1964): *Causal inferences in nonexperimental research* University of North Carolina Press, Chapel Hill.

BRUNSDON, C.–FOTHERINGHAM, A. S.–CHARLTON, M. (2002): Geographically weighted summary statistics – A framework for localized exploratory data analysis *Computers, Environment and Urban Systems* 26 (6): 501–524. https://doi.org/10.1016/S0198-9715(01)00009-6

CASELLI, F.–ESQUIVEL, G.–LEFORT, F. (1996): Reopening the convergence debate: A new look at cross-country growth empirics *Journal of Economic Growth* 1 (3): 363–389. https://doi.org/10.1007/BF00141044

CLARK, W. A. V.–AVERY, K. L. (1976): The effects of data aggregation in statistical analysis *Geographical Analysis* 8 (4): 428–438. https://doi.org/10.1111/j.1538-4632.1976.tb00549.x

CLIFF, A.–ORD, J. (1981): *Spatial processes, models and applications* Pion, London.

CSIZMADIA, A.–BAREITH, T. (2022): Somogy megye lakosságának jövedelmi helyzete, 2012–2019 *Területi Statisztika* 62 (3): 348–373. https://doi.org/10.15196/TS620304

DRUKKER, D. M.–PRUCHA, I. R.–RACIBORSKI, R. (2013): A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables *The Stata Journal* 13 (2): 287–301. https://doi.org/10.1177/1536867X1301300203

DURAN, H. E.–KARAHASAN, B. C. (2022): Heterogenous responses to monetary policy regimes: A regional analysis for Turkey, 2009–2019 *Regional Statistics* 12 (4): 56–91. https://doi.org/10.15196/RS120403

DUSEK, T. (2001): *A területi lehatárolások statisztikai következményei* Magyar Földrajzi Konferencia tudományos közleményei (CD). Szeged.

DUSEK, T. (2003): A statisztikai adatok területi aggregálásának kérdései *Statisztikai Szemle* 81 (2): 127–147.

DUSEK, T. (2004): Spatially aggregated data and variables in empirical analysis and model building for economics *Cybergeo: European Journal of Geography* https://doi.org/10.4000/cybergeo.2654

FERTŐ, I.–VARGA, Á. (2015): Az európai uniós támogatások hatása a kistérségek helyzetére *Tér és Társadalom* 29 (1): 116–131. https://doi.org/10.17649/TET.29.1.2667

FITRIANI, R.–PUSDIKTASARI, Z. F.–DIARTHO, H. C. (2021): Growth interdependence in the presence of spatial outliers: Implementation of an average difference algorithm on East Java regional economic growth, 2011–2016 *Regional Statistics* 11 (3): 119–132. https://doi.org/10.15196/RS110306

FLOWERDEW, R. (2011): How serious is the modifiable areal unit problem for analysis of English census data? *Population Trends* 145 (1): 106–118. https://doi.org/10.1057/pt.2011.20

FOTHERINGHAM, A. S.–WONG, D. W. S. (1991): The modifiable areal unit problem. In: Multivariate statistical analysis E*nvironment and Planning A: Economy and Space* 23 (11): 1025–1044. https://doi.org/10.1068/a231025

GÁL, V.–GÁSPÁR, K.–PARÁDI-DOLGOS, A. (2014): Regional differences in the capital structure of Hungarian SMEs. Acta Universitatis Sapientiae *Economics and Business* 2 (1): 21–36. https://doi.org/10.2478/auseb-2014-0008

GEARY, R. C. (1954): The contiguity ratio and statistical mapping *The Incorporated Statistician* 5 (3): 115–127., 129–146. https://doi.org/10.2307/2986645

GEHLKE, C. E.–BIEHL, K. (1934): Certain effects of grouping upon the size of the correlation coefficient in census tract material *Journal of the American Statistical Association* 29: 169–170. https://doi.org/10.1080/01621459.1934.10506247

GLAESER, E. L.–SCHEINKMAN, J. A.–SHLEIFER, A. (1995): Economic growth in a cross-section of cities *Journal of Monetary Economics* 36 (1): 117–143. https://doi.org/10.1016/0304-3932(95)01206-2

HAJDÚ, D.–KONCZ, G. (2022): Employment data of participants in supported adult training for jobseekers and their territorial pattern in Hungary, 2010–2020 *Regional Statistics* 12 (2): 117–148. https://doi.org/10.15196/RS120205

HORVÁTHNÉ KOVÁCS, B.–BARNA, R.–TITOV, A.–NAGY, M. Z. (2017): A humán tőke indexe a Dél-dunántúli régióban *Közép-Európai Közlemények* 10 (3): 55–72.

ISLAM, N. (1995): Growth Empirics: A panel data approach *Quarterly Journal of Economics* 110 (4): 1127–1170. https://doi.org/10.2307/2946651

JELINSKI, D. E.–WU, J. (1996): The modifiable areal unit problem and implications for landscape ecology *Landscape Ecology* 11 (3): 129–140. http://dx.doi.org/10.1007/BF02447512

KALOGIROU, S.–HATZICHRISTOS, T. (2007): A spatial modelling framework for income estimation *Spatial Economic Analysis* 2 (3): 297–316. https://doi.org/10.1080/17421770701576921

KILGARRIFF, P.–CHARLTON, M. (2020): *A spatial analysis of disposable income in Ireland: A GWR approach* https://doi.org/10.48550/arXiv.2008.11720

KOLBER, K.–PARÁDI-DOLGOS, A.–KOROSECZNÉ PAVLIN, R. (2019): A gazdasági társaságok adóterhét befolyásoló tényezők vizsgálata *Economica* 10 (3–4): 1–7. https://doi.org/10.47282/ECONOMICA/2019/10/3-4/4689

LADOS, G.–HEGEDŰS, G. (2019): Return migration and identity change: A Hungarian case study *Regional Statistics* 9 (1): 150–167. https://doi.org/10.15196/RS090109

LEE, G.–CHO, D.–KIM, K. (2015): The modifiable areal unit problem in hedonic house-price models *Urban Geography* 37 (2): 223–245. https://doi.org/10.1080/02723638.2015.1057397

LESAGE, J.–PACE, K. R. (2009): *Introduction to spatial econometrics* Chapman and Hall/CRC, London.

LIPTÁK, K. (2017): Települési humán potenciál vizsgálata Magyarországon *Észak-magyarországi Stratégiai Füzetek* 14 (2) 55–65.

MALATYINSZKI, SZ. (2008): *Az emberi erőforrás mérése települési szinten Békés megyében* PhD-értekezés. https://doi.org/10.13140/RG.2.2.23778.48322

MANKIW, G.–ROMER, D.–WEIL, N. (1992): A contribution to the empirics of economic growth *Quarterly Journal of Economics* 107 (2): 407–437. https://doi.org/10.2307/2118477

MCCANN, P. (2020): Perceptions of regional inequality and the geography of discontent: Insights from the UK *Regional Studies* 54 (2): 256–267. https://doi.org/10.1080/00343404.2019.1619928

MEDGYESI, M. (2006): Az oktatás társadalmi és gazdasági környezete. In: HALÁSZ, G.–LANNERT, J.(eds.): *Jelentés a magyar közoktatásról 2006* pp. 29–53., Országos Közoktatási Intézet, Budapest.

MINOT, N.–BAULCH, B. (2005): Poverty mapping with aggregate census data: What is the loss in precision? *Review of Development Economics* 9 (1): 5–24. https://doi.org/10.1111/j.1467-9361.2005.00261.x

MOLNÁR, T.–ILK, B. F. (2010): A települések fejlettségének alakulása a Dél-Dunántúli Régióban *Acta Scientiarum Socialium* 32 (1): 29–38.

MORAN, P. (1948): The interpretation of statistical maps *Journal of The Royal Statistical Society B* 10 (2): 243–251. https://doi.org/10.1111/j.2517-6161.1948.tb00012.x

NELSON, J. K.–BREWER, C. A. (2015): Evaluating data stability in aggregation structures across spatial scales: Revisiting the modifiable areal unit problem *Cartography and Geographic Information Science* 44 (1): 35–50. http://dx.doi.org/10.1080/15230406.2015.1093431

NEMES NAGY, J. (1990): Területi egyenlőtlenségek dimenziói – Adalékok egy „kvázi-elmélethez" *Tér és Társadalom* 4 (2): 15–30. https://doi.org/10.17649/TET.4.2.171

OPENSHAW, S. (1978): An empirical study of some zone-design criteria *Environment and Planning A: Economy and Space* 10 (7): 781–794. https://doi.org/10.1068/a100781

OPENSHAW, S. (1984): *The modifiable areal unit problem* CATMOG 38. GeoBooks, Norwich, England.

OPENSHAW, S.–TAYLOR, P. (1979): A million or so correlation coefficients: Three experiments On the modifiable areal unit problem. In: WRIGLEY, N. (ed.): *Statistical applications in the spatial sciences* pp. 127–144., Pion. London.

PIETRZAK, M. B. (2014): The modifiable areal unit problem – Analysis of correlation and regression. Equilibrium *Quarterly Journal of Economics and Economic Policy* 9 (3): 113–131. http://dx.doi.org/10.12775/EQUIL.2014.028

PROUSE, V.–RAMOS, H.–GRANT, J.–L. RADICE, M. (2014): How and when scale matters: The modifiable areal unit problem and income inequality in Halifax *Canadian Journal of Urban Research* 23 (1): 61–82.

QI, Y.–WU, J. (1996): Effects of changing spatial resolution on the results of landscape pattern analysis using spatial autocorrelation indices *Landscape Ecology* 11 (1): 39–49. https://doi.org/10.1007/BF02087112

PÉNZES, J. (2011): A jövedelmi szempontból elmaradott települések területi átrendeződése a rendszerváltozás után *Földrajzi Közlemények* 135 (1): 59–69.

RECHNITZER, J.–BERKES, J.–FILEP, B. (2019): The most important city development initiatives of Hungary *Regional Statistics* 9 (2): 20–44. https://doi.org/10.15196/RS090204

REY, S. J.–MONTOURI, B. D. (1999): US regional income convergence: A spatial econometric perspective *Regional Studies* 33 (2): 143–156. https://doi.org/10.1080/00343409950122945

ROBINSON, W. S. (1950): Ecological correlations and the behavior of individuals *American Sociological Review* 15 (3): 351–357. https://doi.org/10.2307/2087176

SAIB, M-S.–CAUDEVILLE, J.–CARRE, F.–GANRY, O.–TRUGEON, A.–CICOLELLA, A. (2014): Spatial relationship quantification between environmental, socioeconomic and health data at different geographic levels *International Journal of Environmental Research and Public Health* 11 (4): 3765–3786. https://doi.org/10.3390/ijerph110403765

SHAH, T. I.–ASPEN, L.–BELL, S. (2014): Assessment of choice of units of analysis for studying associations between geographic accessibility to PHC services and sociodemographic factors *Spatial Knowledge and Information – Canada 2*.

SZENDI, D. (2017): Települési jövedelmek térbeli kölcsönhatása: felzárkózás vs. tartós különbségek? In: GELEI, A. (szerk.): *Sokszínű gazdálkodástudomány* pp. 141–174., MTA IX. Osztály, Gazdálkodástudományi Bizottság. Budapest.

TAGASHIRA, N.–OKABE, A. (2002): The modifiable areal unit problem in a regression model whose independent variable is a distance from a predetermined point *Geographical Analysis* 34 (1): 1–20. https://doi.org/10.1111/j.1538-4632.2002.tb01072.x

TÓTH, G. (2003): Területi autokorrelációs vizsgálat a Local Moran I módszerével *Tér és Társadalom* 17 (4): 39–49. https://doi.org/10.17649/TET.17.4.914

TÓTH, I. GY. (2006): Jövedelemeloszlás. In: TÓTH, I. GY.–SZIVÓS, P. (eds.): *Feketén, fehéren* pp. 11–32., TÁRKI Monitor Jelentések 2005. TÁRKI, Budapest.

TSIONAS, M. (2019): *Panel data eEconometrics – Empirical applications* Academic Press, London.

TURCZAK, A.–ZWIECH, P. (2014): Variability of household disposable income per capita by types of residence in Poland *Statistics in transition New Series* 15 (4): 573–590.

VARGA, A. (2002): Térökonometria *Statisztikai Szemle* 80 (4): 354–370.

VASTAGH, Z (2015): Rétegződés és iskolázottság. In: *A társadalom rétegződése. 2011. évi népszámlálás* pp. 55–69., 15. Központi Statisztikai Hivatal, Budapest.

WANG, Y.–DI, Q. (2020): Modifiable areal unit problem and environmental factors of Covid-19 outbreak *Science of The Total Environment* 740: 139984. https://doi.org/10.1016/j.scitotenv.2020.139984

WONG, D.–LASUS, H.–FALK, R. (1999): Exploring the variability of segregation indexes with scale and zonal systems: An analysis of thirty US cities *Environment and Planning A: Economy and Space* 31 (3): 507–522. https://doi.org/10.1068/a310507

YULE, G. U.–KENDALL, M. G. (1964): *Bevezetés a statisztika elméletébe* Közgazdasági és Jogi Könyvkiadó, Budapest.

ZHOU, X.–SUN, C.–NIU, X.–SHI, C. (2022): The modifiable areal unit problem in the relationship between jobs-housing balance and commuting distance through big and traditional data *Travel Behaviour and Society* 26: 270–278. https://doi.org/10.1016/j.tbs.2021.11.001

INTERNET SOURCES

CHASCO, C.–RODRÍGUEZ, I.–OTERO, J. (2008): Modeling spatial variations in household disposable income with geographically weighted regression *MPRA Paper (9581)*. https://mpra.ub.uni-muenchen.de/9581/1/MPRA_paper_9581.pdf (downloaded: May 2022)

DEDÁK, I. (2015): Mégis, hogyan néznénk ki EU-támogatások nélkül? *Portfolio.hu* (2015. november 27.) https://www.portfolio.hu/gazdasag/20151127/megis-hogyan-neznenk-ki-eu-tamogatasok-nelkul-223003 (downloaded: May 2022)

LŐCSEI, H. (2004): A *vidéki agglomerációk fejlődési pályája. Műhelytanulmányok MT-DP 2004/24*. Magyar Tudományos Akadémia Közgazdaságtudományi Intézet. Budapest. https://www.econstor.eu/bitstream/10419/108093/1/MTDP0424.pdf (downloaded: May 2022)

MADÁR, I. (2016): Így festenénk az EU-támogatások nélkül *Portfolio.hu* (2016. április 14.) https://www.portfolio.hu/gazdasag/20160414/igy-festenenk-az-eu-tamogatasok-nelkul-230247 (downloaded: May 2022)

MANLEY, D. (2006): *The modifiable areal unit phenomenon: An investigation into the scale effect using UK census data* PhD Thesis. University of St. Andrews. https://research-repository.st-andrews.ac.uk/handle/10023/465 (downloaded: May 2022)

NTHIWA, A. (2011): *Modeling scale and effects of the modifiable areal unit problem on multiple deprivations in Istanbul, Turkey* Master's Thesis: Faculty of Geo-Information Science and Earth Observation of the University of Twente, Enschede, The Netherlands. http://www.itc.nl/library/papers_2011/msc/upm/nthiwa.pdf (downloaded: May 2022)

PAWITAN, G.–STEEL, D. G. (2009): Exploring the MAUP from a spatial perspective, centre for statistical and survey methodology, University of Wollongong *Working Paper 20–09*. https://ro.uow.edu.au/cssmwp/40 (downloaded: May 2022)

SIK, E. (2014): Adalékok a hátrányos helyzetű kistelepülés társadalmi sajátosságainak megismeréséhez T*ÁRKI Regionális műhelytanulmányok 2014/3*. https://www.tarki.hu/hu/publications/WP/regionalis/201403_regionalis_muhely.pdf (downloaded: May 2022)