STUDIES

Quantifying worldwide economic distress

Fernando Delbianco
Universidad Nacional del Sur, Argentina
E-mail: fernando.delbianco@uns.edu.ar

Andrés Fioriti
Universidad Nacional del Sur, Argentina
E-mail: andres.fioriti@uns.edu.ar

Fernando Tohmé
Universidad Nacional del Sur, Argentina
E-mail: ftohme@criba.edu.ar

This paper proposes a new measure for worldwide economic distress, which can be described as the proportion of the gross domestic product (GDP) affected by country-level financial crises. It proves that this measure has the desirable properties of fair and representative indices. It also adequately justifies the beliefs of some economists that financial crises have negative effects on economic performance.

Keywords:
economic crisis,
growth,
index

Introduction

Economic crises are disruptive phenomena that have negative effects on the performance of economies. In contemporary market systems, these crises are mostly of a financial nature, although their origins can often be found in the real sector.[1] Financial channels facilitate the propagation of crises from one country to another, despite policies and other efforts to avoid such events (Reinhart–Rogoff 2009). Therefore, it is of interest for both academic and policy-making reasons to have ways of assessing the impact of these crises. This has been extensively studied at the country level. However, we now pose the question: to what extent can the *world* be said to be in a state of financial distress? To answer this question, a way of measuring the states of world crises is required.

The objective of this study is to answer this question. We present very simple definitions of two *world distress* (WD) indices. Indices are widely used in economics to explain a wide range of situations (e.g. Valkó et al. 2017, Akbash et al. 2018).

[1] The rise of commodity prices, for example, has been described as a contributing cause of the sub-prime crisis of 2007, which later triggered the larger crisis that affected developed economies (Breitenfellner et al. 2015).

Similar indices of economic distress have been developed, but at country level (e.g. Angelopoulou et al. 2014, Bhattacharya–Sinha Roy 2009, Choi–Douady 2012, Louzis–Vouldis 2013). This paper shows that our WD indices satisfy the properties of *non-triviality* (i.e. at least two different values of the indices exist), *symmetry* (i.e. the indices are independent of the names of the countries involved), *normalization* (i.e. the indices are in the range [0,1]) and *weak monotonicity* (i.e. if one economy becomes larger, the impact of a crisis cannot be reduced). Each of these assumptions can be contested in certain contexts. However, they represent the usual properties of fair and representative indices as developed by – for example – Sen (1976) and Alkire–Foster (2011).

Using the large database of financial crises at country level developed by Laeven and Valencia (2013), we create time series for the two WD indices in this paper, statistically illustrating the impact of economic crises from 1980 to 2011.[2] Furthermore, we run a comparison with the series concerning the growth of the world's GDP, showing that – in general – crises tend to have a negative impact on economic performance. This is expected, and therefore serves as an indirect proof of the soundness of our definition of the WD indices.

Both of our indices are constructed on a global scale, but the same methodology could be used to obtain similar information about individual regions or free trade areas. In this study, we focus on the global level, as we aim to construct an index which allows us to capture worldwide effects. Our aim is to provide a simple measure for the current economic state of the world, which does not demand excessive amounts of information that may be hard to collect. Even though it is simple to apply our index on a regional level, for the purposes of this study, we are concerned with explaining world crises and their GDP implications, rather than regional crises.

The remainder of this paper is structured as follows: Section 2 presents the World Distress Index in its two versions, while Section 3 focuses on their properties. Section 4 provides an empirical argument for the soundness of these indices, and Section 5 concludes the paper.

Definition of the world distress indices

Laeven and Valencia's (2013) large database of crises from all over the world from 1980 to 2011 provides data concerning different types of financial crises. The main crisis types are *banking*, *currency,* and *sovereign debt* crises. During the period under consideration, Laeven and Valencia (2013) identified 147 banking crises, 211 currency crises, and 66 sovereign debt crises. Our index will be based on a binary indicator function that takes up the value of 1 in the presence of one of those crises. To accomplish this, we first need to understand the differences among them:

[2] We can consider the likelihood of the data being fuzzy and apply some fuzzy methodology as Sunanta–Viertl (2016).

- A *banking crisis* takes place under the following two conditions:
  1. There exist significant signs of financial distress in the banking system and
  2. the government intervenes in the banks in response to significant losses for these banks.
- A *currency crisis* happens if, during a year, the local currency depreciates vis-à-vis the United States dollar by at least 30% and the rate of depreciation is at least 10 percentage points higher than the previous year.
- A *sovereign debt crisis* occurs if during a year, a country's debt with private investors becomes either defaulted or restructured.

We can now define the indicator function $C\,(i,\ t)$, where the first argument $I = 1, \ldots, I$ denotes a country (Laeven–Valencia 2013) consider $I = 188$ countries) while the second, $t = 1, \ldots, T$ represents the year. Then, $C\,(i,\ t) = 1$ if i suffered one of the aforementioned crises in year t, and $C\,(i,\ t) = 0$ otherwise. For example, since Argentina had all three kinds of crises in 2001, we have that $C\,(i_{\text{Argentina}}, 2001) = 1$ while $C\,(i_{\text{Argentina}}, 2000) = 0$, because none of them took place in 2000, where $i_{\text{Argentina}}$ is obviously the index corresponding to Argentina.

$C\,(i,\ t)$ can be combined with two different GDP measures. $GDP^1_{i,t}$ represents the GDP measured using the *purchasing power parity* (PPP) of country i during year t. Conversely, $GDP^2_{i,t}$ is the GDP of i in t, measured in current United States dollars (USD).

To construct our indices of worldwide distress WD^k_t ($k = 1{,}2$) for a given year t, we take, for each country i, the product of $C(i,t)GDP^k_{i,t}$, and add these terms across countries. We normalize the values by dividing the resulting value by the world $GDP^k = \sum_{i=1}^{I} GDP^k_{i,t}$ at year t. This is expressed as follows:

$$WD^k_t = \frac{\sum_{i=1}^{I} C(i,t)GDP^k_{i,t}}{\sum_{i=1}^{I} GDP^k_{i,t}}. \tag{1}$$

We illustrate the results obtained for WD^1 and WD^2 by taking the $I = 188$ country indices of Laeven and Valencia (2013) from 1980 to 2011. Table 1 presents the whole index, while Figures 1 and 2 summarize this information. We can see that our two indices highlight four notorious episodes: the general crisis of 1988 with its epicentre in the United States, the developing countries crises of the late 1990s, the dot-com crash of the early 2000s, and the financial crisis of 2007/08.[3]

[3] For more information on these events, see Reinhart–Rogoff (2009).

Table 1

Crises index: PPP and USD

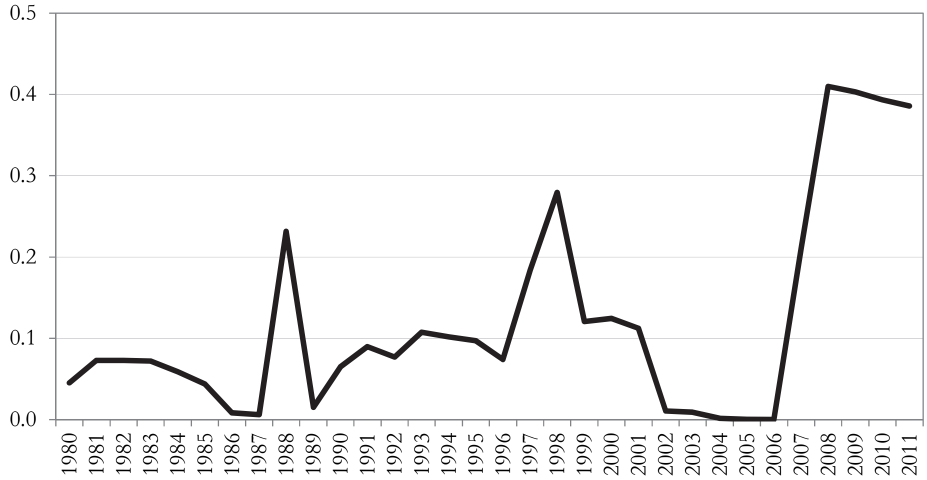
Year	WD^1	WD^2	Year	WD^1	WD^2
1980	0.046	0.045	1996	0.074	0.047
1981	0.073	0.066	1997	0.184	0.210
1982	0.073	0.053	1998	0.279	0.231
1983	0.072	0.046	1999	0.121	0.158
1984	0.059	0.034	2000	0.124	0.166
1985	0.044	0.027	2001	0.113	0.148
1986	0.008	0.003	2002	0.011	0.005
1987	0.006	0.003	2003	0.009	0.004
1988	0.232	0.281	2004	0.002	0.001
1989	0.015	0.007	2005	0.001	0.000
1990	0.065	0.036	2006	0.000	0.000
1991	0.090	0.061	2007	0.211	0.304
1992	0.077	0.051	2008	0.410	0.553
1993	0.107	0.058	2009	0.403	0.544
1994	0.102	0.066	2010	0.393	0.511
1995	0.097	0.070	2011	0.386	0.494

Note: Here and hereinafter, WD^1: world distress index using PPP as the GDP measure; WD^2: world distress index using USD as the GDP measure.

Our indices, which are easy to understand, provide a simplified way of obtaining an economic picture of the world. For example, the value of WD^1 in 2007 was 0.211, which indicates that 21.1% of the world GDP measured using PPP was in crisis. Moreover, the index level increasing to 41% in 2008 means that the world GDP affected by the financial crisis increased sharply. Subsequent years saw a decrease in the exposure of the world GDP to the crisis, with diminishing levels of both WD^1 and WD^2.

Figure 1

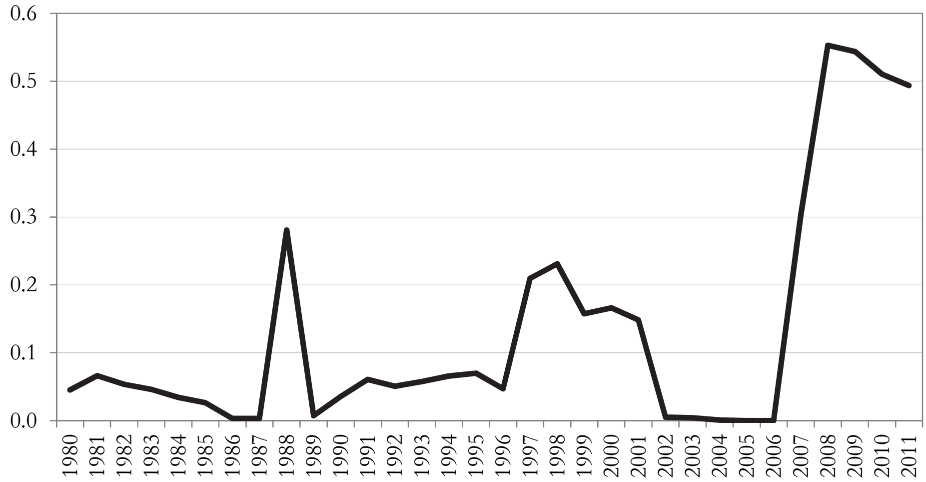
PPP crisis index



Note: Here and hereinafter, PPP: purchasing power parity.

Figure 2

USD crisis index



Institutions such as the International Monetary Fund or the Organisation for Economic Co-operation and Development generate appraisals with similar objectives by estimating the percentage of deviation of the global GDP from its historical trend or by comparing the world GDP with its pre-crisis level. Both of these indices are defined in relation to measures that, by themselves, may be sensitive to how they are defined, such as the world trend or the pre-crisis global GDP. In this regard, our indices are easier to compute, since we require only a dummy of a crisis for each country and country-level GDP measures.

Properties

There are four properties that a fair and balanced index should satisfy. In our case, it means that each country-level crisis should contribute to WD_t^k in proportion to its contribution to the world's GDP.[4] While it is intuitively obvious that WD^k satisfies these properties, we confirm whether this is the case.

Definition 1 (*Non-triviality*). The range of values of each index has at least two distinct values, that is, $|\text{Range}(WD_t^k)| > 1$.

Non-Triviality is a standard condition for indices. If this condition was not satisfied, it would mean that the state of distress of the world (described by the index) remains constant, rendering the index meaningless. The proof follows trivially from the fact that the indices at any t obtain as weighted averages of values $C(i, t) = 1$ or $C(i, t) = 0$ for every i.

This last point also indicates the validity of:

Definition 2 (*Normalization*). The range of both indices has a minimum value of 0 and a maximum value of 1. That is, for every t, $WD_t^k \in [0,1]$.

Normalization allows for a comparison across time, without having to be concerned with the changes of GDP in time. Since (1) has $\sum_{i=1}^{I} GDP_{i,t}^k$ as the denominator, we can guarantee that the index cannot surpass 1. Conversely, since the numerator's lowest value can only be 0 (the denominator is always positive), the index cannot be lower than 0.

Another obvious property satisfied by our indices is:

Definition 3 (*Symmetry*). If the identities of the countries are permuted, the crises indices remain unchanged.

The importance of this requirement lies in that it avoids 'nominal' traps, for instance the status of a country in world affairs that increases or lowers the world distress by having a crisis. While it is true that a crisis started in a country with strategic relevance may have a more significant impact on the rest of the world, usually, this strategic importance is highly correlated to its GDP. Therefore, the change in names would not affect the value of WD^k much.

[4] By a slight abuse of language, we will use k to represent in a single expression both indices, where either $k = 1$ or $k = 2$, measuring the GDP in PPP or in current USD respectively.

Finally, we have

Definition 4 (*Weak monotonicity*). Other factors being equal, an increase in income of a country in crisis cannot decrease the crisis indices. That is, if $C(i, t) = 1$, an increase from $GDP_{i,t}^{k}$ to $GDP_{i,t}^{k}{}'$, cannot lead to a decrease from WD_t^k to $WD_t^{k}{}'$ if $GDP_{i,j}^{k}$ and $C(j, t)$ remain invariant for $j \neq i$.

Weak monotonicity refers to the fact that bigger countries should have a higher impact on the index. Therefore, if a country gets bigger even while in crisis, the index should be at least as high as before, since the 'importance' of this country has grown and with it its contribution to the global crisis level. Our indices satisfy this condition:

Proof: Suppose there exists at least a country j for which we have $C(j, t) = 1$ and assume that $WD_t^k > 0$. Next, increase $GDP_{j,t}^{k}$ by $\epsilon > 0$, leaving all other variables unchanged.[5] This leads to a new value for the index, $WD_t^{k}{}'$.

If

$$WD_t^k = \frac{\sum_{i=1}^{I} C(i,t) GDP_{i,t}^{k}}{\sum_{i=1}^{I} GDP_{i,t}^{k}},$$

we have that

$$WD_t^{k}{}' = \frac{\sum_{i \neq j} C(i,t) GDP_{i,t}^{k} + C(j,t)(GDP_{j,t}^{k} + \epsilon)}{\sum_{i \neq j} GDP_{i,t}^{k} + (GDP_{j,t}^{k} + \epsilon)}$$

$$= \frac{\sum_{i=1}^{I} C(i,t) GDP_{i,t}^{k} + C(j,t)\epsilon}{\sum_{i=1}^{I} GDP_{i,t}^{k} + \epsilon}.$$

Since $C(j, t) = 1$ and calling $\alpha = \sum_{i=1}^{I} C(i,t) GDP_{i,t}^{k}$ and $\beta = \sum_{i=1}^{I} GDP_{i,t}^{k}$, the last expression means that

$$WD_t^{k}{}' = \frac{\alpha + \epsilon}{\beta + \epsilon}$$

as $WD_t^k = \frac{\alpha}{\beta}$ if the claim was false, we could have that $WD_t^k > WD_t^{k}{}'$, that is,

$$\frac{\alpha}{\beta} > \frac{\alpha + \epsilon}{\beta + \epsilon}.$$

Next, a trivial algebraic manipulation shows that this is equivalent to:

$$\alpha > \beta$$

which contradicts the fact that $\frac{\alpha}{\beta} = WD_t^k \leq 1$.

Therefore, $WD_t^k \leq WD_t^{k}{}'$. ***(End proof)***

[5] Notice that if $C(i, t) = 1$ for every i, increasing the income of any country will not increase the index, since it will remain equal to 1. Analogously, if $C(i, t) = 0$ for every i, the index will remain equal to 0.

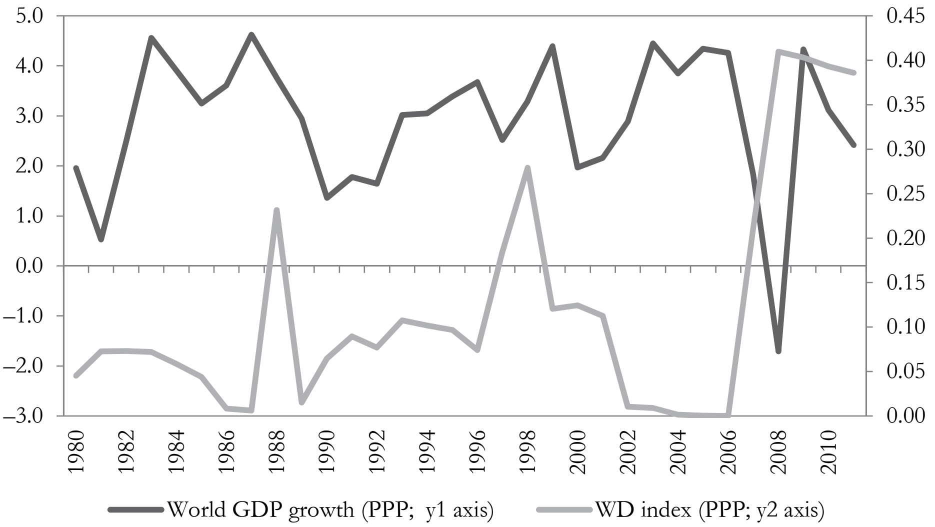
Soundness

To determine whether our definitions capture in some way the impacts of crises, we regress the world's GDP against one of our WD indices. We use the World Bank dataset as a source, selecting the PPP world GDP growth, that is, WD^1.

At first glance, Figure 3 indicates a clear negative relationship for the 2007/2008 crises. However, for previous crises, the relationship is not clear.

Figure 3

World GDP growth and the WD index



To explain this behaviour in former crises, we use the series of first differences of WD^1, because it is $I(1)$ (the Dickey-Fuller statistic is –1.5 in levels and –5.9 for the first differences), while for the worldwide GDP growth, we use the original series, since it is $I(0)$ (the Dickey-Fuller statistic is –4.3, significant at the 1% level). Figure 4 shows that when the differences in WD^1 are positive – indicating that a larger portion of the world GDP is affected by the crises – economic growth slows down. The only exception seems to be around 1998.

If we run a simple regression with WD^1 (in differences, because it is $I(1)$ as the only explanatory variable, we can see that it has a significant and expected negative sign. The results are shown in Table 2. We include a constant (always significant and positive) as well as a trend and lagged growth (neither of which is significant).

Figure 4

World GDP growth and the differenced WD index

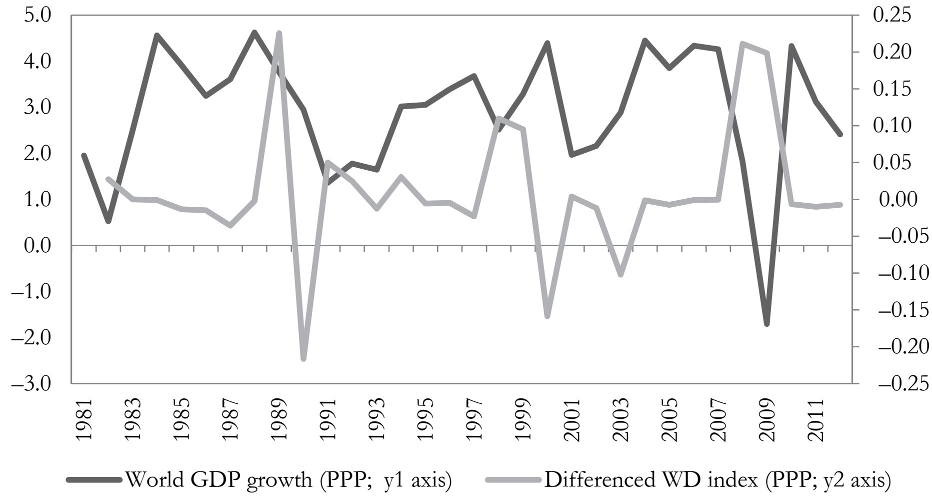


Table 2

Regression results

Variable	(1)	(2)	(3)
	wgrowth		
WD	−5.99**	−5.99**	−6.17**
	(2.59)	(2.65)	(2.55)
Trend		0.00	
		(0.03)	
Lwgrowth			0.24
			(0.17)
Constant	3.02***	3.02***	2.33***
	(0.23)	(0.50)	(0.54)
Observations	31	31	31
R^2	0.16	0.16	0.21

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$. wgrowth: world GDP growth.

This simple exercise seems to indicate that our indices (the case of WD^2 is similar) capture the intended behaviour of crises – that of disruptive events that affect, in this case, the world economy.

Conclusion

This paper proposes a very natural definition of indices that could provide a measure for world economic distress in any given year. We define them as the proportion of the world GDP (either measured in PPP or in current USD) affected by a country-level crisis. By their construction, these indices satisfy the necessary properties that imply that they are faithful and fair representations of how country crises impact world economy by affecting countries' respective GDPs.

In section 4, we related the behaviour of these indices to the growth of the world's GDP. We have shown that they have a degree of explanatory power and can therefore be used in further and more complex analyses. More work is required to ensure that these indices will be efficient explanatory variables in general macroeconomic contexts.

REFERENCES

AKBASH, K. S.–PASICHNYK, N. O.–RIZHNIAK, R. Y. (2018): Generalization of calculation methods for gender indices in demographic and social statistics *Regional Statistics* 8(2): 170–183. https://doi.org/10.15196/RS080205

ALKIRE, S.–FOSTER, J. (2011): Counting and multidimensional poverty measurement *Journal of Public Economics* 95(78): 476–487. https://doi.org/10.1016/j.jpubeco.2010.11.006

ANGELOPOULOU, E.–BALFOUSSIA, H.–GIBSON, H. D. (2014): Building a financial conditions index for the euro area and selected euro area countries: What does it tell us about the crisis? *Economic Modelling* 38: 392–403. https://doi.org/10.1016/j.econmod.2014.01.013

BHATTACHARYA, B.–SINHA ROY, T. N. (2009): Forewarning Indicator System for Banking Crisis in India *SSRN Working Paper.*

BREITENFELLNER, A.–CUARESMA, J. C.–MAYER, P. (2015): Energy inflation and house price corrections *Energy Economics* 48: 109–116. https://doi.org/10.1016/j.eneco.2014.08.023

CHOI, Y.–DOUADY, R. (2012): Financial crisis dynamics: attempt to define a market instability indicator *Quantitative Finance* 12(9): 1351–1365. https://doi.org/10.1080/14697688.2011.627880

LAEVEN, L.–VALENCIA, F. (2013): Systemic Banking Crises Database *IMF Economic Review* 61(2): 225–270.

LOUZIS, D. P.–VOULDIS, A. T. (2013): A Financial Systemic Stress Index for Greece *ECB Working Paper No. 1563.*

REINHART, C. M.–ROGOFF, K. S. (2009): *This Time is Different: Eight Centuries of Financial Folly* Princeton University Press, Priceton.

SEN, A. (1976): Poverty: An Ordinal Approach to Measurement *Econometrica* 44(2): 219–231.

SUNANTA, O.–VIERTL, R. (2016): Fuzzy models in regional statistics *Regional Statistics* 6(1): 104–118. https://doi.org/10.15196/RS06106

VALKÓ, G.–FEKETE-FARKAS, M.–KOVÁCS, I. (2017): Indicators for the economic dimension of sustainable agriculture in the European Union *Regional Statistics* 7(1): 179–196. https://doi.org/10.15196/RS07110

Analysis of the geographical diversification of financial instruments

Bacsosz Sztavrosz
University of Szeged,
Hungary
E-mail:
bacsosz.sztavrosz@gmail.com

The purpose of this research is to examine the effect of the geographical diversification of financial instruments on capital markets. Using geographical diversification in order to create a more effective portfolio would enable investors to use such diversification as a risk mitigation tool. Free capital market data, available to everyone, were primarily used in the research. This study aims to examine geographical diversification and its effectiveness in different regions as well as their historical changes. Historical analysis is not possible in countries lacking developed capital market. A geographically diversified portfolio will result in efficiency gains over the long term. Furthermore, the use of regional and global capital market indexes can be an effective tool for geographic and geopolitical analysis of the relationships between countries and regions.

Keywords:
geographical diversification,
risk,
financial geography,
international portfolio,
geographical exposure,
globalisation

Introduction

The analysis of the risk return of investments is not only of interest to economics researchers but also to anybody who wants to invest and save. The relationship between risk-yield and time to achieve optimum investment performance is a critical factor. Banks, insurance companies and brokerage firms are constantly striving to create the ideal portfolio. Investors are constantly working on models and theories to achieve better performance and better returns. Is the weight of each market news measured efficiently, what price model should be used, and are investors suitably aware of the financial markets they are investing in? They invest significant energy and money to find the best answers to these questions. Complex quantitative models such as machine learning, deep learning, neural networks, big data (Ságvári 2019) and so on are used if necessary (Chatzis et al. 2018). Researchers also make significant efforts to identify the most efficient portfolio (Carrieri et al. 2003).

The use of geographic space and geographical diversification can provide many benefits to investors, and although the benefits of geographical diversification have been recognised, there are relatively few related studies. Some major global financial service providers (e.g. MSCI, Dow Jones, S&P, and Bloomberg) have recognised the importance of geographical diversification and are fighting with their competitors in an attempt to create better models, region indexes to achieve better results. Despite the importance of the topic in most cases the relevance of geographical diversification has been ignored by investors, geographers, and economists.

This research presents the following questions: When creating a portfolio, what level of efficiency could improve the use of geographical diversification in making investment decisions? Who uses risk reduction method of geographical diversification in markets? The study also provides guidance on how financial and capital market indicators can be used by geography researchers. It is important to note that the results of this research can only be applied to countries and regions where there is an at-least-moderately-developed capital market. Where there is no real market, or only a very basic one, capital market indexes may reflect a completely unrealistic picture of a country's economic performance; thus, conclusions drawn on this basis would not be accurate.

The role of geographical diversification in decision-making

The first results of economic and geographic models in which economic events were not represented in a 'single-point economy' (e.g. Thünen, Christaller, Lösch's models) were revolutionary, but these models focused only on the primary and secondary sectors of the economy (Kovács 2014). This is mainly explained by the fact that these theories appeared in the 19th and the first half of the 20th century, when the third sector, including financial services, was undeveloped. Lösch's model contained some innovative elements, such as differences in interest rates or consumer price indexes at different geographic locations (Lösch 1954).

However, despite the successful work of Lösch, classical economic geography focused only on the real estate sector as it is easier to interpret it against the financial sector and it is more closely connected with geographic location. Hence, geography researchers were not interested in words like arbitrage, hedge, and financial diversification.

It should be noted that all aforementioned basic concepts can be organically linked to spatial attributes. Arbitrage can be defined with the following example: when a larger price difference occurs in geographically different markets, some investors (arbitrageur) obtain risk-free profits by buying in one market and selling in another. This process continues until both markets reach equilibrium, wherein prices in the two markets will be almost the same. To further justify the spatial nature of financial markets, consider how current financial markets and their intermediary systems have enabled individuals to invest their savings in markets other than those where the savings originated from.

> '*Consider, for example, a resident of New York City who plans to sell her house and retire to Miami, Florida, in 15 years. Such a plan seems feasible if real estate prices in the two cities do not diverge before her retirement. How can one hedge Miami real estate prices now, short of purchasing a home there immediately, rather than at retirement? One way to hedge the risk is to purchase securities that will increase in value if Florida real estate becomes more expensive. This creates a hedging demand for an asset with a particular risk characteristic. Such demands lead profit-seeking financial corporations to supply the desired goods: observe Florida real estate investment trusts (REITs) that allow individuals to invest in securities whose performance is tied to Florida real estate prices. If Florida real estate becomes more expensive, the REIT will increase in value. The individual's loss as a potential purchaser of Florida real estate is offset by her gain as an investor in that real estate. This is only one example of how a myriad of risk-specific assets are demanded and created by agents in the financial environment.*'
>
> (Bodie et al. 2001., p. 19.)

The citation above has been commonly used in the financial study courses at renowned Hungarian and foreign universities. This simple example can illustrate how risk coverage methods that consider the geographic space may have a bearing on the financial world. The merit of the authors is that the geographical risk sharing techniques are presented in a book under a short chapter called 'international diversification'. However, this does not apply to many other finance instructor's manuals. A number of studies (Solnik 1974) have highlighted that the use of geographical diversification can improve portfolio selection, as many elements of the wider portfolio of financial assets are more beneficial for smaller investors. This fact does not require a long explanation, as it is obvious that a small country, or a small financial market, will be considered small by investors. Investors may not be able to choose from a sufficient variety and quality of securities for obtaining the optimum portfolio, which drives them elsewhere.

The geographical diversification of an investment portfolio, that is, the spread of risk between geographic areas, may have many advantages. However, it can be argued that both foreign and domestic securities have been used by institutions and individual investors to analyse traditional risk sharing methods.[1] The high-level mathematical risk assessment techniques applied in modern advanced capital markets were only developed at the end of the 20th century.[2] This may be explained by the disproportionately minor role of the exploitation of the opportunities offered by the geographic space in the application and research of modern portfolio theories. The following is one proof: in some finance books by famous authors, there are

[1] We primarily mean risk sharing according to company size and industry.

[2] Milton Friedman, the then well-known economist, criticized Harry Markowitz's Study of Portfolio Theory (1952), arguing that what Markowitz wrote about has nothing to do with economics. However, in 1990, he received the Nobel Prize for Economics for developing the theory of portfolio selection. (Bernstein 2005)

only a few pages on a brief outline of the possibilities. Of course, there are some exceptions in the domestic and foreign literature (Pálosi-Németh 2007).

We briefly examined the monthly reports on domestic and foreign funds and found that, in addition to the list of countries that attempt to highlight the geographic exposure of individual securities, they do not play a significant role in geographical diversification. A similar conclusion can be drawn if we examine the fundamental or technical analysis of brokerage firms. In many cases, risk indicators that require a higher level of mathematical background such as VaR, Beta, Alpha, Sharpe, and so on are presented. However, the spread of risk in geographic space is not clearly or comprehensively presented. For example, the breakdown of investments by country is often oversimplified. What could explain the fact that geographical diversification, used as a risk mitigation technique, is not treated as a more important factor?

The main reason for the ignorance regarding geographical diversification may be that its significance is not easily perceived. The increasingly rapid and unrestricted flow of capital in the global economy, online 24-hour trading platforms, and the spread of a wide range of electronic currencies can give the impression that the geographic risk of financial and capital market products is largely meaningless.

The importance and theoretical framework of geographical diversification

Although financial geography has been discussed for more than two decades, only a few related studies are to be found in the Hungarian literature. Economists have also marginalised the study of spatiality in their financial research. O'Brien (1992) predicts the 'death of geography' as a result of globalisation. According to his hypothesis, the significance of the geographical space is lost due to the ongoing deregulation at the time. Instead of physical distance, the cost of obtaining information will be the most important factor in business decision-making as distance is easily overcome by electronic money transfers and dematerialised securities (O'Brien 1992). He believes that the importance of geographic location as a factor for economic development is decreasing. To understand the benefits of diversification, it is essential to have a basic knowledge of portfolio theory as similar tasks and problems are encountered in the analysis, selection, and allocation of securities while using and examining geographical diversification.

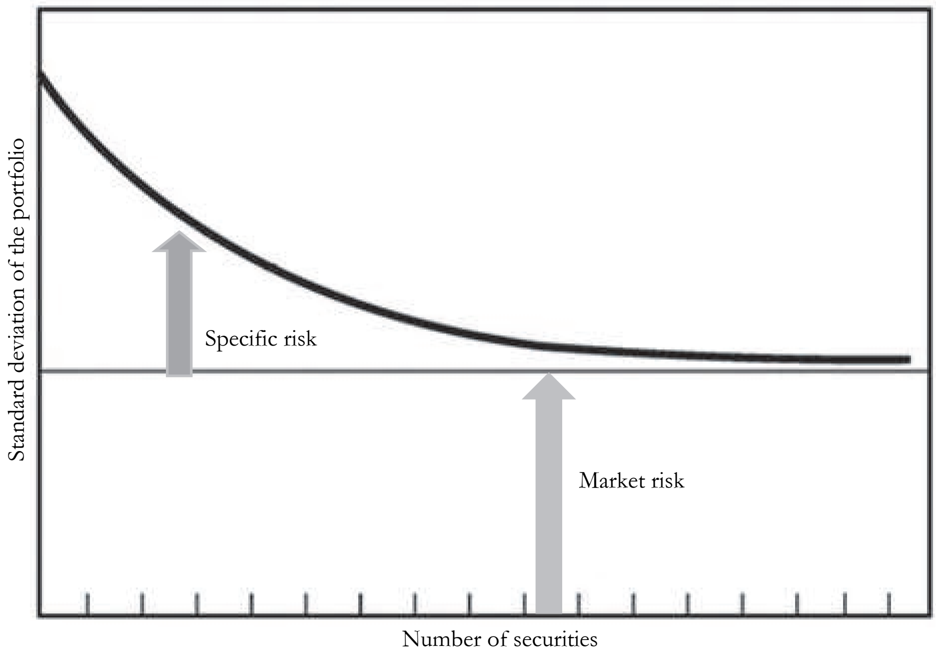
Markowitz's Portfolio Selection (1952), published in the *Journal of Finance*, explains that increasing diversification reduces the variance of the portfolio's yield.[3] Since then, it has also been proven that by increasing the number of risky assets

[3] Variance does not mean anything other than a squared standard deviation. The probability calculation is a common indicator used to characterize the distributions.

(i.e. shares in the portfolio), volatility (i.e. exchange rate volatility) can be significantly reduced, but the so-called market risk cannot be completely eliminated. The risk that cannot be eliminated despite diversification is called market risk or systematic risk. Market risk can be explained by threats across a country's economy, which can affect all companies. In such cases, prices may move regardless of how diversified a portfolio is. On the other hand, specific risk or otherwise diversifiable risk can be reduced through diversification (see Figure 1).

Figure 1

Difference between specific risk and market risk by increasing number of securities



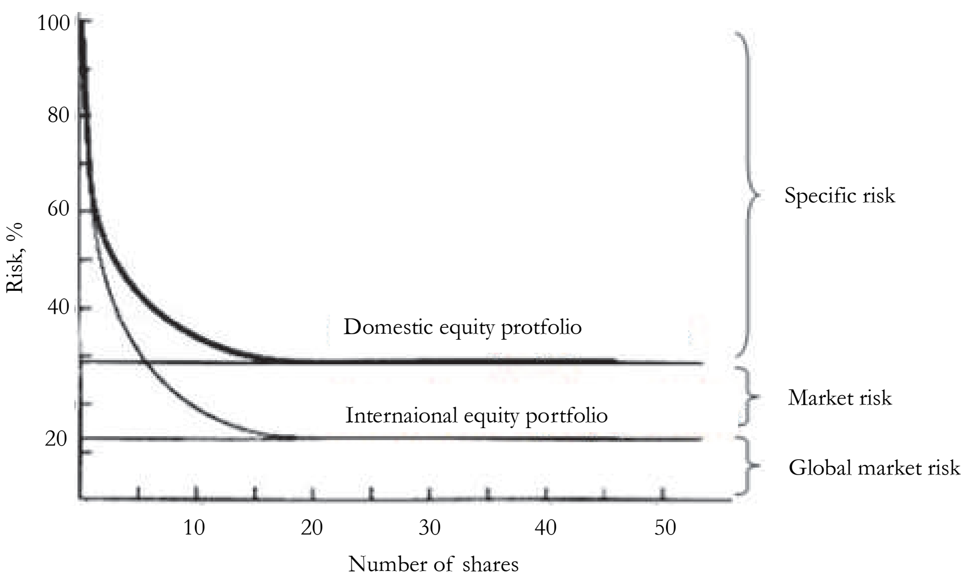
Source: Own elaboration based on Bodie et al. (2001).

Generally speaking, we can eliminate the specific risk of the portfolio by using 15-20 financial instruments. It is also clear from Figure 2 that by increasing the number of securities, the individual risk decreases to a little extent until no increase in efficiency can be achieved by increasing the number of elements (Griffin–Karolyi 1998).

We define market risk as the financial and capital markets of a particular country. A country's national economy may not follow global trends and may slip into recession. For example, an investor places all their savings in a well-diversified portfolio, invests in the Russian stock exchange (e.g. by buying an index product[4]), and then realises that the Russia–Ukraine conflict is devaluing their assets. In this case, the market risk of that country was not eliminated by the investor. Figure 2 also reveals that further efficiency gains can be achieved through a cross-border geographically diversified portfolio.

Figure 2

Domestic vs. international equity portfolio risk



Source: abszoluthozam.hu/images/kockazat.png

In terms of risk, this study focuses on diversification, particularly the sharing of risk between industries. The financial literature rarely refers to other risk diversification opportunities. However, as a truly effective portfolio can be created, above all, by investing in different industries, where we consider the size of the companies as they are at different stages of their lifecycles, our portfolio includes securities with sufficiently differentiated growth potential and stability.

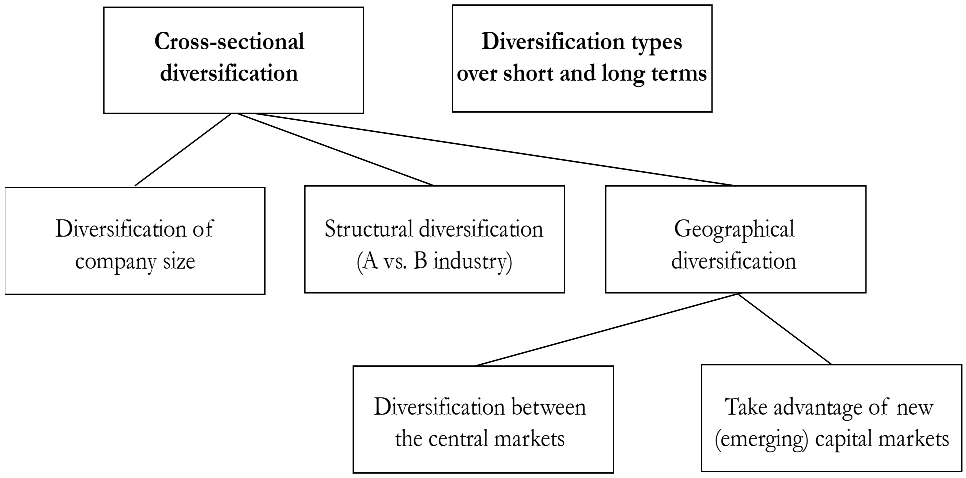
Figure 3 reveals that the use of structural diversification alone cannot be sufficient for investors to create the most effective portfolio in terms of their risk profiles. However, when explaining diversification, the finance literature mainly focuses on structural diversification. Markowitz and Sharpe, the pioneers of risk sharing, explain risk-yield relationships with the correlation between A and B companies in the investment universe. At the same time, if we look more closely at Figure 3, it is

[4] A type of security that tries to copy the performance of some capital market indicator.

striking that cross-sectional diversification is merely one means of spreading risk. There are other options that can be used to spread risk, and efficiency can be increased by careful selection of company size and the use of geographical diversification (Beckers et al. 1996). Moreover, buying and selling financial instruments at different times also has the potential of reducing the risk of the portfolio.

Figure 3

Possible ways of spreading risk



Source: Pálosi-Németh (2007).

Institutions in financial markets, such as banks, insurance companies, brokerages, and funds, offer many financial products and services that attempt to reduce the risks in the aforementioned manner.

Efficiency of geographical diversification

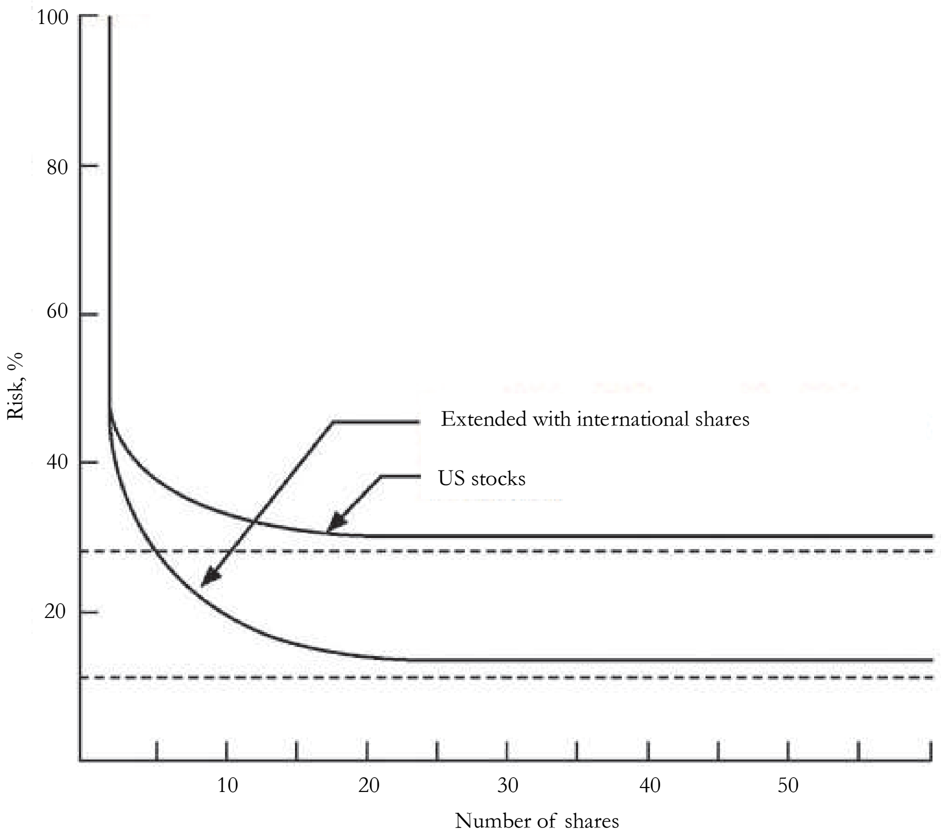
Institutions (banks, insurance companies, pension funds, other funds, etc.) offer not only investment products which one can open speculative positions with, but also financial instruments that can be used to significantly reduce risk; however, as a result, the composition of the savings changes.

One seldom discussed feature of financial intermediation is that it not only collects geographically scattered savings, but also transfers savings through different risk investment products to different points of the geographical space, thereby significantly reducing the cost of resources and risk for each investment.

Solnik (1974) have already demonstrated how the use of geographical diversification can result in significant efficiency gains. As shown in Figure 4, which illustrates the standard deviation of equally weighted portfolios of different sizes, the authors demonstrate the weaknesses of an exclusively US investment-focused portfolio. The value of 20% on the vertical axis means that the deviation of the diversified portfolio is 20% of the standard deviation of a single share portfolio. It is clear that significant risk reduction can be achieved if, for example, in addition to US shares, we also hold foreign shares in our portfolio. Figure 4 also illustrates that a geographically well-diversified portfolio reduces the risk of the domestic portfolio by almost a half.

Figure 4

Advantages offered by international diversification



Source: Solnik (1974).

According to the following table, the market capitalisation[5] of the US accounts for over one third of the total capitalisation. This means that investors who only favour the capital market of their own country cannot exploit the benefits of diversification.

[5] Multiplied by the price and number of shares. This is nothing more than the market value of the company.

The 25 largest capital markets, 2017

Country	Market capitalisation, USD million	Percentage of total capitalisation, %
United States	32,120,703	40.54
China	8,711,267	11.00
Japan	6,222,825	7.85
United Kingdom	3,605,561	4.55
France	2,749,315	3.47
Canada	2,367,060	2.99
India	2,331,567	2.94
Germany	2,262,223	2.86
South Korea	1,771,768	2.24
Switzerland	1,686,497	2.13
Australia	1,508,463	1.90
South Africa	1,230,977	1.55
Sweden (OMX Nordic)	1,122,815	1.42
Netherlands	1,100,105	1.39
Brazil	954,715	1.21
Spain	888,838	1.12
Singapore	787,255	0.99
Russian Federation	623,425	0.79
Italy	561,428	0.71
Thailand	548,795	0.69
Indonesia	520,687	0.66
Malaysia	455,772	0.58
Saudi Arabia	451,379	0.57
Belgium	437,794	0.55
Mexico	417,021	0.53
Total	*79,224,525*	*100.00*

Note: Deviation from 100.00 results from rounding.
Source: Own calculation based on databank.worldbank.org and sdw.ecb.europa.eu

However, we can also see that the global capital market is heavily concentrated and the market capitalization of the top 25 countries is 95.22, suggesting that some markets may be so small that large hedge and pension funds may influence market prices through their investments. Only five countries have market capitalisation exceeding 3 of the total global capitalisation: the US, China, Japan, the United Kingdom, and France.[6]

In the context of the theoretical framework, it has been shown that if we add additional assets to the portfolio, where the correlation between them is not strong our portfolio may become more effective. Thus, the yield per unit may increase or be obtained at a lower risk. In global financial markets, capital flows smoothly between countries; as a result, we can create more diversified portfolios and make these benefits cheaper and faster. However, investors face new issues by spreading risks geographically. Examples include country legislations, political risks, and capital market traditions, as well as currency risks.

However, in their decision-making, investors mainly select the portfolio mix based on their risk preference. In investment decision-making, the size of the portfolio, investor's risk tolerance, time horizon of the investment, and returns on alternative investments can play a decisive role for many factors. It is difficult to determine which of these are the most important.

At the same time, small domestic investors and large international institutional investors have completely different attitudes. They often react differently, and their opinions may be different if they have to place their financial assets in developed or emerging markets. Individual investors often overstate information about their own country, so domestic markets play a greater role in their portfolio. This phenomenon is referred to as '*Home Bias*' (Caprio 2012). This phenomenon exists in almost every market, but by examining its scale, we can obtain different results regarding geographic space and distant time horizons.

Figure 5

The major classes of investments

Asset type	Geographical exposure	Industry	Investment strategy
Equity, bond, property	Developed vs. emerging; Domestic vs. international	Technology, pharmaceuticals, energy	Growth vs. risk aversion

Overweighting domestic markets can be explained by the fact that, despite rapid communication tools, information asymmetry[7] in financial markets increases with the increase of geographical distance.[8] However, '*At the beginning of the 1980s, investors, even institutional investors or individual investors invested only 5 of their money in foreign securi-*

[6] It is important to note that a particular security may not only be traded in one country, so in many cases it is very difficult to determine its geographical location.

[7] The unequal distribution of information among investors in the financial market, which is often due to the fact that one investor has an advantage over the other investor.

[8] The overweighting of the domestic market is not only due to lack of information. As domestic market is free from exchange rate risk, in the most cases the transaction costs are lower.

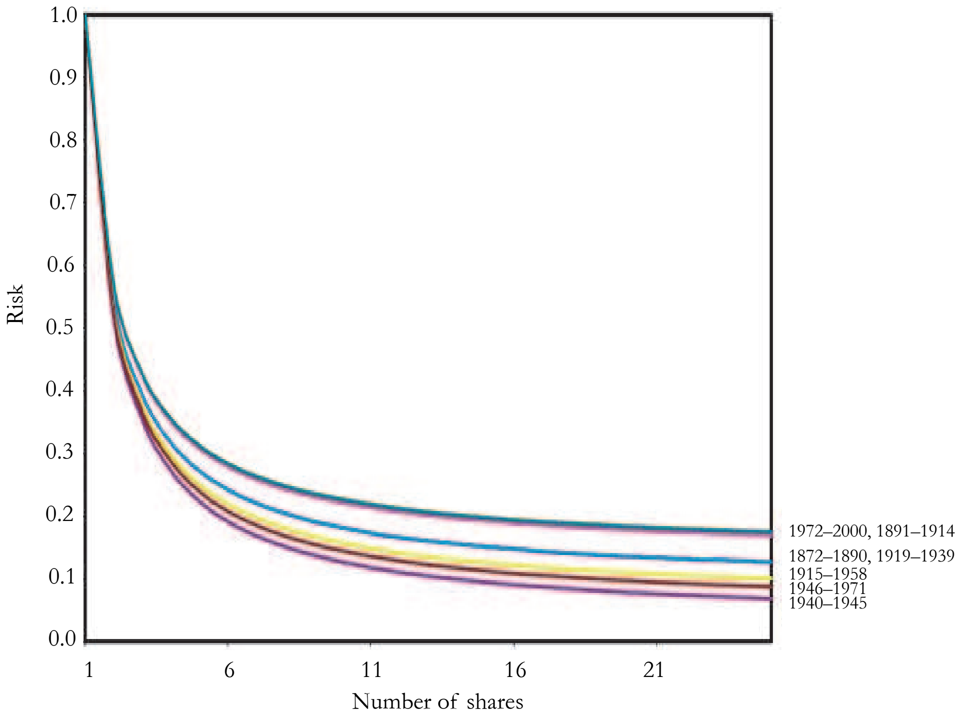
ties. On the other hand, the 15-20 ratio has become quite usual, but in some smaller countries with a developed financial market even one third of the portfolio exceeds the ratio of foreign assets'. (Gál 2010., p. 16.)

This quote demonstrates that cross-border risk sharing techniques are becoming increasingly important. Modern investors purchase fewer direct financial instruments, they are buying instead indirect financial instruments (e.g. investment funds, index funds, ETFs [exchange traded funds], certificates, etc.), largely due to the securitisation[9] revolution over the last few decades.

Figure 6 illustrates that the efficiency of geographical diversification has declined significantly, primarily because of the fact that, as a result of globalisation, a significant proportion of investors favour a larger share in developed markets in their portfolio.

Figure 6

Effectiveness of geographical diversification with increasing number of markets at different times



Source: Goetzmann et al. (2001).

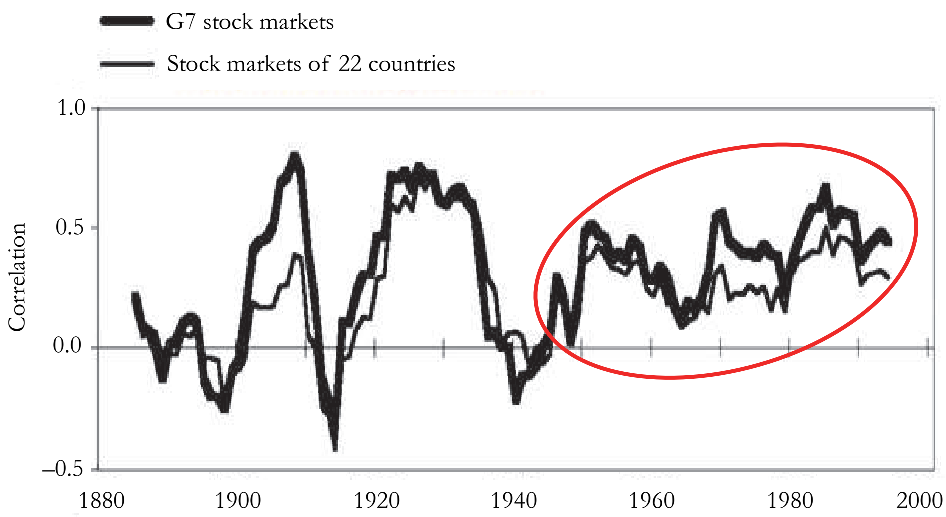
However, dozens of new financial products are available and these financial instruments are becoming increasingly complex or difficult to operate across larger geographical distances. Thus, at the outbreak of the 2008 financial crisis, large geo-

[9] When bypassing lending banks, companies directly receive money as a result of securitization.

graphical distances completely blurred key information for investors and severely impacted the world's money and capital markets (see Figure 7) (Zsibók 2017, Kocziszky et al. 2018).

Figure 7

Historical review of the correlation of capital markets



Note: G7 – Group of Seven (Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States).

Source: Obstfeld–Taylor (2003).

Geographical diversification of the markets

The most important research question in this study is how much more could be achieved if portfolios were developed with greater consideration of geographical risk spreading. Figure 8 illustrates changes in the variability of two diversified portfolios if a geographically diversified portfolio is favoured.

Figure 8

Relative historical change of BUX and MSCI EAFE capital market indexes

Note: Here and in Figure 9, BUX – Budapest Stock Exchange index. MSCI refers to MSCI Inc., EAFE stands for Europe, Australasia and Far East.

Source: Own elaboration based on Yahoo Finance (https://finance.yahoo.com).

The following formula is suitable for the correct comparison of dispersion indicators in different markets. Relative dispersion is measured with the coefficient of variation (CV), which is computed as:

$$CV = \frac{s_x}{\overline{X}} = \frac{\text{standard deviation of } x}{\text{average value of } x}.$$

A direct comparison between the dispersion of the two or more distributions is sometimes not meaningful due to the relatively large difference in their means. The following is a concrete example of what the CV shows.[10] The numerator is the standard deviation of annualised yields, while the denominator is the average of annualised returns. We obtain the following results:

$$CV_{\text{BUX}} = 0.3098/0.1065 = 2.9103$$

$$CV_{\text{MSCI EAFE}} = 0.2010/0.0753 = 2.6694.$$

[10] The period was from 01/02/2002 to 12/31/2014.

A higher CV value is worse because it indicates a higher risk for the same return. That is, the investor obtains a smaller return for the same risk. The efficiency of geographical diversification is well demonstrated by the CV. The geographically well-diversified MSCI EAFE portfolio performs better than the single-market BUX portfolio, meaning there is less risk for one unit of return.

$$CV_{\text{S\&P 500}} = 0.1668/0.0566 = 2.9445$$

$$CV_{\text{MSCI EAFE}} = 0.2010/0.0753 = 2.6694.$$

It is clear that a high number of securities alone is not enough to maximise the efficiency of our portfolio if it is not geographically diversified. The S&P 500 index consists of 500 shares but performs worse than the Budapest Stock Exchange index (BUX)[11] in the examined term. Despite the fact that the S&P 500 index has a high number of shares and a low standard deviation, it is still the worst investment if we consider the other two portfolios. It is also clear that standard deviation alone is not suitable for measuring the risk of an investment (thus, variance is not suitable either).

Figure 9 presents the three most important Eastern European capital market indexes (BUX, WIG[12], PX[13]) in the past 20 years, relative to each other. The budgetary crisis in 2008 led to a significant fall in the capital market indexes of all three countries.[14] However, it is also apparent that although these countries began with a similar level of development after the regime change, the stock indexes of each country perform differently. The yield performance of an investor who preferred the Hungarian stock market instead of the Czech stock exchange since 1998 is three times better. We can also conclude that up until 2011, these three stock indexes were very similar, that is, the correlation between them was significantly positive. The Polish and Hungarian stock markets have outperformed the Czech stock exchange since 2013.[15]

[11] BUX consists of approximately 12-24 companies' shares.

[12] Warsaw Stock Exchange index.

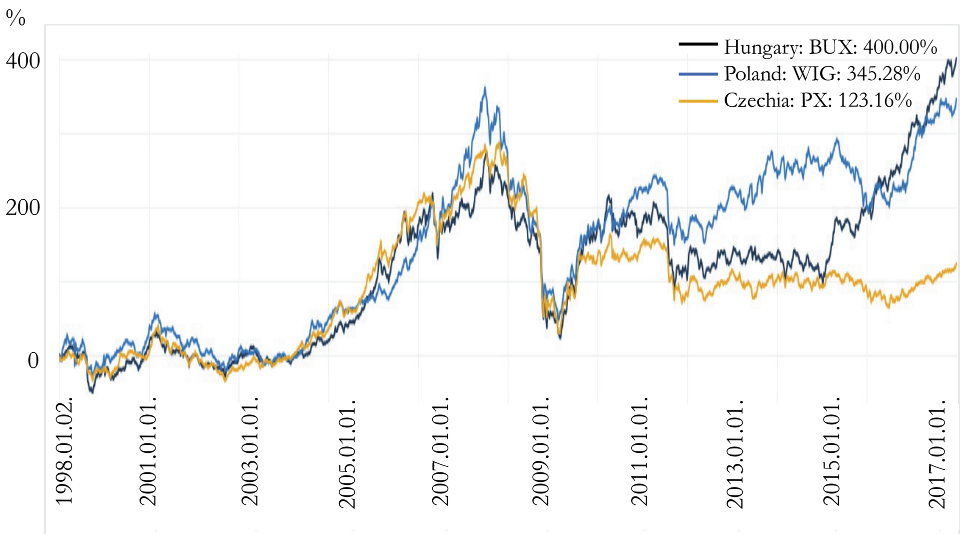
[13] Prague Stock Exchange index

[14] It is important to note that in Hungary the per capita GDP has decreased more in the next few years than in the neighbouring countries (Egedy 2012).

[15] It is important to note that different results can be obtained by changing the study period. By selecting the longest time series this study attempted to draw conclusions from the three most important Eastern European stock indexes. In 2018, the relative performance of the Czech stock market index was similar that of the Polish stock market index.

Figure 9

Relative historical change of BUX, WIG, and PX capital market indexes



Note: WIG – Warsaw Stock Exchange index; PX – Prague Stock Exchange index.
Source: Own elaboration based on *The Wall Street Journal* data.

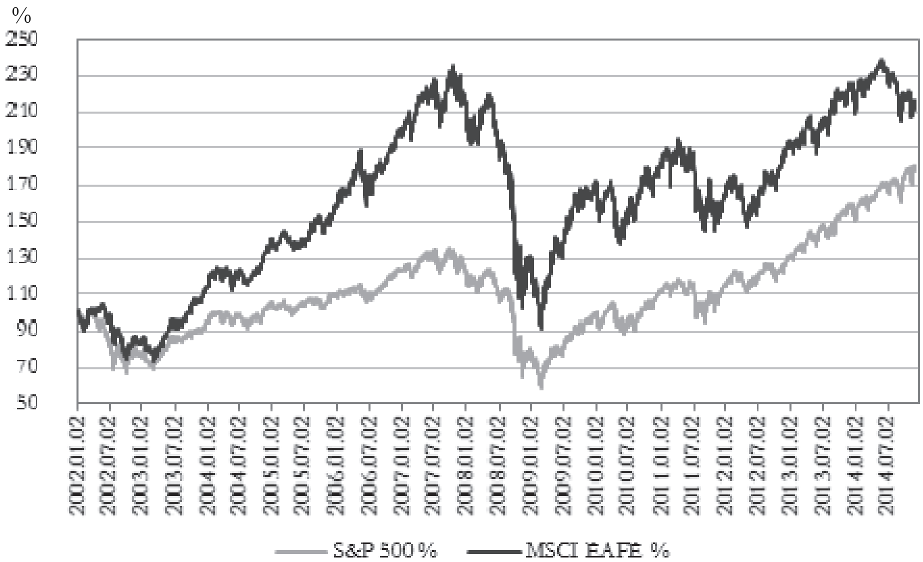
Figure 10 shows that over a thirteen-year period our return is roughly the same as the investors' who invested in the first days of January 2002, but instead of the MSCI EAFE index[16], chose only BUX. We consider an example of what happens when the developing market index, BUX, is replaced by one of the most accepted and well-known developed US capital markets index: the S&P500. The figure clearly shows that the S&P 500 index was not as volatile as BUX and even more stable than the MSCI EAFE index. The same conclusion is reached if we consider the MSCI EAFE 1.0366 Beta indicator. The Beta value of 1.0366 means that in the case of the 1 yield of the S&P 500 index, a 1.0366 shift in the same direction was observed. We know that if Beta > 1, the selected portfolio is more risky than the selected reference index.

This also means that a yield of 1 of the positive return is higher if we keep the MSCI EAFE index in our portfolio instead of the S&P500 index, since the Beta value of 1.0366 shows that the MSCI EAFE index is more risky. This is remunerated by the market with a premium yield.

[16] It is one of the oldest global capital market indexes, which was started by Morgan Stanley in December 1969. The index includes securities from 23 developed countries, but the primary goal was not to include American and Canadian papers.

Figure 10

Relative historical change of S&P 500 and MSCI EAFE capital market indexes



Note: S&P500 – an American stock market index.
Source: Own elaboration based on Yahoo Finance (https://finance.yahoo.com).

Beta is a risk indicator with a regression analysis that can yield different results depending on the benchmark[17]. In addition, different Beta results are obtained by selecting different time intervals. It is very important to clarify that the correlation coefficients can change significantly in the increasingly globalised economy during major market recessions and we can observe a stronger correlation relationship. This means that if there is a fall in prices in some large capital markets, prices are likely to fall in other markets. This spillover effect was observed during the 2008 financial crisis.

As previously mentioned, a negative Beta-denominated portfolio can easily move in the same direction as the Beta reference index during a crisis. For risk analysis with Beta, it is important to note what is chosen as the market reference index. Although the capital market index of the S&P 500 can be a good indication of the mood of global capital markets, there are still many challenges. For example:

– The US capital market mood may have an unavoidable impact on any other capital market even though it only shows the performance of the US capital market.

[17] 'One of the most important, but often overlooked, influences on the asset allocation decision is the choice of the benchmark by which to measure risk. In mean variance optimization, the objective is to maximize return per unit of portfolio risk. The investor's benchmark defines the point of origin for measuring risk. That is, it represents the minimum-risk portfolio.' (Fischer–Litterman 1992, p. 38.)

– We do not always obtain an accurate picture of risk-yield research when we compare the markets in developed and emerging economies.
– Although the S&P 500 index is sufficient, it is not always suitable for global or regional capital market analysis.

Faced with these problems, many news service providers, analysts, and fund managers have attempted to create their own regional and global capital market indexes.[18]

Conclusion

This paper aimed to highlight the fact that the global financial market provides a good environment for diversification as it offers a great deal of variation opportunities to optimise risk-return. Most of the related methods have already been developed, and instead of relying on existing, mainly ex-post data, real-time data may help achieve more relevant and forward-looking conclusions, which are not outdated and are focused on the future.

The importance of the topic is also verified by the fact that geographical diversification can be used in everyday financial decision-making. Indeed, as demonstrated in this study, achieving the best performance is not only possible, but also useful. Several studies have already shown that investors who do not consider the international capital market will create lower performing securities portfolios (Solnik 1982).

The international and cross-border investments presented herein may illustrate the diversification benefits offered by the geographic space used in the selection of the portfolio, which is why there is a much wider choice of financial instruments. However, in terms of international investment, there are a number of problems that rarely occur in domestic markets. One of the most obvious problems is that, with the increase in geographical distance, the investor faces significant information asymmetry, which can generate dozens of added problems in addition to currency risks, political risks, regulatory decisions, and different taxation and accounting practices. However, these kinds of problems can be adequately addressed through the diversification methods already described.

It should be emphasised that a completely different answer to the research questions would have been obtained if the fact that the efficiency of geographical diversification changes over time were not considered. We also experience different levels of efficiency when small, medium, or large companies are included in our portfolio (De Moor–Sercu 2010). However, other authors have previously highlighted that there are substantial benefits to geographical diversification beyond the amounts attributable to industrial or currency diversification (Heston–Rouwenhorst 1994).

[18] For example, BBC Global 30 index, Dow Jones Global Total Stock Market index, FTSE Euro 100.

REFERENCES

BECKERS, S.–CONNOR, G.–CURDS, R. (1996): National versus global influences on equity returns *Financial Analysts Journal* 52(2): 31–39. https://doi.org/10.2469/faj.v52.n2.1978

BERNSTEIN, L. P. (2005): *Capital Ideas – The Improbable Origins of Modern Wall Street* Corvina Press, Budapest.

BODIE, Z.–KANE, A.–MARCUS, J. A. (2001): *Essentials of Investments, 5th Edition* McGraw-Hill Press, New York.

BROOKS, R.–DEL NEGRO, M. (2005): Country versus region effects in international stock returns source *Journal of Portfolio Management* 31(4): 67–72. https://doi.org/10.3905/jpm.2005.570152

CAPRIO, G. (2012): *The Evidence and Impact of Financial Globalization* University Press, Cambridge.

CARRIERI, F.–ERRUNZA, V.–SARKISSIAN, S. (2003): *The Dynamics of Geographic versus Sectoral Diversification: A Causal Explanation* McGill University Press, Montreal.

CHATZIS, S. P.–SIAKOULIS, V.–PETROPOULOS, A.–STAVROULAKIS, E.–VLACHOGIANNAKIS, N. (2018): Forecasting stock market crisis events using deep and statistical machine learning techniques *Expert Systems with Applications* 112: 353–371. https://doi.org/10.1016/j.eswa.2018.06.032

DE MOOR, L.–SERCU, P. (2010): Country vs. sector effects in equity returns and the roles of geographical and firm-size coverage *Small Business Economics* 35(4): 433–448. https://doi.org/10.1007/s11187-008-9170-6

EGEDY, T. (2012): A gazdasági válság hatásai városon innen és túl *Területi Statisztika* 52(4): 335–352.

FISCHER, B.–LITTERMAN, R. (1992): Global Portfolio Optimization *Financial Analysts Journal* 48(5): 28–43.

GÁL, Z. (2010): *Pénzügyi piacok a globális térben: a válság szabdalta pénzügyi tér* Akadémiai Kiadó, Budapest.

GOETZMANN, N. W.–LINGFENG, L.–ROUWENHORST, G. K. (2001): Long-term global market correlations *Journal of Business* 78(1): 1–38. https://doi.org/10.1086/426518

GRIFFIN, J. M.–KAROLYI, G. A. (1998): Another look at the role of the industrial structure of markets for international diversification strategies *Journal of Financial Economics* 50(3): 351–373. https://doi.org/10.1016/S0304-405X(98)00041-5

HESTON, S.–ROUWENHORST, G. (1994): Does industrial structure explain the benefits of international diversification? *Journal of Financial Economics* 36(1): 3–27. https://doi.org/10.1016/0304-405X(94)90028-0

KOCZISZKY, G.–BENEDEK, J.–SZENDI, D. (2018): The impact of the 2008 financial crisis on household income and wealth in Visegrad countries *Regional Statistics* 8(1): 141–167. https://doi.org/10.15196/RS080102

KOVÁCS, S. ZS. (2014): Elérhetőség és kirekesztés Magyarországon a pénzügyi szolgáltatások aspektusából *Területfejlesztés és Innováció* 8(3): 28–35.

LÖSCH, A. (1954): *The Economics of Location* Yale University Press, New Haven.

MALKIEL, B. G. (1998): *A Random Walk down Wall Street* International Banking Training Center, Budapest.

MARKOWITZ, H. (1952): Portfolio Selection *The Journal of Finance* 7(1): 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

O'BRIEN, R. (1992): *Global Financial Integration: The End of Geography* Council on Foreign Relations Press, New York.

OBSTFELD, M.–TAYLOR, A. M. (2003): *Globalization and Capital Markets Globalization in Historical Perspective* University of Chicago Press, Chicago.

PÁLOSI-NÉMETH, B. (2007): *A feltörekvő országok tőkepiaca a pénzügyi globalizációban* PhD dissertation, Debrecen.

SÁGVÁRI, B. (2019): Tér és társadalom Big Data szemüvegen keresztül *Területi Statisztika* 59(1): 27–48. https://doi.org/10.15196/TS590102

SOLNIK, B. (1974): Why not diversify internationally rather than domestically *Financial Analysts Journal* 30(4): 48–54. https://doi.org/10.2469/faj.v51.n1.1864

SOLNIK, B. (1982): Optimal international asset allocation *Journal of Portfolio Management* 9(1): 11–21. https://doi.org/10.3905/jpm.1982.408895

ZSIBÓK, ZS. (2017): Continuing divergence after the crisis: long-term regional economic development in the United Kingdom *Regional Statistics* 7(1): 17–42. https://doi.org/10.15196/RS07102

INTERNET REFERENCES

abszoluthozam.hu/images/kockazat.png

databank.worldbank.org, sdw.ecb.europa.eu

Domestic vs. International equity portfolio risk. http://abszoluthozam.hu/images/kockazat.png (downloaded: January 2, 2016.)

sdw.ecb.europa.eu

Yahoo Finance (https://finance.yahoo.com)

Dynamic measurement of complex phenomena in assessing the Europe 2020 strategy effects

Iwona Müller-Frączek
Nicolaus Copernicus University in Toruń
Faculty of Economic Sciences and Management,
Poland
E-mail: muller@econ.umk.pl

The paper presents a novel construction of dynamic composite indicators modelled on the Human Development Index (HDI), where, instead of scaling, the author applies normalization with respect to anti-pattern: an original method of variable normalization. This method enables the reflection of different environments (reference groups) in dynamic synthetic comparative studies. An empirical example illustrates the utility of the new method. As regards the Europe 2020 strategy, the socio-economic development of 10 countries that joined the European Union (EU) in 2004 is analysed from the perspective of this group of countries as well as the whole EU. The results for the two groups (environments) are similar, but not identical. The study finally carries out an in-depth analysis for the selected country, Poland. Again, both environments show visible differences.

Keywords:
multidimensional data analysis,
composite indicator,
normalization

Introduction

International or regional comparisons often adopt a synthetic approach. This approach combines different aspects of the studied complex (multidimensional) phenomenon into one quantitative characteristic, the so-called composite indicator (or composite index, a synthetic variable). This synthetic comparative research has advantages and disadvantages (Saltelli 2007), but is popular both in science and journalism owing to the simplicity of its interpretation. For a complex phenomenon, many different examples of composite indicators can be found in the literature. For example, the history of creating various composite indicators for the socio-economic development of countries is described in Booysen (2002). Some researchers create their own composite indicators (e.g. Müller-Frączek–Muszyńska 2016, Valkó et al. 2017), but many scientists use already existing ones (Egri–Tánczos 2018, among others).

The creation of composite indicators should be carried out with due diligence. The rules to be followed and stages of construction are presented in the literature; for example, see Zeliaś (2002), Saisana–Saltelli (2011), Casadio Tarabusi–Guarini (2013), and Mazziotta–Pareto (2013).

From the viewpoint of this article, the composite indicators need to be divided into static and dynamic ones. Static composite indicators characterize the phenomenon for a specific unit of time. They can be used for analysis across time, but then we can only compare rankings and not the index values. However, dynamic composite indicators characterize the phenomenon for various time units so that the rankings as well as the index values can be compared.

This article focuses on dynamic composite indicators. A key issue in the construction of dynamic composite indicators is making the variables comparable, that is, normalizing the diagnostic variables (individual indicators). For examples of normalization formulas, see Milligan–Cooper (1988), Jajuga–Walesiak (2000), and Młodak (2006).

Two approaches are used for dynamic normalization. In the first one, all the available observations of an individual variable (in both time and space) are used to determine the parameters needed for normalization (Nardo et al. 2005). The disadvantage of this solution is that once a new unit of time is observed, all the results will have to be recalculated. In the second approach, the parameters needed for normalization are set in advance and also for the future. This solves the problem of recalculating the results but makes it difficult to define the mentioned parameters, especially when very diverse objects are analysed (for some methods, see Mazziotta–Pareto 2016).

This study applies a different approach to dynamic normalization than the ones mentioned above. Indicators are statically transformed, i.e. for each unit of time separately, nonetheless their normalized values are comparable not only in space but also over time. This procedure is generally not applied, but for the author's method of normalization (called normalization with respect to anti-pattern, or anti-pattern normalization in short), it is correct.

The aim of this article is to present a new method for dynamic composite indicators construction using anti-pattern normalization. A prototype of this method is normalization with respect to pattern, which has been also used by Müller-Frączek (2018) for composite indicators construction. Anti-pattern normalization is used for the first time in this study.

Normalization with respect to anti-pattern is adapted for dynamic studies, but unlike popular methods, it uses observations from only the current time unit (a static approach). Therefore, the composite indicators based on this method do not require the results to be recalculated after the appearance of observations for a new time unit.

To analyse a complex phenomenon, we often assess the situation of the object (country, region) against the background of different reference groups (environments). Depending on the normalization method we choose for constructing the composite indicator, the reference group may or may not influence the synthetic results. During anti-pattern normalization, observations for all objects are used, therefore the research environment is really important for constructions based on this method. The normalization results obtained are relative, and the results for another reference group can be very much different.

Note, importantly, that not all dynamic synthetic analyses can be conducted using the proposed composite indicator. Studied complex phenomena should consist of several aspects that are not substitutable with one another. However, each of these aspects can be represented by several mutually substitutable indicators. In addition, the results obtained using the proposed method strongly depend on the analysis environment. They reflect not the objective changes in indicators, but the changes in terms of the analysed reference group. It is worth confronting them with research based on other methods of normalization. Then the image of the studied phenomenon will be more complete.

To illustrate the utility of the proposed method, we analyse the countries in terms of implementation of the Europe 2020 strategy, a program of socio-economic development of the EU countries for 2010–2020 with the following priorities:

- smart development,
- sustainable development,
- inclusive development.

The appropriate indicators to monitor the implementation of the strategy are determined. They are published in the Eurostat database. For more information on the strategy, see the official website of the European Commission.

In this study, we present a dynamic synthetic analysis of the EU countries' development in the context of the Europe 2020 strategy, with focus on 10 countries that joined the EU in 2004 (this group is referred to as the EU10 countries in this article). The development level of the individual countries is assessed from the perspectives of two groups: the EU10 countries and the whole EU. The results obtained for these reference groups are then compared.

To highlight the difference in the results for the different environments, we carry out an in-depth analysis of one country, Poland.[1] Note that we discuss not the development of this country, but the difference between the results obtained for the different environments.

The remainder of the article is structured as follows. Section 2 presents the construction of the dynamic composite indicator. Section 3 describes our empirical study. Section 4 presents the synthetic analysis results of the EU10 countries' development. Section 5 compares the analysis results obtained for the two environments. Section 6 presents an in-depth analysis for Poland. The article ends with conclusions.

[1] The described results were presented at the V. International Scientific Conference 'Spatial Econometrics and Regional Economic Analysis' in Łódź (Poland). Therefore, we selected Poland as our example.

Construction of the dynamic composite indicator

An aggregation of the variables used for the presented construction is modelled on the HDI, a composite indicator used by the United Nations for dynamic comparison of countries around the world in terms of the level of socio-economic development (for more information on HDI, see the official website of the United Nations Development Programme).

In the first step of the construction, we normalize the individual variables (indicators).[2] Next, we divide them into groups, corresponding to the different aspects of the multidimensional phenomenon. We assume more than one variable in each group. We also assume that the variables connected with an aspect are substitutable; that is, the surplus in one indicator can compensate for a deficit in another. Thus, we use a compensatory method of aggregation within groups (the arithmetic mean).[3] However, since the different aspects of the phenomenon are assumed to be non-substitutable, we integrate all aspects into one combined composite indicator using a method that hinders compensation between the aspects (the geometric mean).[4]

The difference between the construction of the HDI and that of the composite indicator proposed in this article is due to the use of different method of variable normalization. As mentioned in Section 1, the presented construction method uses normalization with respect to anti-pattern, whereas the HDI uses scaling min-max normalization or zero unitarization in the version adopted for dynamic analysis.

Let $n \in N$ be the number of considered objects (regions, countries, etc.) and $x = (x^1, x^2, \ldots, x^n) \in R^n$ be a selected diagnostic variable characterising a complex phenomenon (e.g. socio-economic development) in a fixed time unit, for simplicity of notation. For each object $i = 1, \ldots, n$; the scaling used in HDI transforms variable x into variable $s = (s^1, s^2, \ldots, s^n) \in R^n$ as follows:[5]

$$s^i = \frac{x^i - x^{\min}}{x^{\max} - x^{\min}} \quad \text{if} \quad x \in S. \tag{1}$$

In a basic (static) version of the scaling $x^{\min}$ and $x^{\max}$ are the maximum and minimum among the values of the variable x for a given time unit. In a dynamic approach, however, the maximum and minimum among values for all objects and all time units are selected. As mentioned in Section 1, the disadvantage of this method is that all the previous results have to be recalculated when the observations for a new year appear. Therefore, in the HDI, $x^{\min}$ and $x^{\max}$ do not result directly from

[2] The problem of variable selection is beyond the scope of this article.

[3] For more information on substitution in variable aggregation methods, see, for example, Klugman et al. (2011).

[4] For simplicity of presentation, we also assume that all features in each group are equally important and skip the weighting variables step.

[5] Formula (1) is given for stimulants because HDI contains only stimulants. For a general form and other details of this normalization method, see, for example, Kukuła–Bogocz (2014).

the distribution of the variable but are predetermined values ('goalposts') common for all countries and all (also future) years of analysis (Tarantola 2008).

In practice, it is difficult to determine the values of $x^{\min}$ and $x^{\max}$ that are adequate for all countries in the long term, because, first, the countries of the world are very diverse in terms of socio-economic development level, and, second, the values of the indicators used in the HDI are growing in time. The developed countries have already exceeded the maximum level of some of the indicators used in the HDI.

In the context of this study, scaling has one more disadvantage: the value of s for a given country is not influenced by those of other countries, and so the environment of the research is not relevant. Normalization with respect to anti-pattern does not have this disadvantage.

The indicators forming the HDI are stimulants, that is, diagnostic variables, which have a positive impact on socio-economic development. In general, a set of diagnostic variables can also include destimulants, which have a negative impact on the studied phenomenon. To use all the diagnostic variables in the construction of a composite indicator, the destimulants should be converted into stimulants. Such a possibility is given by normalization with respect to anti-pattern, which is another advantage of this method.

In the first step of normalization, we define the anti-pattern for variable $x = (x^1, x^2, \ldots, x^n)$ Î R^n as follows:

$$x^{\text{ant}} = \begin{cases} \min\limits_i x^i & \text{if} \quad x \in S, \\ \max\limits_i x^i & \text{if} \quad x \in D, \end{cases} \tag{2}$$

where S and D are sets of stimulants and destimulants, respectively.

After specifying the anti-pattern, we consider a new variable, $u = (u^1, u^2, \ldots, u^n)$, instead of variable x, given by

$$u^i = \frac{\left|x^i - x^{\text{ant}}\right|}{\sum_{j=1}^{n}\left|x^j - x^{\text{ant}}\right|} = \begin{cases} \dfrac{\left(x^i - x^{\text{ant}}\right)}{\sum_{j=1}^{n}\left(x^j - x^{\text{ant}}\right)} & \text{if} \quad x \in S, \\ \dfrac{\left(x^{\text{ant}} - x^i\right)}{\sum_{j=1}^{n}\left(x^{\text{ant}} - x^j\right)} & \text{if} \quad x \in D. \end{cases} \tag{3}$$

A normalized variable u has a clear interpretation. u^i specifies the share of distance between the i-th object and anti-pattern in the total distance of all objects from the anti-pattern. The i-th object is in a better situation when the value of u^i is higher. Therefore, the variable after normalization becomes a stimulant, regardless of its initial character. Other properties of anti-pattern normalization are presented below.

- Arithmetic mean:

$$\bar{u} = \frac{1}{n}.$$

- Standard deviation:

$$\mathrm{S}(u)=\begin{cases}\dfrac{\mathrm{S}(x)}{n(\bar{x}-x^{\mathrm{ant}})} & \text{if} \quad x\in S,\\ \dfrac{\mathrm{S}(x)}{n(x^{\mathrm{ant}}-\bar{x})} & \text{if} \quad x\in D.\end{cases}$$

- Skewness:

$$\mathrm{A}(u)=\begin{cases}\mathrm{A}(x) & \text{if} \quad x\in S,\\ -\mathrm{A}(x) & \text{if} \quad x\in D.\end{cases}$$

- Kurtosis:

$$\mathrm{K}(u)=\mathrm{K}(x).$$

- Pearson correlation coefficient:

$$\rho(u_1,u_2)=\begin{cases}\rho(x_1,x_2) & \text{if } x_1,x_2\in S \text{ or } x_1,x_2\in D,\\ -\rho(x_1,x_2) & \text{otherwise.}\end{cases}$$

The environmental context is often important in comparative studies. It is helpful to have a tool to consider this context. During scaling in version (1), the situation of other countries does not affect the value of *s* for a given country. This is not the case for normalization with respect to anti-pattern. The values of variable *u* reflect the situation of a country against the background of the whole environment. This is similar as for other forms of normalization, such as standardization, but the environment is specified in a different way. In anti-pattern normalization, the environment is represented by the sum of the distances between the objects and anti-pattern, whereas in common normalizations, the descriptive characteristics of the distribution of *x* are used for this purpose.[6]

The most important advantage of anti-pattern normalization is the comparability of variables in space as well as over time. If for the *i*-th country, the value of a normalized feature increases in time, it means that this country moves relatively away from the worst country such that its situation in the given environment improves. Thus, after a static anti-pattern normalization, the features are 'naturally' comparable over time. Therefore, this method is useful for dynamic studies, although only the observations from a current unit of time are used during normalization. Unlike for dynamic standardization, you do not have to recalculate all the results when you obtain new data for the next year.

[6] Because normalization is not the main topic of the article, we do not compare anti-pattern normalization with the other methods. However, note that many normalization methods exist, and in some of them, the values for other objects have an impact on the normalized value of a given object; for some newer cases, see for example, Lira et al. (2002) and Młodak (2006, 2014).

Description of empirical study

The empirical study illustrates the method of constructing the composite indicators described in Section 2. It highlights the differences in the results obtained for different reference groups.[7]

The relative socio-economic development of the EU10 countries is examined in the context of the Europe 2020 strategy. The development of individual countries is analysed against the background of other group members and the entire EU. An empirical study is conducted for the period 2010–2015. The data are taken from the statistical office of Poland (Statistics Poland 2018). All 11 indicators, that is, four stimulants and seven destimulants monitoring the implementation of the Europe 2020 strategy are used.[8] The indicators characterize the aspects of socio-economic development important from the viewpoint of the strategy.[9] For the study, they are assigned to the respective strategy priorities.

Group 1. Smart development:

x_1 – Gross domestic expenditure on R&D (% of GDP; $x_1 \in S$),

x_2 – Early leavers from education and training (%; $x_2 \in D$),

x_3 – Tertiary educational attainment of persons aged 30–34 (%; $x_3 \in S$).

Group 2. Sustainable development:

x_4 – Greenhouse gas emissions (1990 = 100; $x_4 \in D$),

x_5 – Share of renewables in gross final energy consumption (%; $x_5 \in S$),

x_6 – Consumption of primary energy (kgoe per 1,000 EUR; $x_6 \in D$).

Group 3. Inclusive development:

x_7 – Employment rate of persons aged 20–64 (%; $x_7 \in S$),

x_8 – Share of people at risk-of-poverty or social exclusion (%; $x_8 \in D$),

x_9 – People living in households with very low work intensity (%; $x_9 \in D$),

x_{10} – People at risk-of-poverty rate (after social transfers) (%; $x_{10} \in D$),

x_{11} – Severely materially deprived people (%; $x_{11} \in D$).

Based on the listed indicators, the development of the EU10 countries is examined, applying the methods described in Section 2.[10] The two approaches used for the normalization of variables allow for assessing the phenomenon from different

[7] For this purpose, the prototype of the proposed method, HDI, cannot be used, because some aspects of this index are characterized by only one indicator.

[8] The strategy sets targets for all indicators that should be implemented by individual countries as well as the entire EU. Therefore, none of the indicators can be omitted in the construction of a composite indicator.

[9] The presented results may differ significantly from other studies on socio-economic development of the EU owing to the use of specific indicators (Fura–Wang 2017).

[10] Because of this specific method of normalization, the results obtained are relative. Therefore, it is worth confronting them with results obtained on the basis of other types of normalization. However, this is beyond the scope of this article.

perspectives. The first approach considers only the environment of the EU10 countries, whereas the second approach takes all the EU countries (hereinafter EU28) into account.

Anti-pattern normalization results in 11 stimulants. They are labelled as u_1, u_2, …, u_{11}, respectively. Then the variables inside each group are aggregated using the arithmetic mean.[11] Thus, we obtain the following:

- a composite indicator of smart development

$$I_1 = \frac{u_1 + u_2 + u_3}{3}, \tag{4}$$

- a composite indicator of sustainable development

$$I_2 = \frac{u_4 + u_5 + u_6}{3}, \tag{5}$$

- a composite indicator of inclusive development

$$I_3 = \frac{u_7 + u_8 + u_9 + u_{10} + u_{11}}{5}. \tag{6}$$

Next, the composite indicators for individual groups are aggregated using the geometric mean. Thus, we obtain the following composite indicator of socio-economic development:

$$I = \sqrt[3]{I_1 \cdot I_2 \cdot I_3}. \tag{7}$$

This indicator synthetically characterizes the countries in terms of aspects relevant to the strategy.

Two versions of the composite indicators I_1, I_2, I_3, and I are created. They correspond to the different environments of the study. In one group, the EU10 countries are characterized against their own background, whereas in the other, we look at them from the perspective of the entire EU.

Empirical results

The results of the study described in Section 3 are shown in Appendix Tables 1–5. Table 1 presents the values of indicators monitoring the implementation of the strategy for Poland and the position of Poland in the EU10 countries based on a given indicator. Three versions of the indicator values are presented:

- before normalization,
- after anti-pattern normalization in the EU10 environment,
- after anti-pattern normalization in the EU28 environment.

The results for the other EU10 countries are not presented; however, 2015 showed an improvement in almost all raw indicators (the exception was $x_{10} \in D$) in most of the EU10 countries compared with 2010.

[11] Thus, we adopt a somewhat controversial assumption about the substitutability of variables within the group.

The next four tables present the composite indicators respectively, for the EU10 countries:

- I_1 – smart development,
- I_2 – sustainable development,
- I_3 – inclusive development,
- I – socio-economic development.

The tables contain the composite indicator values for both reference groups as well as the rankings derived from these indicators (in the EU28 environment, only the EU10 countries are classified; the remaining countries are omitted from the classification).

For better visualization, the socio-economic development results for the last year of the analysis (2015) are presented on maps. The countries are divided into three groups by the value of their composite indicator I (the borders are mean value +/– standard deviation). Figures 1 and 2 show the results for the EU10 and EU28 environments, respectively.

Figure 1

Socio-economic development calculated for the reference group of the EU10 countries, 2015

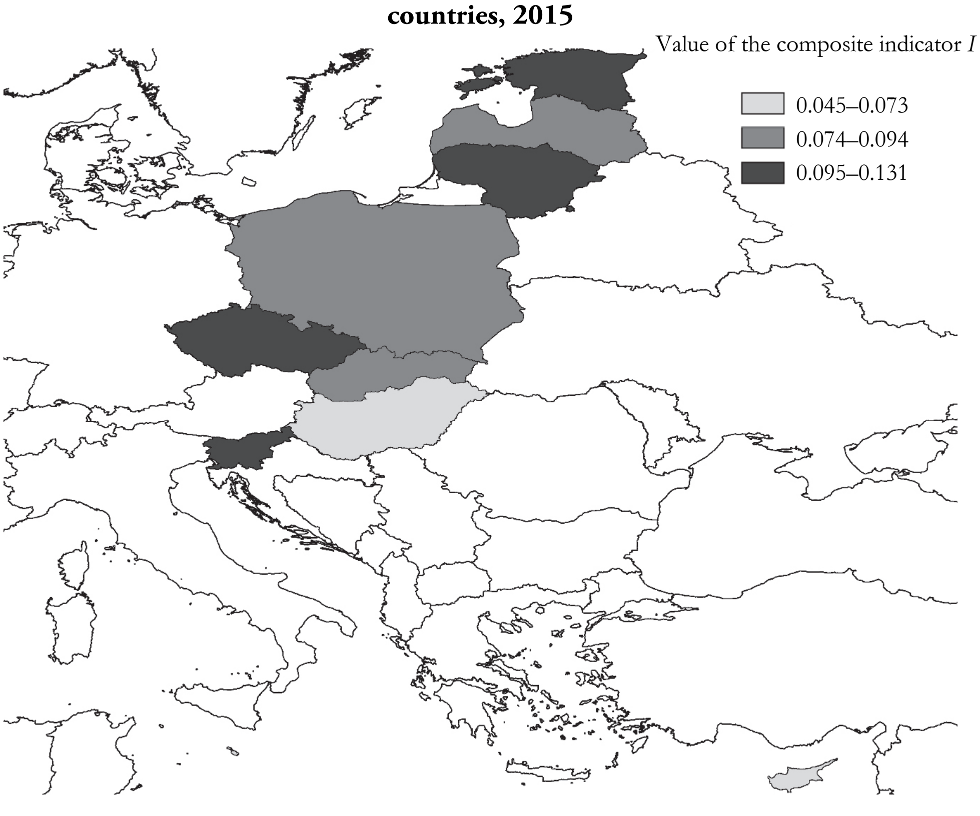
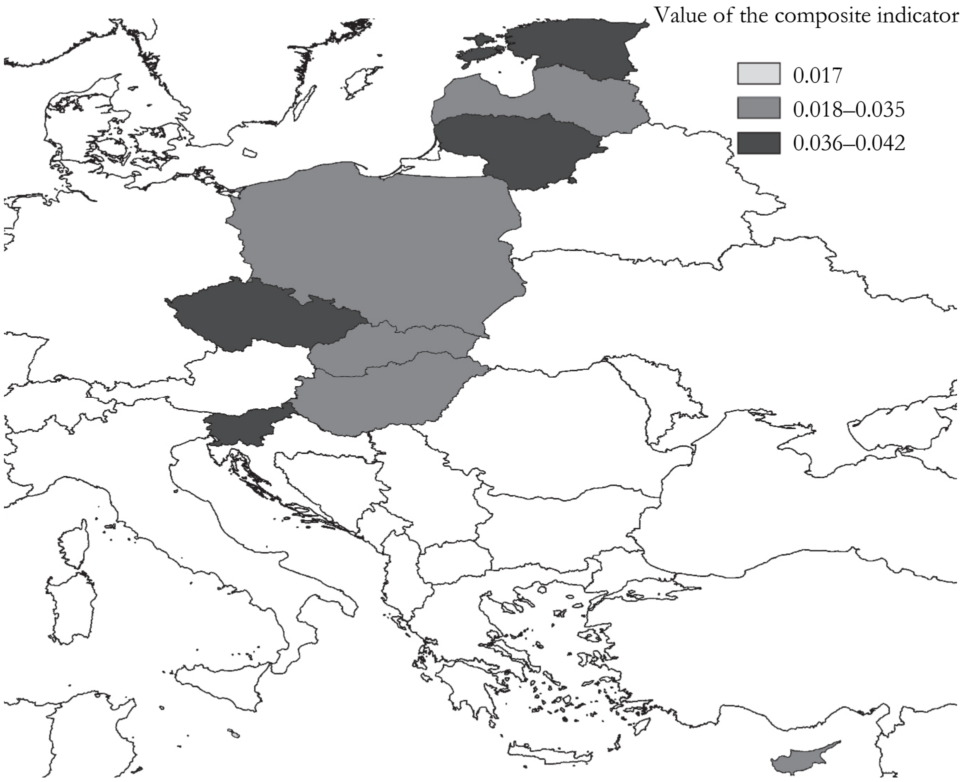


Figure 2

Socio-economic development calculated for the reference group of the EU28 countries, 2015



The synthetic analysis results for the remaining 18 EU countries are not presented here because these countries form only the background of the study. However, note that 12 countries show a decrease in the value of the socio-economic development measure *I* in the extreme years of the study (2010, 2015). For comparison, six out of the ten countries that joined the EU in 2004 showed improvement in this measure (compare Appendix Table 5). When all the EU countries are included in the classification (this classification is not presented), the EU10 countries are found to be a weaker group of countries at the time the strategy was introduced (in 2010). The exceptions are Slovenia in the fifth and Estonia in the tenth EU position. The remaining EU10 countries are in the second 10 (six countries) and third 10 (two countries). Half of the EU10 countries show an improvement in ranking after six years. The Czech Republic and Lithuania as well as Slovenia and Estonia are in the top 10 of the classification.

Discussion of empirical results

The results presented in Section 4 are analysed for similarity of approaches in the two environments. Even a quick analysis shows significantly different results for the two approaches. An obvious example is Cyprus, the first country in Table 5, whose socio-economic-development-based position among the EU10 countries in 2011 and 2012 differs in the two environments. The difference between the two environments is also evident when the countries are divided into groups by the level of development (compare Figure 1 and Figure 2). However, some countries show identical positions in the classifications or group allocations of the two approaches. Descriptive statistics tools are used to examine similarity in general.

For each analysed year and each composite indicator (I_1, I_2, I_3, I), the EU10 country rankings in the two environments are compared. The similarity of rankings is examined using Spearman's rho and Kendall's tau correlation coefficients. The obtained values are presented in Table 1.

Table 1

Compatibility of the results obtained for the two environments of the analysis

Coefficient	Development	2010	2011	2012	2013	2014	2015
Spearman's rho	Smart	0.988	0.976	0.952	0.976	0.891	0.927
	Sustainable	0.988	0.988	0.964	0.952	0.988	0.976
	Inclusive	1.000	0.988	0.988	0.964	0.830	0.855
	Socio-economic	0.915	0.794	0.830	0.927	0.964	0.915
Kendall's tau	Smart	0.956	0.933	0.911	0.933	0.844	0.889
	Sustainable	0.956	0.956	0.933	0.911	0.956	0.933
	Inclusive	1.000	0.978	0.978	0.933	0.822	0.800
	Socio-economic	0.822	0.733	0.778	0.800	0.844	0.778

Both rankings typically show strong (even full) compliance, but in some cases, the ranking compliance is much smaller. For most of the years, the general composite indicators are the least similar in classification. An interesting situation occurs for the third group (inclusive development); in this case, compatibility decreases over time.

In the next step, we analyse the stability of rankings between 2010 and 2015. Separately for each environment, we examine the compatibility of the rankings obtained for the six years. Kendall's W statistic is used for this purpose. The results are presented in Table 2.

Table 2

Stability of rankings over time, 2010–2015

Environment	Smart	Sustainable	Inclusive	Socio-economic
	development			
EU10	0.964	0.982	0.822	0.902
EU28	0.941	0.985	0.894	0.935

Note: We used Kendall's W statistic.

The rankings of the third group (inclusive development) are the least similar in the two environments. Therefore, the question whether the environments influence the stability of the results over time remained unanswered. For sustainable development, both environments showed similar results; for smart development, the EU10 showed more consistent results; and for inclusive and socio-economic development, the entire EU showed more compatible results.

Example of Poland

In order to show clearly the differences between the considered approaches, we carry out an in-depth analysis of the research results for Poland. First, we examine the influence of normalization in the two environments on the country's image based on individual indicators, priorities of strategy, and the level of socio-economic development. Then we analyse the study's initial results.

To begin with, the situation of Poland before and after normalization based on individual indicators is compared. The anti-pattern normalization of indicators does not affect the country's ranking, but it can change its 'dynamic image'. For example, the observed country's value of stimulants may increase over time, and then decrease after normalization. This occurs when the other countries' improvement is greater.[12] The results presented in Appendix Table 1 are examined in this context. The changes in the study's extreme years and the tendency in time are analysed. The raw data, the indicator values after normalization, and Poland's rankings are examined. The results are shown in Table 3.

[12] Such differences would not be visible for the scaling (1) mentioned in Section 2.

Table 3

Impact of normalization in various environments on Poland's dynamic image in terms of particular indicators monitoring the implementation of the Europe 2020 strategy

Indicator	Growth in 2015 compared with 2010				Year-to-year increment		
	Raw data	Normalized data		Rank difference	Raw data	Normalized data	
		EU10	EU28			EU10	EU28
x_1	+	+	+	+1	x	x	x
x_2	+	–	+	–1	x	x	x
x_3	+	0	+	0	+	x	x
x_4	+	–	–	0	x	x	x
x_5	+	–	–	+1	+	x	x
x_6	+	+	+	0	+	x	x
x_7	+	–	+	+2	+	–	+
x_8	+	+	+	–2	+	+	+
x_9	+	+	+	–1	x	x	+
x_{10}	0	+	+	–1	x	x	x
x_{11}	+	+	+	–3	x	+	+

Note: + better; – worse; 0 no change; x not determined (i.e. the tendency in time is not permanent).

From a comparison of the results for the study's extreme years, we find no raw indicator value worsening. However, some normalized indicators show a decline (i.e. decrease in the value of stimulants and increase in the value destimulants). We find several examples for this deterioration. Normalized variables x_4 and x_5 have worsening values in both environments. Thus, for these indicators, the improvement in extreme years for Poland was not as strong as that for other countries. The normalized values of x_2 and x_7 changed for the worse only in the EU10 environment. Thus, the situation (percentage of early leavers from education and training; employment rate of persons aged 20–64) in Poland improves relative to the whole EU, but only smaller improvement can be seen compared with the EU10. An interesting situation is observed for x_{10}. The value of the raw indicator did not change in the extreme years, but improved for both environments after normalisation. This means that in this respect, the situation (percentage of people at risk-of-poverty rate) in other countries is deteriorating.

Some of the described differences (before and after normalization) are reflected in the classification, but not all of them. For example, there is a drop in the ranking for x_5, so it is not surprising that its value falls after anti-pattern normalization. We can take x_{10} as another example. The position of Poland in the ranking for this indicator improved, so an increase in its normalized value is not a surprise. However, in the case of x_4 the deterioration of Poland's situation compared with other countries is not apparent in the classification, but in the fall in the normalized value of this indicator.

A year-by-year comparison of the results (raw data) shows that the situation in Poland is improving for five variables, but after normalization, only two of those variables show improvement in the EU10 environment. The most interesting results are given by x_7, whose raw values are increasing year by year, whereas after normalization, they increase in the EU28 environment but decrease in the EU10 environment. This variable shows especially the utility of normalization with respect to anti-pattern in the creation of composite indicators. Through application of this normalization, we can consider the different environmental contexts of the study and possibly come to different conclusions.

In the next step, we compare the synthetic results in terms of differences resulting from normalization in various environments. Such differences are found in the classification for some years. For example, the position of Poland in 2014 is identical for both environments only for the second priority (sustainable development). Interestingly, in this year, Poland shows a better position for I_1 (smart development) and I_3 (inclusive development) with the whole EU as a background, but generally (socio-economic development) its position is better for a smaller group of countries (EU10).

Then, just as for individual indicators, we analyse the dynamics of the socio-economic development and examine the composite indicators and ranking positions. The extreme years and year-to-year comparison results are shown in Table 4.

Table 4

Impact of normalization in various environments on the dynamic image of Poland in terms of strategy priorities and the level of socio-economic development

Composite indicator	Growth in 2015 compared with 2010				Year-to-year increment			
	Value		Position		Value		Position	
	EU10	EU28	EU10	EU28	EU10	EU28	EU10	EU28
I_1	+	+	+2	0	x	x	x	x
I_2	–	–	+1	0	x	x	x	0
I_3	+	+	–2	–2	x	+	x	x
I	+	+	0	0	x	x	x	0

Note: + better; – worse; 0 no change; x not determined (i.e. the tendency in time was not permanent).

Based on the synthetic analysis, the difference is less clear between the presented approaches than for individual indicators. It is only visible in the classification results when comparing the extreme years of the study, and not the values of synthetic measures. For I_1 and I_2, the position of Poland improves in the EU10 environment but does not change in the EU28 environment.

Conclusion

The article presents a new tool for multivariate comparative analysis – a composite indicator based on normalization with respect to anti-pattern. The proposed construction has important advantages. First, it can be adapted to dynamic research such that you do not have to recalculate the results for observations in a new time unit. Second, the analysis environment is significant, that is you can look at objects (countries, regions, etc.) from different perspectives.

An empirical illustration shows the utility of the new tool. The results obtained for different environments are similar, but not identical.

Anti-pattern normalization is particularly useful in dynamic synthetic studies, where the analysis environment is important. The obtained results are relative, that is, they characterize the situation of the object against the background of an established reference group, therefore it is good to confront them with the results obtained with other types of normalization.

Besides the presented construction, normalization with respect to pattern or anti-pattern can be used with other forms of aggregation. The construction in Müller-Frączek (2017) is based on the arithmetic mean, whereas the aggregation in Müller-Frączek (2018) follows Hellwig (1968). Studies are ongoing on the use of anti-pattern normalization in the Adjusted Mazziatto-Pareto Index (Mazziotta–Pareto 2016). This direction will be developed in future research.

Acknowledgement

I am grateful for the maps courtesy of Géza Tóth.

REFERENCES

BOOYSEN, F. (2002): An overview and evaluation of composite indices of development *Social Indicators Research* 59(2): 115–151. https://doi.org/10.1023/A:1016275505152

CASADIO TARABUSI, E.–GUARINI, G. (2013): An unbalance adjustment method for development indicators *Social Indicators Research* 112(1): 19–45. https://doi.org/10.1007/s11205-012-0070-4

EGRI, Z.–TÁNCZOS, T. (2018): The spatial peculiarities of economic and social convergence in Central and Eastern Europe *Regional Statistics* 8(1): 49–77. https://doi.org/10.15196/RS080108

FURA, B.–WANG, Q. (2017): The level of socioeconomic development of EU countries and the state of ISO 14001 certification *Quality & Quantity* 51(1): 103–119. https://doi.org/10.1007/s11135-015-0297-7

HELLWIG, Z. (1968): Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju i strukturę kwalifikowanych kadr *Przegląd Statystyczny* 15(4): 307–327.

JAJUGA, K.–WALESIAK, M. (2000): Standardisation of data set under different measurement scales In: DECKER, R.– GAUL, W. *Classification and Information Processing at the Turn of the Millennium* pp. 105–112., Springer, Berlin-Heidelberg.

KLUGMAN, J.–RODRÍGUEZ, F.–CHOI, H.-J. (2011): The HDI 2010: new controversies, old critiques *The Journal of Economic Inequality* 9(2): 249–288. https://doi.org/10.1007/s10888-011-9178-z

KUKUŁA, K.–BOGOCZ, D. (2014): Zero unitarization method and its application in ranking research in agriculture *Economic and Regional Studies* 7(3): 5–13.

LIRA, J.–WAGNER, W.–WYSOCKI, F. (2002): Mediana w zagadnieniach porządkowania liniowego obiektów wielocechowych In: PARADYSZ, J. *Statystyka regionalna w służbie samorządu lokalnego i biznesu* pp. 87–99. Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej, Poznań.

MAZZIOTTA, M.–PARETO, A. (2013): Methods for constructing composite indices: one for all or all for one *Rivista Italiana di Economia Demografia e Statistica* 67(2): 67–80.

MAZZIOTTA, M.–PARETO, A. (2016): On a generalized non-compensatory composite index for measuring socio-economic phenomena *Social Indicators Research* 127(3): 983–1003. https://doi.org/10.1007/s11205-015-0998-2

MILLIGAN, G.–COOPER, M. (1988): A study of standardization of variables in cluster analysis *Journal of Classification* 5(2): 181–204. https://doi.org/10.1007/BF01897163

MŁODAK, A. (2006): Multilateral normalizations of diagnostic features *Statistics in Transition* 7(5): 1125–1139. https://doi.org/10.1007/s00357-011-9088-6

MŁODAK, A. (2014): On the construction of an aggregated measure of the development of interval data *Computational Statistics* 29(5): 895–929. https://doi.org/10.1007/s00180-013-0469-7

MÜLLER-FRĄCZEK, I. (2017): Propozycja miary syntetycznej *Przegląd Statystyczny* 64(4): 413–428.

MÜLLER-FRĄCZEK, I. (2018): Dynamic measure of development In: PAPIEŻ, V.–ŚMIECH, S. *12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings* pp. 326–334., Foundation of the Cracow University of Economics, Kraków.

MÜLLER-FRĄCZEK, I. (2019): Pattern normalization – a new tool for dynamic comparisons *Statistika: Statistics and Economy Journal* 99(2): 182–197.

MÜLLER-FRĄCZEK, I.–MUSZYŃSKA, J. (2016): Implementation of Europe 2020 Strategy: the taxonomic analysis In: SIMBEROVA, I.–MILICHOVSKY, F.–ZIZLAVSKY, O. *Smart and Efficient Economy: Preparation for the Future Innovative Economy* pp. 346–355., University of Technology Faculty of Business and Management, Brno.

NARDO, M.–SAISANA, M.–SALTELLI, A.–TARANTOLA, S.–HOFFMAN, A.–GIOVANNINI, E. (2005): *Handbook on Constructing Composite Indicators* OECD Publishing, Washington.

SAISANA, M.–SALTELLI, A. (2011): Rankings and Ratings: Instructions for use *Hague Journal on the Rule of Law* 3(2): 247–268. https://doi.org/10.1017/S1876404511200058

SALTELLI, A. (2007): Composite indicators between analysis and advocacy *Social Indicators Research* 81(1): 65–77. https://doi.org/10.1007/s11205-006-0024-9

STATISTICS POLAND (2018): *Europe 2020 indicators.* https://stat.gov.pl/en/international-statistics/international-comparisons/tables-about-countries-by-subject/europe-2020-indicators/

TARANTOLA, S. (2008): European Innovation Scoreboard: strategies to measure country progress over time *JRC scientific and Technical Reports* EUR - Scientific and Technical Research Reports No. 23526, Brussels.

VALKÓ, G.–FEKETE-FARKAS, M.–KOVÁCS, I. (2017): Indicators for the economic dimension of sustainable agriculture in the European Union *Regional Statistics* 7(1): 179–196. https://doi.org/10.15196/RS07110

ZELIAŚ, A. (2002): Some notes on the selection of normalization of diagnostic variables *Statistics in Transition* 5(5): 787–802.

APPENDIX

Table 1

Individual indicators and corresponding positions – The example of Poland

Indicator		Normalization	2010	2011	2012	2013	2014	2015
x_1	Value	None	0.720	0.750	0.880	0.870	0.940	1.000
		EU10	0.050	0.041	0.056	0.053	0.063	0.071
		EU28	0.009	0.009	0.013	0.014	0.016	0.016
	Rank	Any	6	6	6	6	6	7
x_2	Value	None	5.400	5.600	5.700	5.600	5.400	5.300
		EU10	0.132	0.129	0.130	0.128	0.128	0.131
		EU28	0.051	0.050	0.050	0.049	0.049	0.052
	Rank	Any	4	4	4	3	2	3
x_3	Value	None	34.800	36.500	39.100	40.500	42.100	43.400
		EU10	0.122	0.122	0.128	0.126	0.127	0.122
		EU28	0.037	0.039	0.042	0.042	0.042	0.043
	Rank	Any	4	5	5	5	4	4
x_4	Value	None	87.170	87.050	85.500	84.750	82.110	82.760
		EU10	0.095	0.094	0.093	0.088	0.091	0.087
		EU28	0.037	0.036	0.036	0.035	0.035	0.034
	Rank	Any	7	7	7	7	7	7
x_5	Value	None	9.300	10.300	10.900	11.400	11.500	11.800
		EU10	0.062	0.063	0.060	0.057	0.052	0.051
		EU28	0.020	0.021	0.020	0.019	0.017	0.016
	Rank	Any	7	7	7	7	8	8
x_6	Value	None	278.300	265.300	252.800	250.300	233.300	227.300
		EU10	0.083	0.082	0.084	0.086	0.090	0.087
		EU28	0.024	0.026	0.027	0.025	0.027	0.028
	Rank	Any	8	8	8	8	8	8
x_7	Value	None	64.300	64.500	64.700	64.900	66.500	67.800
		EU10	0.072	0.064	0.054	0.040	0.019	0.003
		EU28	0.019	0.021	0.027	0.028	0.030	0.031
	Rank	Any	6	8	8	8	8	8
x_8	Value	None	27.800	27.200	26.700	25.800	24.700	23.400
		EU10	0.079	0.090	0.091	0.099	0.100	0.107
		EU28	0.031	0.033	0.034	0.035	0.036	0.038
	Rank	Any	7	7	6	6	5	5
x_9	Value	None	7.300	6.900	6.900	7.200	7.300	6.900
		EU10	0.132	0.158	0.145	0.138	0.141	0.144
		EU28	0.042	0.045	0.045	0.046	0.048	0.050
	Rank	Any	4	3	3	2	2	3
x_{10}	Value	None	17.600	17.700	17.100	17.300	17.000	17.600
		EU10	0.058	0.040	0.054	0.068	0.085	0.085
		EU28	0.025	0.027	0.031	0.031	0.035	0.033
	Rank	Any	8	8	7	7	7	7
x_{11}	Value	None	14.200	13.000	13.500	11.900	10.400	8.100
		EU10	0.094	0.104	0.106	0.112	0.115	0.129
		EU28	0.032	0.033	0.034	0.035	0.036	0.038
	Rank	Any	7	7	6	6	6	4

Note: Here and hereinafter, EU10 – countries that joined the EU in 2004; EU28 – all the EU countries. The indicator values are presented in three versions: before normalization, after anti-pattern normalization in the EU10 environment, and after anti-pattern normalization in the EU28 environment.

Source: Here and hereinafter, own calculation based on data from Statistics Poland.

Table 2

The composite indicator of smart development

Country	Composite indicator I_1	Environment	2010	2011	2012	2013	2014	2015
Cyprus	Value	EU10	0.097	0.099	0.100	0.096	0.109	0.114
		EU28	0.032	0.033	0.034	0.033	0.038	0.040
	Rank	EU10	6	4	6	6	5	4
		EU28	5	4	5	6	3	3
Czech Republic	Value	EU10	0.100	0.099	0.106	0.110	0.119	0.113
		EU28	0.029	0.032	0.034	0.034	0.035	0.035
	Rank	EU10	5	5	4	4	3	5
		EU28	6	6	6	5	5	6
Estonia	Value	EU10	0.156	0.170	0.146	0.136	0.115	0.115
		EU28	0.042	0.048	0.044	0.041	0.035	0.035
	Rank	EU10	2	2	2	3	4	3
		EU28	2	2	2	3	6	5
Hungary	Value	EU10	0.090	0.078	0.078	0.085	0.087	0.082
		EU28	0.026	0.026	0.026	0.028	0.028	0.026
	Rank	EU10	7	7	7	7	7	7
		EU28	7	8	8	8	8	7
Latvia	Value	EU10	0.070	0.078	0.076	0.079	0.079	0.072
		EU28	0.024	0.027	0.027	0.029	0.028	0.026
	Rank	EU10	8	8	8	8	8	9
		EU28	8	7	7	7	7	8
Lithuania	Value	EU10	0.124	0.128	0.129	0.136	0.140	0.146
		EU28	0.038	0.040	0.042	0.043	0.045	0.048
	Rank	EU10	3	3	3	2	2	2
		EU28	3	3	3	2	2	2
Malta	Value	EU10	0.015	0.011	0.020	0.013	0.010	0.013
		EU28	0.008	0.008	0.010	0.009	0.007	0.005
	Rank	EU10	10	10	10	10	10	10
		EU28	10	10	10	10	10	10
Poland	Value	EU10	0.101	0.098	0.105	0.103	0.106	0.108
		EU28	0.032	0.033	0.035	0.035	0.036	0.037
	Rank	EU10	4	6	5	5	6	6
		EU28	4	5	4	4	4	4
Slovakia	Value	EU10	0.061	0.054	0.060	0.059	0.058	0.073
		EU28	0.022	0.021	0.022	0.023	0.022	0.025
	Rank	EU10	9	9	9	9	9	8
		EU28	9	9	9	9	9	9
Slovenia	Value	EU10	0.185	0.185	0.181	0.184	0.176	0.164
		EU28	0.047	0.053	0.053	0.053	0.050	0.050
	Rank	EU10	1	1	1	1	1	1
		EU28	1	1	1	1	1	1

Note: In the EU28 environment, only the EU10 countries are classified; the remaining EU countries are omitted from the classification.

Table 3

The composite indicator of sustainable development

Country	Composite indicator I_2	Environment	2010	2011	2012	2013	2014	2015
Cyprus	Value	EU10	0.067	0.065	0.066	0.064	0.061	0.062
		EU28	0.018	0.017	0.017	0.018	0.017	0.017
	Rank	EU10	10	10	9	9	10	10
		EU28	10	10	9	9	10	10
Czech Republic	Value	EU10	0.087	0.087	0.087	0.088	0.090	0.087
		EU28	0.030	0.031	0.031	0.031	0.032	0.031
	Rank	EU10	7	7	7	7	7	7
		EU28	7	7	7	7	7	7
Estonia	Value	EU10	0.105	0.105	0.105	0.100	0.100	0.106
		EU28	0.039	0.041	0.041	0.037	0.037	0.041
	Rank	EU10	4	4	4	5	4	4
		EU28	3	3	3	3	3	3
Hungary	Value	EU10	0.099	0.100	0.104	0.106	0.099	0.091
		EU28	0.034	0.035	0.036	0.036	0.033	0.032
	Rank	EU10	5	5	5	4	5	6
		EU28	5	5	4	4	5	6
Latvia	Value	EU10	0.153	0.163	0.165	0.169	0.170	0.163
		EU28	0.051	0.055	0.056	0.057	0.056	0.055
	Rank	EU10	1	1	1	1	1	1
		EU28	1	1	1	1	1	1
Lithuania	Value	EU10	0.131	0.128	0.130	0.136	0.135	0.134
		EU28	0.044	0.044	0.044	0.046	0.045	0.046
	Rank	EU10	2	2	2	2	2	2
		EU28	2	2	2	2	2	2
Malta	Value	EU10	0.070	0.067	0.061	0.061	0.063	0.081
		EU28	0.020	0.018	0.016	0.018	0.018	0.023
	Rank	EU10	9	9	10	10	9	8
		EU28	9	9	10	10	9	9
Poland	Value	EU10	0.080	0.080	0.079	0.077	0.078	0.075
		EU28	0.027	0.028	0.027	0.026	0.026	0.026
	Rank	EU10	8	8	8	8	8	9
		EU28	8	8	8	8	8	8
Slovakia	Value	EU10	0.093	0.094	0.095	0.091	0.094	0.093
		EU28	0.032	0.033	0.033	0.032	0.032	0.033
	Rank	EU10	6	6	6	6	6	5
		EU28	6	6	6	6	6	5
Slovenia	Value	EU10	0.115	0.111	0.109	0.107	0.110	0.108
		EU28	0.036	0.035	0.035	0.034	0.035	0.035
	Rank	EU10	3	3	3	3	3	3
		EU28	4	4	5	5	4	4

Note: In the EU28 environment, only the EU10 countries are classified; the remaining EU countries are omitted from the classification.

Table 4

The composite indicator of inclusive development

Country	Composite indicator I_3	Environment	2010	2011	2012	2013	2014	2015
Cyprus	Value	EU10	0.150	0.150	0.120	0.096	0.081	0.038
		EU28	0.045	0.045	0.040	0.036	0.035	0.031
	Rank	EU10	3	2	4	6	8	10
		EU28	3	2	4	6	7	8
Czech Republic	Value	EU10	0.173	0.179	0.186	0.192	0.191	0.202
		EU28	0.052	0.052	0.051	0.053	0.052	0.053
	Rank	EU10	1	1	1	1	1	1
		EU28	1	1	1	1	1	1
Estonia	Value	EU10	0.110	0.105	0.118	0.127	0.126	0.147
		EU28	0.036	0.038	0.039	0.039	0.037	0.040
	Rank	EU10	5	5	5	3	2	2
		EU28	5	5	5	4	5	4
Hungary	Value	EU10	0.055	0.048	0.030	0.024	0.031	0.053
		EU28	0.028	0.026	0.027	0.027	0.027	0.033
	Rank	EU10	8	8	10	10	10	9
		EU28	8	9	9	10	9	7
Latvia	Value	EU10	0.014	0.021	0.032	0.054	0.057	0.062
		EU28	0.017	0.021	0.026	0.028	0.027	0.030
	Rank	EU10	10	10	9	9	9	8
		EU28	10	10	10	9	10	10
Lithuania	Value	EU10	0.048	0.044	0.054	0.066	0.098	0.069
		EU28	0.022	0.026	0.029	0.029	0.034	0.030
	Rank	EU10	9	9	8	8	5	7
		EU28	9	8	8	8	8	9
Malta	Value	EU10	0.092	0.097	0.100	0.097	0.085	0.085
		EU28	0.031	0.034	0.036	0.037	0.037	0.038
	Rank	EU10	6	6	6	5	7	6
		EU28	6	6	6	5	6	6
Poland	Value	EU10	0.087	0.091	0.090	0.091	0.092	0.094
		EU28	0.030	0.032	0.034	0.035	0.037	0.038
	Rank	EU10	7	7	7	7	6	5
		EU28	7	7	7	7	4	5
Slovakia	Value	EU10	0.119	0.125	0.127	0.124	0.121	0.122
		EU28	0.040	0.040	0.041	0.042	0.044	0.044
	Rank	EU10	4	4	3	4	3	4
		EU28	4	4	3	2	2	2
Slovenia	Value	EU10	0.151	0.140	0.144	0.128	0.118	0.128
		EU28	0.046	0.043	0.043	0.042	0.042	0.043
	Rank	EU10	2	3	2	2	4	3
		EU28	2	3	2	3	3	3

Note: In the EU28 environment, only the EU10 countries are classified; the remaining EU countries are omitted from the classification.

Table 5

The composite indicator of socio-economic development

Country	Composite indicator *I*	Environment	2010	2011	2012	2013	2014	2015
Cyprus	Value	EU10	0.099	0.099	0.093	0.084	0.081	0.065
		EU28	0.029	0.029	0.029	0.028	0.028	0.028
	Rank	EU10	4	4	5	8	8	9
		EU28	7	8	9	9	9	9
Czech Republic	Value	EU10	0.115	0.115	0.120	0.123	0.127	0.126
		EU28	0.036	0.037	0.038	0.038	0.039	0.039
	Rank	EU10	3	3	3	2	2	2
		EU28	3	3	4	4	3	3
Estonia	Value	EU10	0.122	0.124	0.121	0.120	0.113	0.121
		EU28	0.039	0.042	0.041	0.039	0.036	0.039
	Rank	EU10	2	2	2	3	4	3
		EU28	2	2	2	3	4	4
Hungary	Value	EU10	0.079	0.072	0.062	0.060	0.065	0.073
		EU28	0.029	0.029	0.029	0.030	0.029	0.030
	Rank	EU10	8	8	9	9	9	8
		EU28	8	9	8	8	8	8
Latvia	Value	EU10	0.054	0.064	0.074	0.090	0.091	0.090
		EU28	0.028	0.031	0.034	0.036	0.035	0.035
	Rank	EU10	9	9	8	6	6	7
		EU28	9	5	5	5	5	5
Lithuania	Value	EU10	0.092	0.090	0.097	0.107	0.123	0.111
		EU28	0.033	0.036	0.038	0.039	0.041	0.041
	Rank	EU10	5	5	4	4	3	4
		EU28	4	4	3	2	2	2
Malta	Value	EU10	0.045	0.041	0.050	0.043	0.038	0.045
		EU28	0.017	0.017	0.018	0.018	0.017	0.017
	Rank	EU10	10	10	10	10	10	10
		EU28	10	10	10	10	10	10
Poland	Value	EU10	0.089	0.089	0.091	0.090	0.091	0.091
		EU28	0.030	0.031	0.032	0.032	0.033	0.033
	Rank	EU10	6	6	6	5	5	6
		EU28	6	6	6	6	6	6
Slovakia	Value	EU10	0.088	0.086	0.090	0.087	0.087	0.094
		EU28	0.030	0.030	0.031	0.031	0.032	0.033
	Rank	EU10	7	7	7	7	7	5
		EU28	5	7	7	7	7	7
Slovenia	Value	EU10	0.148	0.142	0.141	0.136	0.132	0.131
		EU28	0.043	0.043	0.043	0.042	0.042	0.042
	Rank	EU10	1	1	1	1	1	1
		EU28	1	1	1	1	1	1

Note: In the EU28 environment, only the EU10 countries are classified; the remaining EU countries are omitted from the classification.

A literature review and categorisation of sustainability-aimed urban metabolism indicators: a context, indicator, mechanism, outcome analysis

Yan Song
Delft University of Technology
Department of Urbanism,
the Netherlands
E-mail: yan.song@tudelft.nl

Arjan van Timmeren
Delft University of Technology
Department of Urbanism,
the Netherlands
E-mail: A.vanTimmeren@tudelft.nl

Alexander Wandl
Delft University of Technology
Department of Urbanism,
the Netherlands
E-mail: A.Wandl@tudelft.nl

Keywords:
urban metabolism,
urban metabolism indicator,
sustainability,
CIMO approach

Urban metabolism has been advanced as an approach to quantifying energy and resource use and supply in the modern urban system. It is a multidisciplinary approach focused on providing insight into the behaviour of cities for drafting effective proposals for a more humane and ecologically responsible future. Urban metabolism indicators could play an important role in promoting the science and practice of urban metabolism for sustainability. This paper presents a systematic review of literature centred on defining sustainability-aimed urban metabolism indicators to improve the integration of urban metabolism and urban sustainability. Furthermore, this paper concentrates on two indicator sets (emergy synthesis and material flow analysis [MFA]), examining the relationship between these indicators and the three dimensions of sustainability (environment, economy, and society) in the literature. The paper thus builds a bridge between urban metabolism and urban sustainability in the hope that urban metabolism indicators can be used to measure and assess urban sustainability.

Introduction

With the onset of the Industrial Revolution and the rise of capitalism, the modern world has moved into an era of resource exploitation and intensity never seen before. To bring modern society's demands for energy, water, air, and other resources in line with the finite reserves of the earth, more needs to be done to quantify resource usage and to understand its political, economic, and ecological context.

One promising framework that has been advanced as an approach to quantifying energy and resource use and supply in modern society is "urban metabolism" (Ferrão–Fernandez 2013, Acebillo 2012). Wolman (1965) was the first to claim that the "metabolism" of a city comprises all the resources required by an urban system for economic production processes and the sum of waste streams emitted as a consequence of consumption. Urban metabolism can be defined as "the sum total of the technical and socio-economic processes that occur in cities, resulting in economic growth, production of energy, and elimination of waste" (Kennedy et al. 2007). In modern reference, urban metabolism has been distinguished as an analytical tool used to understand the essential energy, material, and waste streams between cities, their surrounding regions, and the planet. It is tangential to concepts of regenerative design, cradle-to-cradle design, and the emerging academic fields of industrial ecology and biomimicry (Richards et al. 1994, Benyus 2009, McDonough–Braungart 2002, van Timmeren 2013, Decker et al. 2000). Urban metabolism is an approach to modelling complex urban systems' material and energy streams as if the cities were organisms in the ecosystem (Fischer-Kowalski 2002, van Timmeren 2013). Urban metabolism thus forms a multi-disciplinary research domain that focuses on providing insights into the behaviour of cities for the purpose of advancing effective proposals for a more humane and ecologically responsible future.

Methodology and review of urban metabolism indicators for improving knowledge integration

The main objective of this paper is to present a systematic review of literature centred on categorising sustainability-aimed urban metabolism indicators to improve the integration of urban metabolism and urban sustainability. To achieve this, relevant research articles on urban metabolism were reviewed after searching through the Scopus database twice in October 2018. The literature was selected using three filters. First, 144 articles focusing on urban metabolism indicators and the concepts of sustainability were selected based on the content of their abstracts, titles, and keywords. Subsequently, these articles were filtered by subject area (environmental science and social science), source type (journals), document type (article), and language (due to language competence, two Spanish articles were excluded), resulting in 84 articles. Finally, these articles were read in depth and only those that provided specific indicator sets and mechanisms with sustainability concepts were selected, using qualitative content analysis. The articles that were excluded fell into the following categories: 1. no specific indicator set proposed; 2. sole focus on indicators in a limited research area; 3. indicator set only suitable in a specific site; and 4. an urban sustainability indicator set proposed rather than an urban metabolism indicator set. This filter process resulted in a total of 23 articles. Next, the adapted context, indicator, mechanism, outcome (CIMO) approach was applied to systematically capture the article information related to the main objective of the research. In our case, context (C) includes the research background and objective; indicator (I) is the quantifying item of each aspect, which is the intervention part of the original CIMO approach; mechanism (M) refers to the method of measuring or evaluating the indicator; and outcome (O) comprises the expected effects, which can be implemented in the general cases. Based on the results of the CIMO analysis, this paper concentrates on two indicator sets (emergy synthesis and MFA), analysing the relationship between these indicators and the three dimensions of sustainability (environment, economy, and society) in the literature. This could be the selection basis of sustainability-aimed urban metabolism indicators in future research.

The CIMO approach was used to process systematically the information in the 23 articles according to the objective of each paper. This approach originated from the domain of planning research (Soria-Lara et al. 2016, Straatemeier et al. 2010). The CIMO approach states that in a problematic context (C), the mechanism (M) can be used to explore generative intervention (I) to deliver some outcome (O) (Denyer et al. 2008). It offers a useful framework to identify and assess the mechanism and indicator sets in the selected literature. In this paper, we adopt the CIMO approach by using urban metabolism and sustainability indicators to represent I as shown in Table 1.

Table 1

Summary of mechanisms and indicators of urban metabolism and sustainability in the selected articles

	Context	Indicators	Mechanism	Outcomes
Barles (2009)	Presents the results of a research project aimed at a. examining the feasibility of MFA on a regional and urban scale in France; b. selecting the most appropriate method; c. identifying the available data; and d. calculating the material balance for a specific case	Balancing inputs and outputs, domestic material consumption, direct material input and output, local and exported processed output, the net addition to stock, total domestic output, total material input, total material output, and requirement	MFA	It reveals the need for new public policies, especially concerning waste management – to reduce construction material imports – and urban planning – to reduce their consumption. In addition, it states the need for more research and the development of action plans to link urban and agricultural policies to improve the use of urban fertilizers to favour local food supply.

(Continued on the next page.)

(Continued)

	Context	Indicators	Mechanism	Outcomes
Browne et al. (2012)	Seeks to apply a number of biophysical sustainability metrics to an Irish city-region to evaluate the effect of methodological pluralism when measuring urban sustainability and to determine the outcome of using more than one method when measuring the sustainability of the same system boundary at a city-region level	1. Measuring energy flows for a. solid fuels, including coking coal, steam coal, sub-bituminous coal, lignite/ brown coal, peat, oven and gas coke, patent fuel, and brown coal peat briquettes, b. oil, including crude oil, refinery feedstocks, and petroleum products, c. liquid natural gas; 2. Measuring energy and emissions metabolism estimates: a. TFC of energy in a particular sector, disaggregated by fuel type, b. total emissions from that sector, including greenhouse gas emissions and air pollution, c. the ratio of total emissions to TFC in a particular year	1. Energy flow accounting 2. Energy flow-metabolism ratio analysis	It develops an approach to measuring energy metabolism by outlining and applying the 'energy flow-metabolism ratio analysis' methodology, which is used to measure the ratio of greenhouse gas emissions as a function of energy material inputs.
Chen–Chen (2014)	Investigates a way to balance economic development and ecosystem health within a workable framework	1. Sets of MFA, life cycle analysis, exergy-based analysis, and emergy analysis; 2. Ecological network analysis sets	1. Element-based method 2. Structure-based method	It is an up-to-date inspection of integrating eco-indicators, which has both wide academic interest among interdisciplinary scientific boards and realistic application meaning for better urban management.
Chen–Wang (2014)	Gathers insights from global cities, identifies best practices internationally, and discusses how cities and regions can play a leading role in creating a sustainable society	1. A new multi-layered indicator set for urban metabolism studies: definition information (spatial boundaries, constitu-ent cities, population, economy), biophysical characteristics (climate, population density, building floor area), and metabolic flows (water, waste, materials, and all types of energy) of megacities; 2. Accounting scheme and its indicators from 13 flow elements and 9 fund elements	1. Multi-layered urban metabolism 2. MuSIASEM	It probes into the regulatory measures to optimise the configuration of water resources and to realize the integration of fundamental research innovation and management practice, thus, providing reasonable decision support for the nexus of water security, ecological security, and sustainable socio-economic development of cities and regions.

(Continued on the next page.)

(Continued)

	Context	Indicators	Mechanism	Outcomes
Chifari et al. (2017)	Presents a useful method for organizing a process of production and the use of scientific information in which both scientists and other social actors can have a bidirectional and constructive exchange of information	Occupied land, power capacity electrical machinery, power capacity thermal machinery, process heat consumption, electricity consumption, fuel consumption, water consumption, fixed investments, running costs, cost of exports, electricity revenue, recyclables revenue, subsidies for electricity production	MuSIASEM	Its approach provides a detailed characterization of the material balance of waste flows through the Municipal Solid Waste Management System.
Chrysoulakis et al. (2013)	Improves the communication of new biophysical knowledge to end-users (such as urban planners, architects, and engineers) with a focus on sustainable urban metabolism	The indicators set used in BRIDGE evaluations: a. energy, b. thermal comfort, c. water, d. greenhouse gases, e. land use, f. mobility/accessibility, g. social inclusion, h. human well-being, j. cost of proposed development, and k. effects on the local economy (employment and revenue)	Based on sustainability objectives and associated indicators addressing specific aspects of urban metabolism	It shows how a tool like the BRIDGE DSS may not simplify the urban planning process, but can help urban planners deal more adequately with its complexity. Although implementation of the DSS during planning processes may be constrained by lack of resources and skills at municipalities, practitioners can gain significant insight for more informed decision-making.
Geng et al. (2011)	Employs the MSIASM approach to evaluate regional societal and ecosystem metabolism in China	Hour-based human time, Joule-based exosomatic energy throughput, exosomatic metabolism rate, and bio-economic pressure	MSIASM	It indicates that the MSIASM method provides a feasible way for different levels of government to recognize the main barriers and challenges to development.
Goldstein et al. (2013)	Advances the ability to quantify environmental impacts of cities by modelling pressures embedded in the flows upstream (entering) and downstream (leaving) of the actual urban systems studied, and by introducing an advanced suite of indicators	Indicators of environmental exchanges (material and energy inputs, air, soil, water emissions, etc.) for the modelled processes	UM–LCA	It shows that the urban metabolism approach can be embedded within the process-based LCA framework, yielding a hybrid UM–LCA model that can provide a complete measurement of the environmental pressures exerted by a city.

(Continued on the next page.)

(Continued)

	Context	Indicators	Mechanism	Outcomes
González et al. (2013)	Enables the formulation of planning and policy recommendations to promote the efficient use of resources and enhance environmental quality in urban areas	Water (i.e. water balance, including evapotranspiration and run-off, and risk of flooding); air and climate (i.e. air quality in terms of pollutant concentration and dispersion; as well as CO_2 emissions, carbon sinks, and energy balance); and material assets (i.e. energy/fuel consumption and associated heat fluxes, including heat island effects)	Analytical hierarchical process multi-criteria assessment technique	It shows how the DSS can support impact assessment processes associated with the development and implementation of plans and projects, as well as contribute to monitoring and forecasting indicator performance in a planning context.
Hoekman–von Blottnitz (2017)	Contributes to the number of urban metabolism case studies using a standardized methodology	Domestic extraction used, imports, exports, domestic processed output, direct material input, domestic material consumption, physical trade balance, and direct material output	Economy-wide MFA	The study provides insights into the city's metabolism through various indicators including direct material input, domestic material consumption, and direct material output, among others.
Hoornweg et al. (2012)	Presents urban metabolism case studies, the data gathering challenges outlined, and the recommendations made as to how local governments can institutionalize the collection of metabolism information and use it to inform local sustainability programs and projects	Inflows, outflows, internal flows, storage and production of biomass, minerals, water, and energy	Abbreviated urban metabolism (a standardized listing of urban metabolism measures that ideally should be included in basic level reporting)	It states that by making citizens and companies more aware of their own impact on their city's metabolism, advances in information and communications technology and open data can help promote society-wide collaboration, smarter public decision-making, and a 'race to the top' to improve a city's resource efficiency and sustainability.
Huang–Hsu (2003)	Incorporates resource and MFA to investigate the Taipei area's urban sustainability due to urban construction	Indicators include the categories of a. intensity of resource consumption; b. inflow/outflow ratio; c. urban liveability; d. efficiency of urban metabolism; and e. emergy evaluation of urban metabolism	MFA and emergy synthesis analysis	It shows that the material flow accounting approach and the emergy evaluation of urban construction have important implications for evaluating the sustainability of urban development.

(Continued on the next page.)

(Continued)

	Context	Indicators	Mechanism	Outcomes
Inostroza (2014)	Proposes a new indicator to measure this process of material accumulation, namely, technomass	Technomass aspects (e.g. buildings, roads, cars, furniture, clothes, machines, and technological assets) and flows (e.g. water, food, energy, and supporting flows)	MFA	The study shows, in metabolic terms, how the indicator looks into the black box, providing the possibility of linking metabolic behaviours with urban forms and attempting to fill the gap between urban planning, urban metabolism, and MFA. This new indicator offers a broad scope of applications. Further possibilities and links to urban research and policy-making are explored in the discussion section.
Kennedy–Hoornweg (2012)	Presents a standardized, comprehensive urban metabolism framework and some degree of agreement on which parameters, out of the many possible, should ideally be included in basic level reporting	Inflows, outflows, internal flows, storage, and production of biomass, minerals, water, and energy	Urban metabolism framework	The study results indicate that the urban metabolism methodology is sufficiently robust, standardized, and practical to allow quick uptake by cities and ease of continued monitoring.
Kennedy et al. (2014)	Proposes a new 'multi-layered' indicator set for urban metabolism studies in megacities	Information on the definition (spatial boundaries, constituent cities, population, economy), biophysical characteristics (climate, population density, building floor area), and metabolic flows (water, waste, materials, and all types of energy) of megacities	Multi-layered urban metabolism indicator set	It shows that use of the standardized indicator set will ease inter-city comparisons of urban metabolism, while enhancing knowledge of megacities and their transformation into sustainable systems.
Kennedy et al. (2015)	Quantifies the energy and material flows of the world's 27 megacities, based on 2010 population, and identifies physical and economic characteristics that underlie the resource flows at multiple scales	Resource flows of electricity consumption, heating and industrial fuel use, ground transportation energy use, water consumption, waste generation, and steel production in terms of heating-degree-days, urban forms, economic activity, and population growth	MFA	It shows that overall energy and material flows vary considerably among megacities. It provides previously unidentified insights into the relationship between electricity consumption and urban forms.

(Continued on the next page.)

(Continued)

	Context	Indicators	Mechanism	Outcomes
Li et al. (2016)	Applies MFA in conjunction with specific socio-economic indicators to model urban metabolism and evaluate appropriate urban metabolism changes for the study case	Four major component inputs and outputs of the city: metals and industrial minerals, energy consumption, construction materials and biomass (predominantly from the surrounding farming areas)	MFA	The study shows that MFA techniques can be used as valuable tools for understanding urban metabolism, evaluating urban sustainability, and suggesting strategies for timely addressing urban sustainability issues.
Rosado et al. (2016)	Contributes to the discourse on urban area typology as well as to identifying urban metabolism characteristics	Eight urban metabolism characteristics: needs, accumulation, dependency, support, efficiency, diversity of processes, self-sufficiency, and pressure on the environment	MFA	It presents the extent of the imbalance between the types of materials extracted, consumed, and stocked, which makes urban areas vulnerable to external changes in resource supplies.
Sun et al. (2017)	Develops an integrated MFA and emergy evaluation model to investigate the environmental and ecological benefits of urban industrial symbiosis implementation	Urban statistics (urban level input and output flows), and micro level material and energy flow analysis (input and output flow within the symbiotic network)	Integrated MFA	This paper provides a useful modelling app-roach to understand the ecological benefits and trade-offs of local circular economy prac-tices and fundamental insights on natural capital accounting.
Yang et al. (2012)	Assesses resource exchanges and environmental emissions, urban household metabolism is investigated using an emergy synthesis framework	The emergy self-sufficiency ratio and the emergy investment ratio	Emergy synthesis analysis	It helps foster alternative household consumption strategies that could result in more equitable resource allocation and effective mitigation of cross-boundary environmental influences.
Yang et al. (2014)	Presents how creating sustainable cities has led to increasing concern over achieving healthy spatial metabolic interactions and system sustainability	Emergy-based indicators: renewable resources, non-renewable resources, local agriculture products, agricultural consumption, agricultural pollutants, residents' consumption, imports, exports	Emergy synthesis analysis	It shows how emergy synthesis can effectively integrate economic, social, and ecological dimensions and provide insights into cross-boundary metabolic interactions and system metabolic sustainability.

(Continued on the next page.)

(Continued)

	Context	Indicators	Mechanism	Outcomes
Zhai et al. (2018)	Combs through input-output analyses with ecological network analysis to help academics shed light on complicated system interactions and interior energy flows	Embodied ecological energy element intensity, direct integral flow control intensity, average mutual information, residual uncertainty	Energy ecological network model and Input-output analysis	This is a detailed study on the direction of energy; the flows uncover the relationship between social production activities and energy circulation. A thorough insight into robustness creatively provides a reference for improving the system efficiency.
Zhang et al. (2013)	Identifies the main metabolic actors responsible for these problems and analyses the characteristics of their metabolic structure.	Metabolic evaluation indicators: metabolic scale; metabolic intensity; metabolic efficiency (resources); metabolic impact (wastes).	The urban metabolic network model	It states that this improved resolution would provide a clearer picture of the network's characteristics, which cannot be represented accurately by small networks, such as the one example in the study, and would provide a more realistic simulation of an urban metabolic system.

Note: Results are listed alphabetically. MFA – material flow analysis; TFC – total final consumption; MuSIASEM – multi-scale integrated analysis of societal and ecosystem metabolism; MSIASM – multi-scale integrated analysis of societal metabolism; DSS – decision support system; UM–LCA – urban metabolism–life cycle assessment

From the CIMO literature review, there are two basic accounting and assessment mechanisms for urban metabolism indicators based on MFA and energy (emergy synthesis) analysis. Most recent urban metabolism mechanisms expand on or supplement these two mechanisms, such as the integrated MFA, the multi-layered urban metabolism indicator framework, the abbreviated urban metabolism, and the energy flow-metabolism ratio analysis (Sun et al. 2017, Kennedy et al. 2014, Kennedy–Hoornweg 2012, Hoornweg et al. 2012, Chen–Wang 2014, Browne et al. 2012). The MFA begins with material classification and concludes with a balance sheet that accounts for the categorised materials (Zhang et al. 2015). Similarly, emergy synthesis analysis starts with multiplying each flow of energy by its solar transformity and results in the assessment of emergy flow analysis (Zhang et al. 2009). Recently, researchers have begun to explore the possibilities of using life cycle assessment to account for and assess urban metabolism, which will be a further development of the model using consequential life cycle inventories (Zhang et al. 2013, Goldstein et al. 2013).

An approach for integrating urban metabolism indicators and sustainability

As urbanisation develops, so do environmental problems associated with it (Yang et al. 2017). Therefore, cities are seeking transformative methods to support sustainability in the future. To date, there are several urban-centric approaches that attempt to initiate radical innovations in this area such as the compact city (Dempsey 2010), smart growth (Kolbadi et al. 2015), the eco-city (Caprotti 2014), the zero-carbon city (Abbasi et al. 2012), the smart city (Townsend 2013), and the just city (Fainstein 2010). All of these schemes contain urban sustainability characteristics (Wei 2011, van Timmeren et al. 2015). The term 'sustainability' refers to a particular relationship between the human and environmental systems – one that ensures meeting human needs in the long term (World Commission on Environment and Development 1987, Alberti 1996). From the perspective of urban metabolism, a sustainable city is one in which the inflow of material and energy resources and the disposal of waste do not exceed the capacity of the city's surrounding environment (Kennedy et al. 2007). The aim of sustainability is to create the smallest possible ecological footprint and to produce the lowest quantity of pollution possible, to use land efficiently, compost used materials, recycle or convert waste to energy, and to make the city's overall contribution to climate change minimal (McCormick et al. 2013, Yang et al. 2017, Nassauer et al. 2014, Rotmans 2006). As a focus of sustainable development, urban sustainability has become increasingly prominent on political agendas and among scientific studies during recent decades, especially the indicator study that became a pronounced requirement of decision-makers (Huang et al. 2015, Shen et al. 2011, Wu 2014, Valkó et al. 2017). Based on the current study, researchers agree that sustainability depends on social, economic, and environmental factors (INTRASOFT International 2015, Sustainable Cities International 2012, Wu 2014). In the literature, many researchers have begun to explore urban metabolism within the context of urban sustainability (Li et al. 2016, Kennedy et al. 2014) (see Table 2).

Table 2

Summary of the relationship between urban metabolism indicator sets and urban sustainability

Urban metabolism indicator set	Urban sustainability factor		
	Environmental	Social	Economic
Material flow analysis	The efficient urban metabolism would first result in the built environment of the city (Voskamp et al. 2016, Huang–Hsu 2003, Kennedy et al. 2014). Resource and waste management are two key aspects among the environmental factors that are also MFA concerns (INTRASOFT International 2015, Huang et al. 2015, Mori–Christodoulou 2012, Kennedy et al. 2014).	Li et al. (2016) and Zhang (2013) use the structural decomposition of material flows to build a relationship between input/output with social wealth, which can depict the interindustry relationship of the whole economy (Szabó 2015). Dinarès (2014) also proposes social metabolism to question the apparent separation between human beings and their environment, the society-nature duality. Barles (2009) and Broto et al. (2011) attempt to integrate social aspects and influences on material and energy flows.	In the comprehensive framework for evaluating sustainability, Ness et al. (2010) and Li et al. (2016) implement economy-wide MFAs based on regional flows and non-integrated environmental pressure indicators. Furthermore, the decoupling model is a widely used method to analyse economic activities and their dependence on material consumption which can be utilized to build the relation between urban metabolism and the economy (Falb–Wolovich 1967, Li et al. 2016, Tapio 2005).
Emergy synthesis analysis	Metabolic flux references the structure of the metabolic flux in terms of resource consumption (Zhang et al. 2009, Huang–Hsu 2003). It expresses the amount of material and energy from within the urban metabolic system's internal environment as well as from its external environment.	Yang et al. (2014) and Lei et al. (2016) indicate that emergy synthesis can be adapted to quantify the flow of resources through complex ecological-socioeconomic systems. The indicator of metabolic efficiency reflects the resource utilization efficiency (i.e. the economic cost) of urban development (Zhang et al. 2009).	Economic metabolic activities can result in energetic interactions (Yang et al. 2014, Zhang et al. 2015). In addition, emergy products are useful to the economic system in the form of fuels, lubricants, and so on (Ulgiati et al. 1995).

Conclusion and future directions

After over 60 years of research, urban metabolism has been advanced as a promising approach for quantifying energy and resource use and supply in modern society. This paper investigates the most relevant urban metabolism mechanisms and indicators for improving urban sustainability. To that end, a literature review of relevant mechanisms and indicators in the field of urban metabolism and sustainability was conducted. The literature selection shows that a sizeable number of studies focus on urban metabolism, but only a limited number (23) explore indicators related to sustainability. These studies were reviewed following an adapted version of the CIMOapproach.

In analysing the literature, several findings arise: 1. most of the study objectives for urban metabolism lie in ecosystem health, energy, environmental technology, urban planning, waste management, and water technology; 2. in these articles, most of the proposed indicators are subject-oriented, which means there is still a lack of systematic indicator frameworks; 3. the most common methods for accounting urban metabolism are MFA and emergy synthesis analysis, which represent two main research streams in urban metabolism studies; 4. in the outcomes, most case studies do not explore the universal application of their research. The review also shows that there are relationships between urban metabolism and sustainability among environmental, social, and economic factors. Based on the integration of urban metabolism and sustainability, urban metabolism indicators can be used to build a connection between the two. This can provide a promising model for guiding urban development towards sustainability. Take MFA and emergy synthesis analysis as examples, the urban metabolism indicators can reflect urban sustainability in terms of environmental, social, and economic aspects.

The body of knowledge around urban metabolism is still growing. Indicator analysis, as one of the most common ways to assess organisational sustainable performance by municipalities, can collect specific quantitative and qualitative information on cities to enable comparisons of multiple areas (Mapar et al. 2017). Therefore, urban metabolism indicator analysis can be applied as an approach to assess sustainability. Future research directions on urban metabolism indicators could move in the following directions:

1. Quantitative correlation research on urban metabolism indicators with sustainability factors. The amount of the extant research implies the potential for using urban metabolism indicators to assess sustainability, for example, using energy flow accounting to measure urban sustainability (Browne et al. 2012). However, the correlation is not explored yet for all urban metabolism indicators.

2. Develop a standard classification system for stocks and flows, as Kennedy et al. (2011) mentions. Based on the review of urban metabolism literature, we found that the accounting methods and units vary among different studies. The non-standard classification differences can be a big barrier when comparing urban metabolism among multiple cities/regions.

3. Explore the application in urban design and planning. Several researchers attempt to connect urban metabolism to urban design and planning but most still focus on the process optimisation rather than quantifying resource flows using indicators. However, urban metabolism could be used to develop an approach that informs the design process for sustainability.

4. A comprehensive sustainability-aimed urban metabolism indicator system. To date, there is no comprehensive urban metabolism indicator list. However, an urban metabolism indicator list could be identified after a comprehensive selection by experts. It could be a useful tool for assessing the performance of urban metabolism to measure the shift in urban development towards sustainability.

Acknowledgements

The funding for this study was provided by the China Scholarship Council (CSC No. 201606260044). We would also like to thank the members of the Horizon 2020 Research and Innovation Action project REPAiR (REsource Management in Peri-urban Areas: Going Beyond Urban Metabolism) who helped collect the literature material and provided constructive comments.

REFERENCES

ABBASI, T.–PREMALATHA, M.–ABBASI, S. A. (2012): Masdar City: A Zero Carbon, Zero Waste Myth *Current Science* 102(1): 12.

ACEBILLO, J. (2012): *A New Urban Metabolism: Barcelona/Lugano Case Studies*. I. CUP, Accademia di architettura, Università della Svizzera Italiana; AB Publishers; Mendrisio, Barcelona, New York.

ALBERTI, M. (1996): Measuring Urban Sustainability *Environmental Impact Assessment Review* 16(4–6): 381–424. https://doi.org/10.1016/S0195-9255(96)00083-2

BARLES, S. (2009): Urban Metabolism of Paris and Its Region *Journal of Industrial Ecology* 13(6): 898–913. https://doi.org/10.1111/j.1530-9290.2009.00169.x.

BENYUS, J. M. (2009): *Biomimicry: Innovation Inspired by Nature* Harper Collins Publishers, New York. https://www.overdrive.com/search?q=D9D9103D-AD2D-41C4-B953-41DC68EF3899.

BROTO, V. C.–ALLEN, A.–RAPOPORT, E. (2011): Interdisciplinary Perspectives on Urban Metabolism *Journal of Industrial Ecology* 16(6): 851–861. https://doi.org/10.1111/j.1530-9290.2012.00556.x

BROWNE, D.–O'REGAN, B.–MOLES, R. (2012): Comparison of Energy Flow Accounting, Energy Flow Metabolism Ratio Analysis and Ecological Footprinting as Tools for Measuring Urban Sustainability: A Case-Study of an Irish City-Region *Ecological Economics* 83: 97–107. https://doi.org/10.1016/j.ecolecon.2012.08.006.

CAPROTTI, F. (2014): Critical Research on Eco-Cities? A Walk through the Sino-Singapore Tianjin Eco-City, China *Cities* 36: 10–17. https://doi.org/10.1016/j.cities.2013.08.005.

CHEN, B.–CHEN, S. (2014): Eco-Indicators for Urban Metabolism *Ecological Indicators* 47: 5–6. https://doi.org/10.1016/j.ecolind.2014.09.021.

CHEN, B.–WANG, R. (2014): Integrated Ecological Indicators for Sustainable Urban Ecosystem Evaluation and Management *Ecological Indicators* 47: 1–4. https://doi.org/10.1016/j.ecolind.2014.09.020.

CHIFARI, R.–RENNER, A.–LO PIANO, S.–RIPA, M.–BUKKENS, S. G.F.–GIAMPIETRO, M. (2017): Development of a Municipal Solid Waste Management Decision Support Tool for Naples, Italy *Journal of Cleaner Production* 161: 1032–1043. https://doi.org/10.1016/j.jclepro.2017.06.074.

CHRYSOULAKIS, N.–LOPES, M.–SAN JOSÉ, R.–GRIMMOND, C. S. B.–JONES, M. B.–MAGLIULO, V.– KLOSTERMANN, J. E. M. (2013): Sustainable Urban Metabolism as a Link between Bio-Physical Sciences and Urban Planning: The BRIDGE Project *Landscape and Urban Planning* 112(1): 100–117. https://doi.org/10.1016/j.landurbplan.2012.12.005.

DECKER, E. H.–ELLIOTT, S.–SMITH, F. A.–BLAKE, D. R.–ROWLAND, F. S. (2000): Energy and Material Flow Through the Urban Ecosystem *Annual Review of Energy and the Environment* 25: 685–740. https://doi.org/10.1146/annurev.energy.25.1.685.

DEMPSEY, N. (2010): Revisiting the Compact City? *Built Environment* 36(1): 5–8. https://doi.org/10.2148/benv.36.1.5.

DENYER, D.–TRANFIELD, D.–VAN AKEN, J. E. (2008): Developing Design Propositions through Research Synthesis *Organization Studies* 29(3): 393–413. https://doi.org/10.1177/0170840607088020.

DINARÈS, M. (2014): Urban Metabolism: A Review of Recent Literature on the Subject *Documents d'anàlisi Geogràfica* 60(3): 551–571. http://dx.doi.org/10.5565/rev/dag.134.

FAINSTEIN, S. S. (2010): *The Just City* Cornell University Press, Ithaca. https://login.proxy.bib.uottawa.ca/login?url=http://ebookcentral.proquest.com/lib/ottawa/detail.action?docID=3138037.

FALB, P. L.–WOLOVICH, W. A. (1967): Decoupling in the Design and Synthesis of Multivariable Control Systems *IEEE Transactions on Automatic Control* 12(6): 651–659. https://doi.org/10.1109/TAC.1967.1098737.

FERRÃO, P.–FERNANDEZ, J. E. (2013): *Sustainable Urban Metabolism* The MIT Press, Cambridge, London. http://site.ebrary.com/id/10752785.

FISCHER-KOWALSKI, M. (2002): Exploring the History of Industrial Metabolism. In: AYRES, R. U.–AYRES, L.: *A Handbook of Industrial Ecology* 1st Edition., pp. 35–45.: Edward Elgar. Cheltenham, Northampton.

GENG, Y.–LIU, Y.–LIU, D.–ZHAO, H.–XUE, B. (2011): Regional Societal and Ecosystem Metabolism Analysis in China: A Multi-Scale Integrated Analysis of Societal Metabolism (MSIASM) Approach *Energy* 36(8): 4799–4808. https://doi.org/10.1016/j.energy.2011.05.014.

GOLDSTEIN, B.–BIRKVED, M.–QUITZAU, M.-B.–HAUSCHILD, M. (2013): Quantification of Urban Metabolism through Coupling with the Life Cycle Assessment Framework: Concept Development and Case Study *Environmental Research Letters* 8(3): 1–14. https://doi.org/10.1088/1748-9326/8/3/035024.

GONZÁLEZ, A.–DONNELLY, A.–JONES, M.–CHRYSOULAKIS, N.–LOPES, M. (2013): A Decision-Support System for Sustainable Urban Metabolism in Europe *Environmental Impact Assessment Review* 38: 109–119. https://doi.org/10.1016/j.eiar.2012.06.007.

HOEKMAN, P.–VON BLOTTNITZ, H. (2017): Cape Town's Metabolism: Insights from a Material Flow Analysis *Journal of Industrial Ecology* 21(5): 1237–1249. https://doi.org/10.1111/jiec.12508.

HOORNWEG, D. A.–CAMPILLO, G.–LINDERS, D.–SUGAR, L.–SALDIVAR-SALI, A. N. (2012): Mainstreaming Urban Metabolism: Advances and Challenges in City Participation In: *World Bank Sixth Urban Research and Knowledge Symposium – Rethinking Cities* Barcelona.

HUANG, L.–WU, J.–YAN, L. (2015): Defining and Measuring Urban Sustainability: A Review of Indicators *Landscape Ecology* 30(7): 1175–1193. https://doi.org/10.1007/s10980-015-0208-2.

HUANG, S-L.–HSU, W-L. (2003): Materials Flow Analysis and Emergy Evaluation of Taipei's Urban Construction *Landscape and Urban Planning* 63(2): 61–74. https://doi.org/10.1016/S0169-2046(02)00152-4.

INOSTROZA, L. (2014): Measuring Urban Ecosystem Functions through 'Technomass' – A Novel Indicator to Assess Urban Metabolism *Ecological Indicators* 42: 10–19. https://doi.org/10.1016/j.ecolind.2014.02.035.

INTRASOFT INTERNATIONAL (2015): *Science for Environment Policy IN-DEPTH REPORT: Indicators for Sustainable Cities* European Commission, Bristol. https://doi.org/10.2779/61700.

KENNEDY, C.–STEWART, I. D.–IBRAHIM, N.–FACCHINI, A.–MELE, R. (2014): Developing a Multi-Layered Indicator Set for Urban Metabolism Studies in Megacities *Ecological Indicators* 47: 7–15. https://doi.org/10.1016/j.ecolind.2014.07.039.

KENNEDY, C. A.–CUDDIHY, J.–ENGEL-YAN, J. (2007): The Changing Metabolism of Cities *Journal of Industrial Ecology* 11(2): 43–59. https://doi.org/10.1162/jie.2007.1107.

KENNEDY, C. A.–HOORNWEG, D. (2012): Mainstreaming Urban Metabolism *Journal of Industrial Ecology* 16(6): 780–782. https://doi.org/10.1111/j.1530-9290.2012.00548.x.

KENNEDY, C. A.–STEWART, I.–FACCHINI, A.–CERSOSIMO, I.–MELE, R.–CHEN, B.–UDA, M. (2015): Energy and Material Flows of Megacities *Proceedings of the National Academy of Sciences* 112(19): 5985–5990. https://doi.org/10.1073/pnas.1504315112.

KENNEDY, C.–PINCETL, S.–BUNJE, P. (2011): The Study of Urban Metabolism and Its Applications to Urban Planning and Design *Environmental Pollution* 159(8–9): 1965–1973. https://doi.org/10.1016/j.envpol.2010.10.022.

KOLBADI, N.–MOHAMMADI, M.–NAMVAR, F. (2015): Smart Growth Theory as One of the Main Paradigms of Sustainable City *International Journal of Review in Life Sciences* 5(9): 209–219.

LEI, K.–LIU, L.–HU, D.–LOU, I. (2016): Mass, Energy, and Emergy Analysis of the Metabolism of Macao *Journal of Cleaner Production* 114: 160–170. https://doi.org/10.1016/j.jclepro.2015.05.099.

LI, Y.–BEETON, R. J. S.–HALOG, A.–SIGLER, T. (2016): Evaluating Urban Sustainability Potential Based on Material Flow Analysis of Inputs and Outputs: A Case Study in Jinchang City, China *Resources, Conservation and Recycling* 110: 87–98. https://doi.org/10.1016/j.resconrec.2016.03.023.

MAPAR, M.–JAFARI, M. J.–MANSOURI, N.–ARJMANDI, R.–AZIZINEJAD, R.–RAMOS, T. B. (2017): Sustainability Indicators for Municipalities of Megacities: Integrating

Health, Safety and Environmental Performance *Ecological Indicators* 83: 271–291. https://doi.org/10.1016/j.ecolind.2017.08.012.

MCCORMICK, K.–ANDERBERG, S.–COENEN, L.–NEIJ, L. (2013): Advancing Sustainable Urban Transformation *Journal of Cleaner Production* 50: 1–11. https://doi.org/10.1016/j.jclepro.2013.01.003.

MCDONOUGH, W.–BRAUNGART, M. (2002): *Cradle to Cradle: Remaking the Way We Make Things*. 1st Edition. North Point Press., New York.

MORI, K.–CHRISTODOULOU, A. (2012): Review of Sustainability Indices and Indicators: Towards a New City Sustainability Index (CSI) *Environmental Impact Assessment Review* 32(1): 94–106. https://doi.org/10.1016/j.eiar.2011.06.001.

NASSAUER, J. I.–WU, J.–XIANG, W.-N. (2014): Actionable Urban Ecology in China and the World: Integrating Ecology and Planning for Sustainable Cities *Landscape and Urban Planning* 125: 207–208. https://doi.org/10.1016/j.landurbplan.2014.02.022.

NESS, B.–ANDERBERG, S.–OLSSON, L. (2010): Structuring Problems in Sustainability Science: The Multi-Level DPSIR Framework *Geoforum* 41(3): 479–488. https://doi.org/10.1016/j.geoforum.2009.12.005.

RICHARDS, D. J.–ALLENBY, B. R.–FROSCH, R. A. (1994): The Greening of Industrial Ecosystems: Overview and Perspective In: ALLENBY, B. R.–RICHARDS, D. J *The Greening of Industrial Ecosystems* pp. 1–19. The National Academy Press, Washington D.C. https://doi.org/10.17226/2129.

ROSADO, L.–KALMYKOVA, Y.–PATRÍCIO, J. (2016): Urban Metabolism Profiles. An Empirical Analysis of the Material Flow Characteristics of Three Metropolitan Areas in Sweden *Journal of Cleaner Production* 126: 206–217. https://doi.org/10.1016/j.jclepro.2016.02.139.

ROTMANS, J. (2006): Tools for Integrated Sustainability Assessment: A Two-Track Approach *The Integrated Assessment Journal* 6(4): 35–57.

SHEN, L.-Y.–OCHOA, J. J.–SHAH, M. N.–ZHANG, X. (2011): The Application of Urban Sustainability Indicators – A Comparison between Various Practices *Habitat International* 35(1): 17–29. https://doi.org/10.1016/j.habitatint.2010.03.006.

SORIA-LARA, J. A.–BERTOLINI, L.–TE BRÖMMELSTROET, M. (2016): Towards a More Effective EIA in Transport Planning: A Literature Review to Derive Interventions and Mechanisms to Improve Knowledge Integration *Journal of Environmental Planning and Management* 60(5): 755–772. https://doi.org/10.1080/09640568.2016.1180282.

STRAATEMEIER, T.–BERTOLINI, L.–TE BRÖMMELSTROET, M.–HOETJES, P. (2010): An Experiential Approach to Research in Planning *Environment and Planning B: Planning and Design* 37(4): 578–591. https://doi.org/10.1068/b35122.

SUN, L.–LI, H.–DONG, L.–FANG, K.–REN, J.–GENG, Y.–FUJII, M.–ZHANG, W.–ZHANG, N.–LIU, Z. (2017): Eco-Benefits Assessment on Urban Industrial Symbiosis Based on Material Flows Analysis and Emergy Evaluation Approach: A Case of Liuzhou City, China *Resources, Conservation and Recycling* 119: 78–88. https://doi.org/10.1016/j.resconrec.2016.06.007.

SUSTAINABLE CITIES INTERNATIONAL (2012): *Indicators for Sustainability: How Cities Are Monitoring and Evaluating Their Success* Vancouver, Canada. www.cashewstory.com.

SZABÓ, N. (2015): Methods for Regionalizing Input-Output Tables *Regional Statistics* 5(1): 44–65. https://doi.org/10.15196/RS05103.

TAPIO, P. (2005): Towards a Theory of Decoupling: Degrees of Decoupling in the EU and the Case of Road Traffic in Finland between 1970 and 2001 *Transport Policy* 12(2): 137–151. https://doi.org/10.1016/j.tranpol.2005.01.001.

TOWNSEND, A. M. (2013): *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia* W. W. Norton & Company, Inc., New York.

ULGIATI, S.–BROWN, M. T.–BASTIANONI, S.–MARCHETTINI, N. (1995): Emergy-Based Indices and Ratios to Evaluate the Sustainable Use of Resources *Ecological Engineering* 5(4): 519–531. https://doi.org/10.1016/0925-8574(95)00043-7.

VALKÓ, G.–FEKETE-FARKAS, M.–KOVÁCS, I. (2017): Indicators for the Economic Dimension of Sustainable Agriculture in the European Union *Regional Statistics* 7(1): 179–196. https://doi.org/10.15196/RS07110.

VAN TIMMEREN, A. (2013): *Reciprocities: A Dynamic Equilibrium.* 2nd Edition. Delft University of Technology, Delft. https://doi.org/10.13140/2.1.3708.0964.

VAN TIMMEREN, A.–HENRIQUEZ, L.–REYNOLDS, A. (2015): *Ubikquity and the Illuminated City.* 2nd Edition. TU Delft Publication, Delft.

VOSKAMP, I. M.–STREMKE, S.–SPILLER, M.–PERROTTI, D.–VAN DER HOEK, J. P. –RIJNAARTS, H. H M. (2016): Enhanced Performance of the Eurostat Method for Comprehensive Assessment of Urban Metabolism: A Material Flow Analysis of Amsterdam *Journal of Industrial Ecology* 21(4): 887–902. https://doi.org/10.1111/jiec.12461.

WEI, H. (2011): Strategy of Urban Transformation in China *Journal of Urban and Regional Planning* 1: 3.

WOLMAN, A. (1965): The Metabolism of Cities *Scientific American* 213(3): 179–190. https://doi.org/10.1038/scientificamerican0965-178.

WORLD COMMISSION ON ENVIRONMENT AND DEVELOPMENT (1987): *Our Common Future* Oxford University Press, Oxford, New York.

WU, J. (2014): Urban Ecology and Sustainability: The State-of-the-Science and Future Directions *Landscape and Urban Planning* 125: 209–221. https://doi.org/10.1016/j.landurbplan.2014.01.018.

YANG, B.–XU, T.–SHI, L. (2017): Analysis on Sustainable Urban Development Levels and Trends in China's Cities *Journal of Cleaner Production* 141: 868–880. https://doi.org/10.1016/j.jclepro.2016.09.121.

YANG, D.–GAO, L.–XIAO, L.–WANG, R. (2012): Cross-Boundary Environmental Effects of Urban Household Metabolism Based on an Urban Spatial Conceptual Framework: A Comparative Case of Xiamen *Journal of Cleaner Production* 27: 1–10. https://doi.org/10.1016/j.jclepro.2011.12.033.

YANG, D.–KAO, W. T. M.–ZHANG, G.–ZHANG, N. (2014): Evaluating Spatiotemporal Differences and Sustainability of Xiamen Urban Metabolism Using Emergy Synthesis *Ecological Modelling* 272: 40–48. https://doi.org/10.1016/j.ecolmodel.2013.09.014.

ZHAI, M.–HUANG, G.–LIU, L.–SU, S. (2018): Dynamic Input-Output Analysis for Energy Metabolism System in the Province of Guangdong, China *Journal of Cleaner Production* 196: 747–762. https://doi.org/10.1016/j.jclepro.2018.06.084.

ZHANG, Y. (2013): Urban Metabolism: A Review of Research Methodologies *Environmental Pollution* 178: 463–473. https://doi.org/10.1016/j.envpol.2013.03.052.

ZHANG, Y.–LIU, H.–CHEN, B. (2013): Comprehensive Evaluation of the Structural Characteristics of an Urban Metabolic System: Model Development and a Case Study of Beijing *Ecological Modelling* 252: 106–113. https://doi.org/10.1016/j.ecolmodel.2012.08.017.

ZHANG, Y.–YANG, Z.–YU, X. (2009): Evaluation of Urban Metabolism Based on Emergy Synthesis: A Case Study for Beijing (China) *Ecological Modelling* 220: 1690–1696. https://doi.org/10.1016/j.ecolmodel.2009.04.002.

ZHANG, Y.–YANG, Z.–YU, X. (2015): Urban Metabolism: A Review of Current Knowledge and Directions for Future Study *Environmental Science and Technology* 49(19): 11247–11263. https://doi.org/10.1021/acs.est.5b03060.

The degree of urbanisation in Brazil*

Maria do Carmo Dias Bueno
Brazilian Institute of Geography and Statistics
Center for Information Documentation and Dissemination,
Brazil
E-mail: maria.bueno@ibge.gov.br

Ricardo Neves de Souza Lima
Brazilian Institute of Geography and Statistics
Center for Information Documentation and Dissemination,
Brazil
E-mail: ricardo.lima@ibge.gov.br

Urban and rural are two central concepts used by a wide range of policymakers, researchers, national administrations, and international organisations. An option for defining urban areas is the use of the degree of urbanisation, a measure that classifies an area within a region based on a range of factors including population size, population density, the degree and extent of the built-up area, and many other concepts.

A new approach proposed jointly by the Directorate-General for Regional and Urban Policy, European Commission (DG REGIO) and the Organisation for Economic Co-operation and Development (OECD) not only uses the traditional criterion of population density to determine the degree of urbanisation, but it also includes a contiguity rule, and it is applied to spatial units of the same size (1 km^2 grid square cells).

The authors have performed a comparison study using this new approach and two other methodologies, one proposed by the OECD in 2011 and the other developed by the Brazilian Institute of Geography and Statistics (IBGE) in 2017. The results demonstrate that the degree of urbanisation for 42% of the Brazilian municipalities was the same using any of the three methods. The results using the DG REGIO/OECD approach and the IBGE methodology matched for approximately 70% of the municipalities.

Concerning the differences observed in the results of the test, it is necessary to consider the purpose and scale of each method: while the IBGE methodology intends to provide local and detailed results, the purpose of the other two methodologies is global and more generic.

Keywords:
degree of urbanisation,
Brazil,
population grid

* Disclaimer: This paper should not be reported as representing the views of the Brazilian Institute of Geography and Statistics; the views expressed here are those of the authors.

Introduction

The traditional distinction between urban and rural areas within a country has assumed that urban areas, no matter how they are defined, provide a different way of life and usually a higher standard of living than rural areas. However, the criteria for defining areas that are urban vary widely from one country to another.

The simplest criterion used is population density, calculated by total population divided by the total land area. For small geographic areas, this measure relates closely to observable patterns of development and settlement. At higher levels of geography (municipalities, regions, and others), population density becomes a gross measure. Consequently, areas of low density can be masked by areas of very high density.

An alternative measure of defining urban areas is the degree of urbanisation. Areas within a region are classified as urban based on a range of factors including population size, population density, the degree and extent of built-up area, and many other concepts.

Another option for defining urban areas is using measurements of the extent or concentration of urban land use and land cover, such as impervious surfaces, built-up areas, or transportation network density. Unfortunately, these land use/land cover-based definitions are rarely linked to population-based definitions and limit the ability to bridge urban areas defined using these different approaches (Rain et al. 2007).

Usually, urban areas are defined by individual national governments according to internal administrative rules. This lack of a standard makes international data incomparable or at least increases the degree of uncertainty in measures derived from this information. Concerning the geographic unit used in rural-urban classification, the UN recommend, both for national purposes and for international comparability, the use of the locality. If this is not possible, the recommendation is the smallest administrative division of the country (UN 2014). However, these units can vary in character as well as in size between countries, within countries, and over time. This dependence on administratively defined geographies for classification of rural and urban sites makes comparisons very difficult.

Despite the differences in the definition of urban areas, some agencies that act globally need to make comparisons and evaluations between countries to define their policies in social and economic fields. In such cases, the establishment of an internal definition is an option. Therefore, to account for differences between rural and urban areas, the OECD established a regional typology, classifying local administrative areas as predominantly urban, intermediate, or predominantly rural, according to the share of the rural and urban population (OECD 2011). This typology is based essentially on the percentage of the population living in urban or rural localities and has proved to be meaningful in explaining the regional differences in economic and labour market performance. The problem with this methodology is that it relies on the population size of local administrative areas as the main criterion. If a local administrative area is large, even the presence of a dense city will lead to it being classified as thinly populated or a predominantly rural area.

A new approach proposed jointly by the DG REGIO and the OECD uses the criteria of population density and contiguity applied to units of analysis of the same size (1 km^2 square cells). As the grid cells are identical, this method eliminates the distortion of using areas that vary in size and produce results that are less distorted and more comparable (Dijkstra–Poelman 2014). Besides that, the grid-based perspective offers a useful basis for regional statistical analysis since many economic, social, and environmental phenomena do not strictly follow administrative boundaries and needs more flexible geographies (Brandmueller et al. 2017). This new methodology needs to be tested and validated to demonstrate that its global adoption is both valuable and suitable. The details and results of this assessment are detailed in the next section.

Study case

Goal

The goal of this study is to test the methodology for classifying the level of urbanisation proposed by the DG REGIO/OECD using the best available data in Brazil. The study compares the results of this new methodology to two others: the regional typology proposed by the OECD in 2011 and the recent national classification for municipalities released by the IBGE in 2017.

Data and methods

OECD Regional Typology

The OECD Regional Typology is based on population density and the size of the urban centres located within a region. The methodology is composed of three steps (OECD 2011):

1. the classification of local administrative units as rural if their population density is below 150 inhabitants/km^2;
2. the classification of local administrative units as predominantly urban, intermediate, and predominantly rural if the percentage of the rural population is lower than 15%, is equal or higher than 15% and lower than 50%, or is higher than 50%, respectively;
3. adjusting the classification based on the following rules:

– A unit classified as predominantly rural becomes intermediate if it contains an urban centre with more than 200,000 inhabitants, representing at least 25% of the regional population.

– A unit classified as intermediate becomes predominantly urban if it contains an urban centre with more than 500,000 inhabitants, representing at least 25% of the regional population.

Data used in this methodology is the number of people in the enumeration areas of the 2010 Population Census, classified as rural or urban. Following the recommendations of the UN for the population census, the IBGE adopts a simple classification of the enumeration areas for statistical purposes – rural and urban. Most of the inputs to establish this classification come from the current local legislation. For municipalities lacking any local law concerning rural-urban boundaries, the classification is based on observed local patterns of settlement. The sum of the population for each class of enumeration area inside a municipality gives the total urban and rural population for this administrative level.

DG REGIO/OECD methodology

The approach recommended by the European Commission and developed by DG REGIO with the cooperation of the OECD brings a new degree of urbanisation that uses not only population density as a criterion but also a spatial criterion of contiguity applied to square grid cells sized 1 km^2. The first step of the methodology is the definition of three regions (Dijkstra–Poelman 2014):

– high-density cluster: contiguous grid cells with a density of at least 1,500 inhabitants/km^2 and a minimum population of 50,000 inhabitants;

– urban cluster: contiguous grid cells with a density of at least 300 inhabitants/km^2 and a minimum population of 5,000 inhabitants;

– rural grid cells: grid cells outside urban and high-density clusters.

After the definition of these three regions, we can achieve the degree of urbanisation for an administrative area using the rules below (Dijkstra–Poelman 2014):

– densely populated area: an area in which at least 50% of the population is living in high-density clusters.

– intermediate density area: an area in which less than 50% of the population is living in rural grid cells, or less than 50% of the population lives in a high-density cluster.

– thinly populated area or rural area: an area in which at least 50% of the population is living in rural grid cells.

Population data used to apply this method comes from the Brazilian population grid released in 2015, with information from the 2010 Population Census. The population grid was generated using a hybrid method, a mix of bottom-up and top-down approaches. The bottom-up or aggregation approach uses population census microdata and their location attributes. The approach was put into practice through the combination of two types of data: geographic coordinates of households in rural areas and street block segments properly geocoded in urban areas. The disaggregation or top-down approach was used in cases where the aggregation was not possible due to the absence of location data. The disaggregation approach uses spatial and statistical methods, with or without ancillary data, for the spatial transfer of data from enumeration areas to grid cells. The ancillary data used were layers of land use-land cover classification in rural areas and layers of the road system in urban areas. The Brazilian Population Grid can be considered excellent for 80% of the country, where the bottom-up approach was used, and quite good for the other 20%, where the top-down approach was used (IBGE 2016).

Urban and rural sites of Brazil – IBGE classification for Brazilian municipalities

Recently, in July 2017, the IBGE published a national classification for municipalities concerning their degree of urbanisation (IBGE 2017) that aligned with the OECD Extended Regional Methodology (Brezzi et al. 2011) and other experience.

The new national classification was made according to successive groupings and matrix crossings based on the total population living in high-density areas, the share of the population in high-density areas, and location. Furthermore, to expand the understanding of urban and rural contexts, the methodology also uses accessibility to goods and services as a criterion for rural and urban space integration.

The share of the population living in high-density areas was calculated using the 1 km^2 population grid with figures of the 2010 Population Census. High-density areas are defined as areas with 300 inhabitants/km^2 and where the sum of the population of eight neighbouring cells reaches a minimum of 3,000 inhabitants.
Table 1 resumes the degree of urbanisation concerning the classes of municipalities.

After the application of these rules, another criterion was considered: the proximity to urban centres. This new rule allows the classification of intermediate and predominantly rural municipalities as adjacent or remote. This subclassification was not used in the test conducted.

Data input for this classification is the 2010 Population Census, the Brazilian 1 km^2 grid cells, and the commuting time using the transportation logistic map released by the IBGE in 2014.

Table 1

IBGE degree of urbanisation by classes of municipalities

Population in high-density areas, inhabitants	Share of population in high-density areas			
	More than 75%	More than 50 to 75%	25 to 50%	Less than 25%
More than 50,000	Predominantly urban	Predominantly urban	Predominantly urban	Predominantly urban
More than 25,000 to 50,000	Predominantly urban	Predominantly urban	Intermediate	Predominantly rural
More than 10,000 to 25,000	Predominantly urban	Intermediate	Predominantly rural	Predominantly rural
More than 3,000 to 10,000	Intermediate	Predominantly rural	Predominantly rural	Predominantly rural
Less than 3,000	Predominantly rural	Predominantly rural	Predominantly rural	Predominantly rural

Results

Table 2 resumes the results for the degree of urbanisation of Brazilian municipalities using the three methodologies according to the rules presented previously.

Using the OECD Regional Typology (Methodology A), we can observe in Table 2 that most of the Brazilian municipalities are classified as intermediate (49.9%) and the percentage of municipalities classified as intermediate is in the range from 40 to 60 in all the regions. Urban municipalities are the minority in Brazil (21.0%), as well as in the Northern, Northeast, and Southern regions. Only 29.1% of Brazilian municipalities are classified as rural, but these figures are expressive in the Northern, Northeast, and Southern regions, varying from 35.2 to 42.3%. The Southeast and Central-West regions present the lowest number of rural municipalities, 214 (12.8%) and 61 (13.1%), respectively.

Comparing these results with the reality, we can conclude that the urban class is over-estimated while the rural class is under-estimated. This disagreement is caused in most cases by the misrepresenting of the local legislation that tries to promote personal or commercial interests through the raising of municipal taxes which, in turn, increase the municipality budget. (In general, areas defined as urban pay higher taxes than those defined as rural.)

If the DG REGIO/OECD methodology (Methodology B) is applied, most of the municipalities are rural (56.2%), and the percentage of municipalities classified as rural is lower than 50% only in the Southeast region (47.4%), as we can see in Table 2. There are only 390 (7.0%) urban or densely populated municipalities in Brazil. All the regions of the country present low percentages of urban municipalities and only the Southeast region has a figure higher than 10%. An important number of municipalities are classified as intermediate: 2,047 (36.8%) and all the Brazilian regions present figures similar to this value.

Table 2

Degree of urbanisation of Brazilian municipalities, by three methodologies, 2010

Region	Degree of urbanisation					
	Rural		Intermediate		Urban	
	Number	Percentage	Number	Percentage	Number	Percentage
	of municipalities					
	Methodology A					
North	167	37.2	244	54.3	38	8.5
Northeast	759	42.3	899	50.1	136	7.6
Southeast	214	12.8	827	49.6	627	37.6
South	418	35.2	521	43.9	249	21.0
Central-West	61	13.1	285	61.2	120	25.8
Brazil	*1,619*	*29.1*	*2,776*	*49.9*	*1,170*	*21.0*
	Methodology B					
North	269	59.9	151	33.6	29	6.5
Northeast	1,106	61.6	610	34.0	78	4.3
Southeast	791	47.4	694	41.6	183	11.0
South	725	61.0	389	32.7	74	6.2
Central-West	237	50.9	203	43.6	26	5.6
Brazil	*3,128*	*56.2*	*2,047*	*36.8*	*390*	*7.0*
	Methodology C					
North	292	65.0	72	16.0	85	18.9
Northeast	1,236	68.9	251	14.0	307	17.1
Southeast	793	47.5	250	15.0	625	37.5
South	767	64.6	109	9.2	312	26.3
Central-West	275	59.0	64	13.7	127	27.3
Brazil	*3,363*	*60.4*	*746*	*13.4*	*1,456*	*26.2*

Note: Methodology A – OECD Regional Typology; Methodology B – DG REGIO/OECD methodology; Methodology C – IBGE national classification for municipalities. Here and in Table 4, deviations from 100.0% result from rounding.

The results for both densely and sparsely populated classes appear to align with reality. However, the intermediate class is very wide and encompass different urbanisation 'nuances'. For instance, Mathias Lobato, a municipality with little more than 3,000 inhabitants in the state of Minas Gerais, is classified as intermediate as well as Palmas, the capital of the state of Tocantins, which is home to 230,000 people.

The IBGE national classification for municipalities (Methodology C) demonstrates that most of the municipalities in Brazil are rural (60.4%) (see Table 2). This predominance can be observed in the whole country, with percentages higher than 50% except for the Southeast region where that is 47.4%. Intermediate municipalities are the minority in the country as well as in all the Brazilian regions. The values are approximately 9.2% for the Southern region and 16% for the Northern region. The percentage of municipalities classified as urban varies from 17.1 in the Northeast region to 37.5 in the Southeast region; the value for Brazil is 26.2.

The results suggest that the urban class is over-estimated. We can observe municipalities with little more than 11,000 inhabitants classified as rural in Methodologies A and B – such as Canápolis in the state of Minas Gerais – being considered predominantly urban in Methodology C alongside the largest city of the country, São Paulo, with more than 11 million inhabitants.

According to Table 3 that compares the matching between the three methodologies evaluated, A and B match C in approximately 70% cases and A matches B to a lower degree (approximately 60%).

For 2,328 municipalities (41.83%) the degree of urbanisation using the three methods are the same, while 3,237 municipalities (58.17%) present differences in the results.

Table 3

Matching of the three methodologies

	B		C	
	Match	No match	Match	No match
A	58.4	41.6	69.1	30.9
B	–	–	68.6	31.4

Note: Methodology A – OECD Regional Typology; Methodology B – DG REGIO/OECD methodology; Methodology C – IBGE national classification for municipalities.

We can suggest that the differences between A and B, as well as A and C, are due to imprecisions in rural-urban site definitions as the rules used in methodology A are based on different criteria, which are not precisely equal for all municipalities.

Surprisingly the greatest matching is between A and C: 69.1% or 3,843 municipalities and we believe that further investigation is necessary to explain this number.

The matching between B and C is also great (68.6%) and this means that 3,816 municipalities have the same degree of urbanisation using both methodologies. Looking deeply into the differences between B and C, we computed the changes in each level of the degree of urbanisation (see Table 4).

The differences between the rural and intermediate degrees calculated by methodologies B and C can be due to a tough limit line between classes. A density slightly higher or slightly lower than 300 inhabitants/km^2 can probably change the degree of urbanisation from rural to intermediate (1-2) or from intermediate to rural (2-1).

The values for the change of the degree of urbanisation from intermediate to urban for the Southeast and Central-West regions (22.0% and 20.0%, respectively) are striking, although all the regions present high values in this respect. We suggest that these differences are due to the rules used to define the clusters in C, but it should be further investigated.

Table 4

Changes in the degree of urbanisation using methodologies B and C

Region	1-2		1-3		2-1		2-3		0-0	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
	of changes									
North	14	3.1	2	0.4	39	8.7	54	12.0	340	75.7
Northeast	41	2.3	18	1.0	189	10.5	211	11.8	1,335	74.4
Southeast	64	3.8	75	4.5	141	8.5	367	22.0	1,021	61.2
South	22	1.9	45	3.8	109	9.2	193	16.2	819	68.9
Central-West	9	1.9	8	1.7	55	11.8	93	20.0	301	64.6
Brazil	*150*	2.7	*148*	2.7	*533*	9.6	*918*	16.5	*3,816*	68.6

Note: Methodology B – DG REGIO/OECD methodology; Methodology C – IBGE national classification for municipalities. Changes: 1-2 – from rural to intermediate; 1-3 – from rural to urban; 2-1 – from intermediate to rural; 2-3 – from intermediate to urban; 0-0 – the same degree of urbanisation.

Conclusion

As a general conclusion, we can point out that approaches using grid population data and rules based on population density and contiguity - like the methodology developed by the DG REGIO/OECD - offer a useful basis for regional statistics, once that comparisons between different countries or regions are crucial. The data generated using this methodology could be useful, for instance, for generating indicators for the UN's Sustainable Development Goals and for functional geographies, like large urban clusters that transpose administrative borders, as well as for studying the causes and effects of many socioeconomic phenomena, such as commuting and labour, and environmental phenomena, such as flooding and air quality.

Although the DG REGIO/OECD and the IBGE methodologies have many similarities, there were differences in the results. It is suggested that the main reason for these differences is conceptual, as the rules used in each methodology tries to materialize different concepts of urban and rural. Additionally, it is necessary to consider the purpose of the development and scale of each method. The Brazilian methodology intends to provide local detailed results that are adjusted to the national reality. On the other side, the purpose of the DG REGIO/OECD methodology is global and generic, as they intend to be applied globally and therefore needs to be suitable to many different configurations of human settlements.

Despite these differences, the methodology is useful and can help in understanding and solving different challenges (economic, demographic, social or environmental). The strength of the methodology is the possibility of applying it to any region of the world - as there is freely available population data for the whole world - and to a specific location, using more accurate population data released by official data producers.

We intend to continue the evaluation of the methodology proposed by the DG REGIO/OECD as it moves beyond the determination of the degree of urbanisation: it also deals with the definition of cities and functional urban areas.

REFERENCES

BRANDMUELLER, T.–SCHÄFER, G.–PETRI, E.–MÜLLER, O. (2017): Territorial indicators for policy purposes: NUTS regions and beyond *Regional Statistics* 7(1): 78–89. https://doi.org/10.15196/RS07105

BREZZI, M.–DIJKSTRA, L.–RUIZ, V. (2011): *OECD Extended Regional Typology: The Economic Performance of Remote Rural Regions* OECD Regional Development Working Papers 2011/06, OECD Publishing. http://dx.doi.org/10.1787/5kg6z83tw7f4-en

DIJKSTRA, L.–POELMAN, H. (2014): *A harmonized definition of cities and rural areas: the new degree of urbanisation* European Commission Regional Working Paper 01/2014. http://ec.europa.eu/regional_policy/sources/docgener/work/2014_01_new_urban.pdf

IBGE (2016): *Grade Estatística* IBGE, Rio de Janeiro. ftp://geoftp.ibge.gov.br/recortes_para_fins_estatisticos/grade_estatistica/censo_2010/grade_estatistica.pdf

IBGE (2017): *Classificação e caracterização dos espaços rurais e urbanos do Brasil: Uma primeira aproximação* IBGE, Rio de Janeiro. https://biblioteca.ibge.gov.br/visualizacao/livros/liv100643.pdf

OECD (Organisation for Economic Co-operation and Development) (2011): *Regional Typology* Directorate for Public Governance and Territorial Development. Available on http://www.oecd.org/governance/regional-policy/42392595.pdf

RAIN, D. R.–LONG, J. F.–RATCLIFFE, M. R. (2007): Measuring population pressure on the landscape: comparative GIS studies in China, India, and the United States *Population and Environment* 28: 321–336 https://doi.org/10.1007/s11111-007-0055-4

UN (United Nations) (2014): *Principles and Recommendations for a Vital Statistics System*, Rev. 3. n https://unstats.un.org/unsd/demographic/standmeth/principles/M19Rev3en.pdf

Appendix

Figure 1

Degree of urbanisation using OECD Regional Typology (Methodology A)

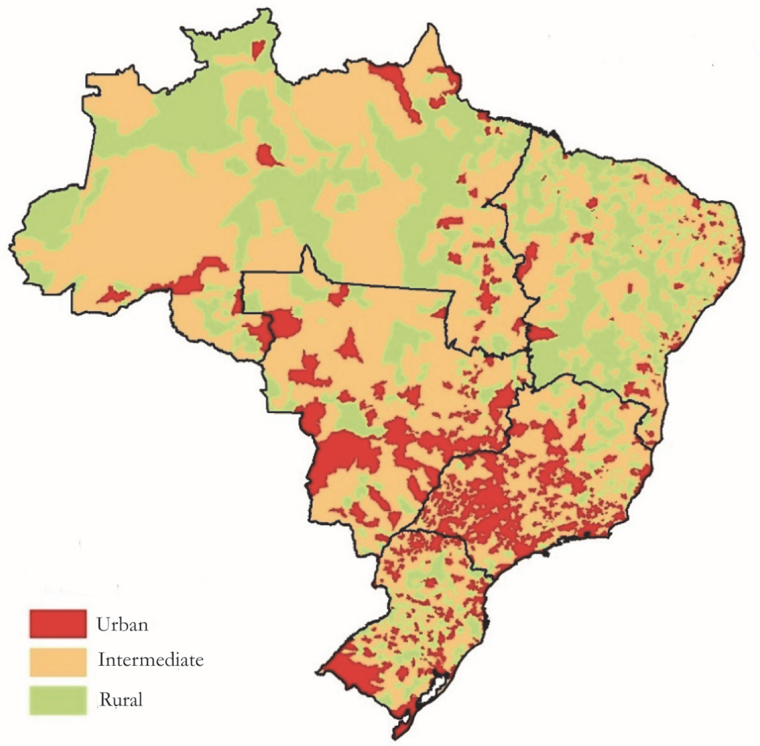


Figure 2

Degree of urbanisation using DG REGIO/OECD methodology (Methodology B)

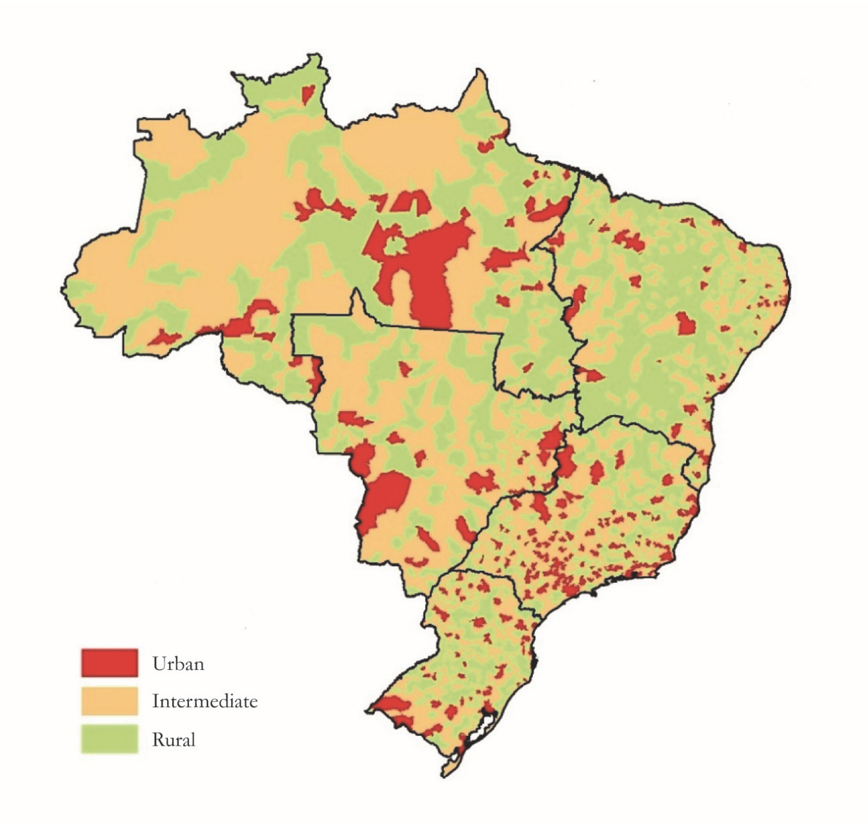
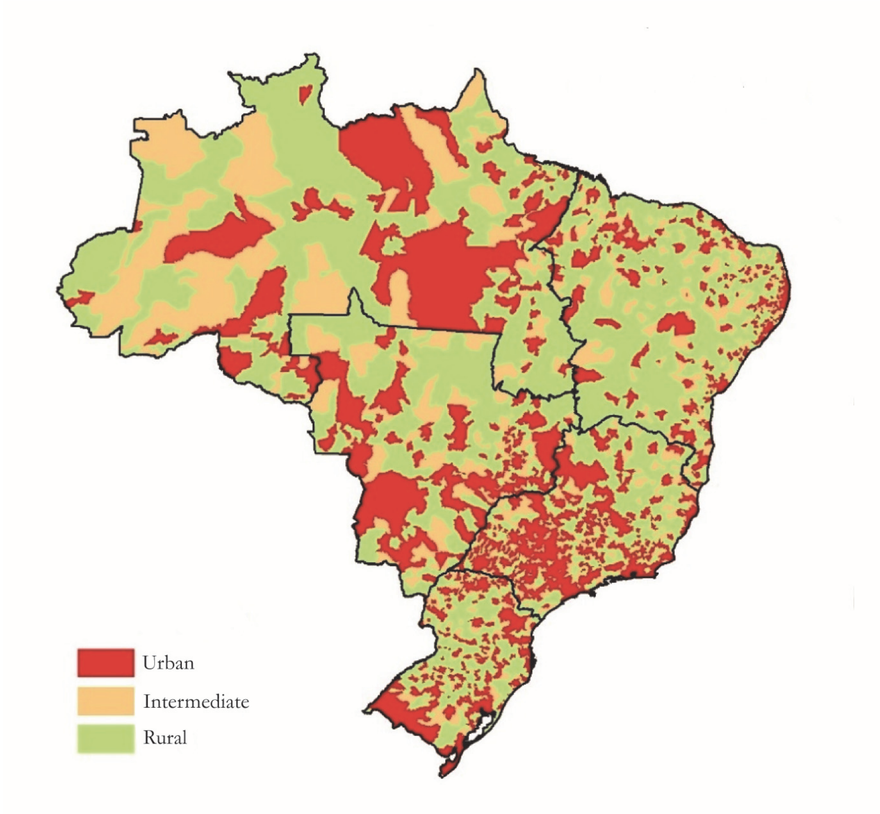


Figure 3

Degree of urbanisation using the IBGE national classification for municipalities (Methodology C)



Deprivation amidst affluence in 'rising India': Impacts of the National Rural Employment Guarantee Scheme*

Buddhadeb Ghosh
Evaluation Study on Mahatma Gandhi National Rural Employment Guarantee Act and Border Area Development Programme Planning Commission, Economic Research Unit Indian Statistical Institute Government of India, India
E-mail: buddhadeb_ghosh@yahoo.com buddhadebghosh.om@gmail.com

Sankar Narayan Samanta
Atos Global IT Solutions and Services Pvt. Ltd, India
E-mail: sankar.econ@gmail.com

Nilachal Ray
Ministry of Statistics and Programme Implementation Government of India, India
E-mail: nray111@yahoo.co.in

This study examines regional variations in spatial inequality measured by consumption expenditure using National Sample Survey Office data and social indicators from the census of India. It classifies rural people from survey-based, unit-level household data into bottom 25%, middle 50%, and top 25% economic quartiles. The observations include 559 common districts over time across rural India. Contemporary studies on regional development are based on the 'state' as the unit of analysis. The main objective of this study is to investigate whether sub-state disparities increased during the highest growth phase in India since 2004/2005, thereby helping determine whether the Mahatma Gandhi National Rural Employment Guarantee Act had any lasting impact on the living standards of utterly deprived people beyond the state capitals. In addition to some non-parametric tests, the study conducts alternative multivariate regressions to investigate the factors responsible for rising deprivation and the impact of policy. The statistical evidence is highly alarming and warrant urgent policy initiatives to reach the poor living in alienated regions.

* The paper was presented at Dibrugarh University, Assam in August 2014; at IIT Kharagpur, West Bengal in November 2014; at the Institute of Rural Management, Anand, Gujarat in December 2014; at Chennai Presidency College on its 175th Year Celebration, Tamil Nadu in March 2015; at the Department of Business Management, Tezpur University, Assam in November 2015; and at the Department of Rural Management, Tripura University, Tripura, in January 2016. The authors are indebted to all commentators. They are responsible for any remaining errors.

Arpita Chatterjee
University of New South Wales,
Australia
E-mail: chatterjee.arp@gmail.com

Keywords:
bottom quartile,
consumption expenditure,
disparity,
social development index,
Cook's distance,
Indian economy

Introduction

When 'On Economic Inequality' was first published in 1973, Amartya Sen dedicated the book to his daughters, as follows: 'To Antara and Nandana with the hope that when they grow up they will find less of it no matter how they decide to measure it.' There is no doubt that he was acutely aware of the extreme poverty and inequality in his home country of India, particularly in Bengal, during the 1940s and 1950s. However, the tragedies of deprivation and rising social unrest in vast rural regions have remained unchanged. The traditional approach of poverty (the percentage of people below the government-defined poverty line) has lost relevance since it is based on the daily availability of two meals and 2,400 calories and does not include, for example, the costs of education, health, clothing, etc. Thus, Sen's 'hope' remains unfulfilled, irrespective of whether we choose to measure inequality in terms of 'economic', 'social', or 'participatory' factors.[1] Interestingly, India is the largest 'electoral' democracy with the largest number of poor people and the highest 'invisible foreign exchange earnings'.[2]

This study is the first attempt to investigate the impact of the largest rural development project in human history, the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), undertaken since 2005/2006 by the Government of India to address the economic conditions of rural people, particularly the vulnerable class, across all rural districts in India. Additionally, it seeks to identify

[1] Eswaran and Kotwal (1999) show systematically how and why India has long suffered from the 'vicious circle of poverty' trap.

[2] Many studies prove that conventional models of formal economic development nurtured in textbooks should not be forcefully applied across India. India's diversities accommodate all possible disparities known to the world. The average does not hold, either for the nation or for the states. Professor Joan Robinson correctly remarked in the 1960s, 'Whatever you can rightly say about India, the opposite is also true'. Shashi Tharoor (2007) also rightly stated, '…the paradoxes that is today's India – a country that, as I wrote many years ago, manages to live in several centuries at the same time' (p. 15.).

the mobility of people between various economic classes before and after the MGNREGA. Finally, the study explores the relationship between economic indicators and social development indices (SDIs) with emphasis on the deprived class.

The paper is structured as follows. The next section briefly presents the basic features and achievements of the MGNREGA at the national level, as well as selected literature on rising global and Indian inequality. Section 'Purpose, data, and methodology' analyses not only the purpose, data, and methodology, but also the geographical coverage of the research. Section 'Deprivation' illustrates the intensity of deprivation at the sub-state level; that is, among districts within each state in terms of both the proportion of people in the three economic classes (bottom 25%, middle 50%, and top 25%) and their mean consumption expenditure. It also presents the economic mobility of districts between 2004/2005 and 2009/2010. Section 'Relationship between social infrastructure facilities and deprivation' explores the relationship between economic indicators from the 2004/2005 (61st Round) and 2009/2010 (66th Round) Household Consumer Expenditure Survey of the National Sample Survey Office (NSSO), and SDIs from the Indian census data for 2001 and 2011. The last section presents the concluding remarks.

Literature

The MGNREGA is a rural employment guarantee scheme implemented by the Government of India and enacted in 2005. The act was implemented in a phased manner, prioritising the most backward districts. In Phase I (in 2006), it was introduced in the 200 most backward districts of India. Then it was implemented in an additional 130 districts in Phase II (2007–2008), and in the remaining rural districts in Phase III (from 1 April 2008). Since then, the MGNREGA has covered all rural districts. The most important feature of the programme is that it gives a legal guarantee of 100 days of work to each rural household at the minimum rural wage rate prevailing in their respective states.

It is the largest public work program in the world, with annual average outlays of about Rs. 40,000 crores[3] (400 billion INR) and generates about 50 million person-days of work each year since the entire country was brought under the programme. The programme ensures that payments for work are made within 15 days and the distance between the home and workplace is no more than 5 km. The programme also gives priority to women and people from backward castes. It aims to reduce poverty by increasing employment opportunities and raising rural wages, especially for unskilled rural casual workers.

[3] Here and hereinafter, crore means 10^7.

Major features and achievements of the MGNREGA[4]

- Ensuring social protection for the most vulnerable people living in rural areas by providing employment opportunities to the tune of 100 days per household;
- Ensuring a secure livelihood for the poor through the creation of durable assets such as water supply, soil conservation, rural connectivity, strengthening drought-proofing, flood control, etc.;
- Aiding in the empowerment of marginalised communities, especially women, Scheduled Castes (SC), and Scheduled Tribes (ST) through the process of rights-based reservation;
- Strengthening decentralised, participatory planning through the convergence of various anti-poverty and livelihood initiatives;
- Deepening democracy at the grass root level by strengthening Panchayati Raj Institutions;
- Ensuring greater transparency and accountability in governance through social auditing.

As evident from Table 1, the total available funds for the implementation increased from Rs. 12,073.55 crores in 2006/2007 to Rs. 37,084.76 crores in 2013/2014, reaching a peak of Rs. 54,172.14 crores in 2010/2011, thereby making the total cumulative fund Rs. 303,469.62 crores from 2006/2007 to 2013/2014. If one adds the 2014/2015 and 2015/2016 budget and allotment, the total is more than Rs. 3.75 lakh[5] crores (INR 3,750 billion). Against this, the total expenditure increased from Rs. 8,823.35 crores in 2006/2007 to Rs. 24,848.75 crores in 2013/2014, while the peak expenditure occurred in 2012/2013 (Rs. 39,657.04 crores, which was 88% of the available budget). Given such significant figures, MGNREGA is, clearly, the largest rural development project in human history. Recent research (GOI 2012) shows that poverty in India declined, albeit slowly, over this period (2006/2007–2013/2014). However, the challenge is to sustain and accelerate this trend across all regions equally.

[4] See GOI (2005), (2012).
[5] Lakh means 10^5.

Table 1

MGNREGA performance – national overview

Denomination	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014* (up to December 2013)
Total number of job cards issued (crores)	3.78	6.48	10.01	11.25	11.98	12.50	12.79	12.72
Number of jobs provided to the households (crores)	2.10	3.39	4.51	5.26	5.49	5.06	4.98	3.81
Person-days per household (days)	43	42	48	54	47	43	46	35
Budget outlay (Rs. crores)	11,300	12,000	30,000	39,100	40,100	40,000	33,000	33,000
Total available fund (Rs. crores)	12,073.55	19,305.81	37,397.06	49,579.19	54,172.14	48,805.68	45,051.43	37,084.76
Expenditure (Rs. crores)	8,823.35	15,856.89	27,250.10	37,905.23	39,377.27	37,072.82	39,657.04	24,848.75
Percentage of expenditure in the available fund	73	82	73	76	73	76	88	67
Expenditure on wages (Rs. crores)	5,842.37	10,738.47	18,200.03	25,579.32	25,686.53	24,306.22	27,128.36	17,832.19
Percentage of expenditure on wages in the available fund	66	68	67	70	68	70	72	76
Total number of jobs taken up (lakhs)	8.35	17.88	27.75	46.17	50.99	80.77	106.51	111.64
Total number of jobs completed (lakhs)	3.87	8.22	12.14	22.59	25.90	27.96	25.60	11.17

* From 1st April 2013 to 31st December 2013.
Note: Crore – 10^7; lakh – 10^5.
Source: GOI (2012).

There is no doubt that the MGNREGA has attracted significant attention in India since the debate about poverty began in the 1970s and 1980s. However, rigorous empirical research with reliable data on the MGNREGA is still limited. One possible reason may be that it is essentially a direct interventionist policy to transfer funds to rural India under an act of the Parliament instead of a typical, textbook development model. Consequently, ideal textbook testing may not yield uniform results across such a diverse country. Yet, it proves that the Indian economy, society, and polity are imperfect, even after 70 years of democratic governance. Much of the concern about MGNREGA is related to corruption, and governance failures in the implementation of the programme have been widely discussed in Indian media. Here also, one of the main reasons for the scope of the diversion of funds is related to the decentralised process of funds transfer from the central government to the state governments, and then, sequentially, to the districts, blocks, and villages (Gram Panchayat). It is also true that the central government cannot be held responsible for any mis-governance in the long route it takes for funds to reach the vulnerable classes in extremely rural areas of such a vast and diverse nation. Therefore, unlike the National Expressway projects undertaken during the term of the first NDA (National Democratic Alliance) government and governed by the National Highway Authority of India, governance failure or corruption in the MGNREGA is an essential part of the much talked about 'democratic decentralisation' in India. In contrast to higher-level corruption in global projects, the achievements of and corruption in the MGNREGA at the lower-level administration within most of the states are as diverse as the country itself (Ghosh 2016a, 2016b). Of course, corruption is hardly unique to this scheme (Ravallion 2012, Shankar–Gaiha 2013, Dreze–Sen 2013, Ghosh–Roy 2015). In a recent paper, Das (2015) clearly shows that the politically active households and supporters of the local ruling party are more likely to receive benefits under MGNREGA in terms of participation, number of days of work, and earnings from the project using data from the state of West Bengal. The study also corroborates the quantitative findings with qualitative evidence. Similarly, Marcesse (2018) shows that the demand for work benefits does not emerge spontaneously from self-selecting rural citizens but is articulated by local elected officials under pressure to accommodate demands for rents from the bureaucracy. Specifically, local elected officials are compelled to disburse benefits selectively to ensure that they generate a surplus that will form the basis of bureaucratic and political rent payments. The study of Marcesse (2018) relies on qualitative data, specifically interviews with past beneficiaries of the scheme, bureaucrats tasked with policy implementation, and elected village leaders (the Gram Pradhans in the Indian state of Uttar Pradesh) to document the conditions under which a surplus is extracted and rent payments are made. The relative performance of the states in terms of corruption under the scheme is naturally of much interest. However, any attempt at generalising about such a far-reaching programme may result in losing the benefits of disaggregation in exchange for generality.

Despite such a massive transfer of funds for uniform rural development across all districts of India, rising regional disparity is the most important menace to the Indian government. Recent research shows that inequality in most countries is rising at faster rates in the new millennium. Basu (2006) alerted the world to rising inequality, particularly in China, India, and several other developing countries under the wave of globalisation. He suggests an internationally coordinated approach to solve the problem. Basu (2010) argues against contemporary capitalism and calls for collective actions to build a fair society. Ortiz and Cummins (2011) study economic inequality in 141 countries, and find that 20% of the rich enjoy 70% of the global income. Alkire et al. (2014) show that the bottom billion of the global poor live in 28 countries; India alone is home to 63.6% of the bottom billion.[6]

The problem in India is incredibly complex: the country exhibits a new dualism between rising inequality among a majority and rapid growth among a few of its citizens (Dreze–Sen 1995, 2002). India still cannot rid itself of the fundamental problems of social backwardness, agricultural stagnation, abject poverty, poor communication, poor health and sanitary conditions, vulnerable housing conditions, a poor education system, unemployment, and so on. Official survey statistics on poverty suggest that more than a quarter of the population lives below the calorie poverty line, which is significantly lower than the USD 1 norm.[7] Gupta and Ghosh (2007) estimate that another quarter is equally vulnerable, bringing the proportion close to the bottom 50%. There is enough evidence to show that rural India always remained poorer than urban areas, and the difference has persisted over time (Dubey–Gangopadhyay 1998, Deaton–Dreze 2002, Sundaram–Tendulkar 2003, Sen–Himansu 2004, Ghosh–De 2005, Chaudhury–Gupta 2007, Gupta–Ghosh 2007, Ghosh, 2010). Households that depend on agriculture as landless agricultural labourers, small and marginal farmers, and rural artisans constitute the bulk of the rural poor. In terms of social groups, SCs and STs account for a proportionately large number of the poor. They can easily slip into the category of the poor due to different 'shocks', including misfortunes such as disease and the death of the earning member of the household, loss of assets due to natural disasters, or a downward impact on current income due to market-induced factors (Vyas 2004).

Although recent growth in India's gross national income is above that of its neighbours, Dreze and Sen (2013) demonstrate that progress in key social indicators

[6] Kofi Annan (2016) recently posted on his foundation's site: 'In the economic and diplomatic sphere as well, we see this failure. Clamping down on tax evasion and using the money raised to invest in public services are two simple measures that would counteract the growing and damaging inequality in our world. This year will see the wealth of the top 1% exceed that of the remaining 99%.' p. 2.

[7] If one uses the USD 1 (approximately INR 67.00) per person per day poverty line, then it equals INR 2010 for 30 days in the current period. By this standard, no less than 60% Indians will lie below the poverty line.

is slower. Alkire and Seth (2015) note that India has been partly able to reduce monetary poverty, particularly in poorer states. However, this is not uniform across regions, castes, or religions, leading to rising inequality. Studies by Chaudhury and Gupta (2007) and Thomas (2007) provide adequate evidence to challenge the conventional state-level analysis that formed the basis for past policies to address poverty and deprivation in the post-2000 period. The first district-level study by Ghosh (2010) also shows intense variations in developmental performance across the districts within each state.

Spatial disparity is not unique in such a diverse and large country as India. Recent studies on Europe show that in a free-market global economy, spatial inequalities across central and eastern European countries such as Hungary exist, even though these countries are much smaller, and culturally and linguistically much less diverse compared with India (Nemes Nagy–Tagai 2011, Pénzes 2012, Azis 2013, Siposné Nándori 2014, Páger–Zsibók 2014, Egri–Tánczos 2015, Nagy–Nagy–Dudás 2016).

Against a backdrop of widespread disagreement on poverty and inequality relentlessly covered by the free press in India, the government could no longer rely on textbook development models for rural India and enacted the MGNREGA in 2005 to reinforce its commitment to livelihood security at the village level across all districts.[8] Proponents of the MGNREGA thought that this would significantly reduce the bottom quartile population in rural areas through a 'collective development spirit' (GOI 2012). Unfortunately, no one suspected the inner rift in the governance system of India.

GOI (2007) describes the situation as follows: 'As the Eleventh Plan commences, a widespread perception all over the country is that disparities among States, and regions within States, between urban and rural areas, and between various sections of the community, have been steadily increasing in the past few years and that the gains of the rapid growth witnessed in this period have not reached all parts of the country and all sections of the people in an equitable manner.' p. 137.

Purpose, data, and methodology

In this context, the present study investigates the performance of districts from the viewpoint of the bottom quartile in terms of inequality, purchasing power, and SDIs. An emphasis on the sub-state level is inevitable because the major thrust of the central government is rightly placed on the 'inclusion' of 'excluded people'. The present approach will help clarify how heterogeneous and alarming the deprivation scenario among the districts within each state of India is. The best way to cap-

[8] Note that there were more than 600 districts in India in 29 states in 2009/2010. However, in order to create a set of common districts for 2001, 2004/2005, 2009/2010, and 2011, we ignore some newly formed ones.

ture such a rift is to emphasise the lowest 'quartile'. This is a simple and useful way to capture the spatial location of deprived people at the district level.

More specifically, this study attempts to estimate both the state- and district-level proportions of the population in the following three economic classes in terms of monthly per-capita consumption expenditure (MPCE) at the household level from NSSO unit-level consumption data for the 61st (2004/2005) and 66th rounds (2009/2010) of the Household Income Expenditure Survey (HIES): 'destitute' class (bottom 25%), 'vulnerable' class (middle 50% people), and 'rich' class (top 25%). We also estimate the mean MPCE for each class and for the district as a whole.[9] We focus on 559 common rural districts[10] for which detailed information are available from the Census of India 2001 and 2011, and NSSO HIES 2004/2005 and 2009/2010. The main mission is to investigate whether sub-state level disparities reached alarming proportions relative to inter-state disparities measured by economic and social indicators. We estimate the SDI from a linear combination of work participation (work participation index [WPI]), human capital (human capital index [HCI]), health and housing (health and housing index [HHI]), and the transport and communication for rural districts (transport and communication index [TCI]). We follow the human development index method of the United Nations Development Programme (UNDP) (hereinafter UNDP method) using 44 common development indicators from the Census of India for 2001 and 2011.[11]

First, we measure the percentage of people at the district level corresponding to each class for each state separately by identifying the state's bottom quartile as the baseline to obtain the corresponding MPCE, because each Indian state has a separate official cost of living index. Then, we use the state's MPCE corresponding to the bottom 25% for each district in the state, so the district-level quartiles differ from that of the mother state as if the state mean is the weighted average of the districts. We repeat the same process for the middle 50% and top 25%. In addition, we derive the mean MPCE of the bottom and top quartiles and the middle 50% separately for each district. Finally, we also estimate each district's overall mean MPCE for the two points (2004/2005 and 2009/2010) without considering the classes. We summarise the arithmetic approach below.

[9] Counting the poor alone and leaving out a large proportion of people who live marginally above the conventional poverty line, but much below the stable level of income, from the core analysis would inevitably make any policy intervention ineffective. Many studies are revealing the rising social cost of land acquisition and new development initiatives by most state governments from all corners of the country (e.g. Nandigram and Singur in West Bengal, and Nandagudi in Karnataka) as the main outcome of the failure of 'trickle down' processes and the lack of dynamism in the rural sector, where the bulk of the people live.

[10] The number of the Indian districts was increased between 2001 and 2011 through bifurcation of the biggest ones based on social, economic, ethnic, and topographical factors. Thus, we included only those (common) districts in the research whose territory size remained unchanged.

[11] UNDP method of indexation: Index = (Own value – Minimum value)/ (Maximum value – Minimum value).

If B_i is the upper bound of MPCE for the i-th state corresponding to the bottom 25%, then the same B_i line is applied to each of the districts in the i-th state. This yields different proportions of people across districts in the i-th state. As each district is ruled by the same head of i-th state, these varying proportions directly represent the state's performance in terms of the bottom quartile. We apply the same methodology to estimate M_i (for the middle 50%) and T_i (for the top 25%) as well as the proportions of people corresponding to the middle 50% and top 25%, respectively. This approach does not require any monetary poverty line. Moreover, this approach makes it possible to precisely identify where the bulk of the deprived and better-off people live, with the corresponding MPCEs.

There is no official income data in India. One of the main reasons for this could be that about 94% of employment occurs in the informal, unorganised, and unrecorded sectors in India. The present approach is unique and the first of its kind in the sense that even such rich-friendly data can reveal the sub-state-level disparities without relying on any controversial definition of the poverty line.[12] Finally, we normalise the MPCE data by an appropriate deflator (consumer price index for agricultural labourers) between 2004/2005 and 2009/2010, so the values are directly comparable over time and space.[13] As the sample size is large, it approximately follows the standard normal distribution.[14]

Here, we aim to investigate the real differences for the three defined classes based on population proportions and MPCE using an inequality test (Z for the test statistic of MPCE and τ for the test statistic of population proportion) between comparable datasets generated from the NSSO. This dataset is not typically cross sectional; it is rather a comparative static analysis. Additionally, it uses only two types of indicators for this purpose, namely MPCE and population proportions, for the three classes as mentioned before. We must also mention here that the total population of India and per district is unusually high, so one could use decennial classes, as in the NSSO reports. However, that would make the analysis enormously complex and cumbersome and require a huge amount of space to address. The bottom 25% is the target population of most public policies of the Indian government, because the national poverty level hovered above the first quartile for a long time. More importantly, since there are significant differences in MPCE and

[12] The idea of the poverty line in India is simply an insult to the dignity of the poor who also have one vote just like the rich or the celebrities. The official poverty line in India dates back to the late 1960s. The conventional poverty line does not include health, housing, education, and accident, which are essential elements of a multidimensional poverty study.

[13] Although this dataset is highly reliable, it is widely accepted that the poor do not have much to hide, while richer people's voluntary disclosures are highly suspicious. As a result, the Gini coefficient estimated from such one-sided under-reported data cannot capture the actual degree of disparity in India.

[14] Type 1 error is the error committed in rejecting a null hypothesis by the test when it is actually true. The critical region is determined such that the probability of Type 1 error does not exceed the level of the significance of the test.

the corresponding population proportions across the four quartile groups, the mean differences across the nation, state, and district as a whole do not make any sense in India. Therefore, the average represents neither India overall nor the states. Even district level differences exist. Unfortunately, there is no such data below the district level.

Economic deprivation analysis of MPCE and the proportions of various classes of people

Table 2 presents the coefficients of variation (CVs) of the proportions of people in the bottom, middle, and top economic classes within the districts in each state for 2004/2005 and 2009/2010. The figures reveal numerous alarming features about the growth of MPCE and rising deprivation within most states. First, the CV of the proportions of people in the bottom quartile varies enormously across states, from 54.65 in Himachal Pradesh to 137.55 in Orissa in 2004/2005. These extreme values increased to 73.19 in Himachal and 156.03 in Orissa in 2009/2010. Second, such high and rising values of the CVs of the bottom 25% people among districts within states simply suggest that even the state-level average is meaningless in understanding the vertical rift among people at the sub-state level during the highest growth phase. It also proves that a whole host of poverty removal policies could not have any perceptible impact on the lives of the most deprived people in rural India. Third, the last two columns also show that the bottom quartile is similar to the top quartile; that is, the values of the CVs are too significant to ignore. On the other hand, there is no significant change in the composition of the middle 50%. Undeniably, a large majority of districts are home to extremely poor persons and the remaining districts accommodate those growing richer in each state. In a sense, rising deprivation and limited affluence have been the order in India at the sub-state level during the last decade, even after the complete implementation of the MGNREGA. This means that direct interventionist development projects like the MGNREGA failed to reduce rising disparity at the sub-state level.

Table 3 presents the inter-state disparities measured by the CV of mean MPCE for the three classes in each state. We calculate the CVs from the district level MPCEs for each state. In sharp contrast to CVs of the proportions of people in the three classes, those for mean MPCE present a completely different picture. First, there are no perceptible statistical variations in mean consumption in any of the classes from 2004/2005 to 2009/2010. In fact, except for the rich class, the MPCE of Uttaranchal in 2009/2010, all CVs are negligible. This suggests that the rising vertical rift among people (see Table 2) accompanied stagnant average expenditures among the classes in each state. Second, disparities in mean MPCEs were rising at faster rates for the top 25% between 2004/2005 and 2009/2010 in Andhra Pradesh, Assam, Gujarat, Kerala, Madhya Pradesh, Jammu and Kashmir, and Uttaranchal.

Table 2

Coefficients of variation of the proportion of people in the bottom, middle and top economic classes among the districts in each state, 2004/2005, 2009/2010

State	Economic class					
	Lowest (25%)		Middle (50%)		Top (25%)	
	CV of 61R	CV of 66R	CV of 61R	CV of 66R	CV of 61R	CV of 66R
Andhra Pradesh	98.24	113.75	24.93	44.15	65.58	101.12
Assam	101.91	114.21	32.65	46.73	75.33	156.79
Bihar	94.08	88.27	30.60	35.80	95.52	115.94
Chhattisgarh	137.15	75.51	43.76	44.00	92.99	103.01
Gujarat	131.75	119.31	41.93	53.56	123.91	135.39
Haryana	84.76	90.90	35.83	38.06	67.64	85.51
Himachal Pradesh	54.65	73.19	10.55	22.85	53.55	93.20
Jammu and Kashmir	113.14	117.15	27.06	26.74	113.11	64.34
Jharkhand	91.74	102.87	31.77	32.70	93.15	88.49
Karnataka	118.95	148.65	42.83	47.94	104.08	157.12
Kerala	92.56	84.49	21.60	24.60	87.01	88.45
Maharashtra	115.54	120.28	34.39	38.30	102.96	76.80
Madhya Pradesh	122.80	164.20	46.18	65.64	117.22	157.79
North East	94.23	93.82	35.95	37.35	88.37	96.10
Orissa	137.55	156.03	36.62	45.63	97.84	105.34
Punjab	84.84	68.97	20.04	18.36	72.12	66.51
Rajasthan	88.02	113.92	37.56	42.53	66.91	75.10
Sikkim	80.48	94.00	18.54	18.11	101.96	60.87
Tamil Nadu	76.57	118.05	26.00	47.56	63.94	89.01
Uttar Pradesh	90.51	86.44	30.24	38.18	87.23	78.67
Uttaranchal	70.82	109.86	32.71	55.23	50.67	87.50
West Bengal	94.56	64.23	28.15	35.36	60.35	58.58

Note: Here and in the following tables and figures, CV – coefficient of variation; and 61R and 66R refer to the 61st (2004/2005) and 66th (2009/2010) rounds of the Household Income Expenditure Survey. CV measures inter-district variation across the three economic classes within each state.

Table 3

Coefficients of variation among the districts in each state for the mean MPCE of the bottom, middle, and top classes, 2004/2005, 2009/2010

State	Economic class					
	Lowest (25%)		Middle (50%)		Top (25%)	
	CV of 61R	CV of 66R	CV of 61R	CV of 66R	CV of 61R	CV of 66R
Andhra Pradesh	5.04	8.80	4.15	5.49	8.15	21.08
Assam	4.75	4.47	4.29	5.86	6.68	30.32
Bihar	3.62	8.21	3.08	4.26	8.28	9.05
Chhattisgarh	10.41	7.27	3.68	3.82	20.77	8.54
Gujarat	6.93	8.13	5.08	5.72	13.01	22.73
Haryana	6.52	8.13	3.86	4.57	37.50	17.92
Himachal Pradesh	3.57	5.22	3.32	4.85	10.55	11.15
Jammu and Kashmir	3.85	6.63	3.79	3.12	6.87	18.31
Jharkhand	5.06	8.00	3.13	4.01	11.21	10.84
Karnataka	5.06	11.65	3.27	5.97	21.72	14.60
Kerala	6.96	6.82	3.63	3.29	9.65	31.54
Maharashtra	5.95	7.35	4.10	4.41	14.66	14.65
Madhya Pradesh	6.64	9.35	4.91	5.38	14.44	25.54
North East	24.23	13.24	21.53	12.06	27.54	23.57
Orissa	7.23	12.44	5.08	4.87	10.12	12.19
Punjab	4.84	4.88	3.89	4.22	10.73	16.78
Rajasthan	5.16	6.00	3.09	3.10	14.67	12.60
Sikkim	0.27	2.15	5.07	2.56	7.878	4.31
Tamil Nadu	6.12	7.17	3.10	3.41	31.76	21.51
Uttaranchal	3.78	7.77	3.78	3.74	21.28	41.05
Uttar Pradesh	5.74	6.10	3.41	4.29	20.63	26.33
West Bengal	4.64	4.94	2.46	2.93	13.68	7.02

Note: Here in the following tables and figures, MPCE – monthly per-capita consumption expenditure.

Table 4

Trend in the proportion of the three population classes by state, 2004/2005–2009/2010

State	Value of τ for the lowest 25%	Value of τ for the middle 50%	Value of τ for the top 25%	Lowest 25% in 66R compared with 61R	Middle 50% in 66R compared with 61R	Status of the top 25% in 66R compared with 61R
Andhra Pradesh	–1.56	1.25	17.06	Did not decrease	Did not increase	Did not increase
Assam	–152.76	175.32	–28.48	Did not decrease	Did not increase	Increased
Bihar	–36.43	–29.23	70.75	Did not decrease	Increased	Did not increase
Chhattisgarh	–15.92	–5.90	27.48	Did not decrease	Increased	Did not increase
Gujarat	2.10	6.51	13.81	Decreased	Did not increase	Did not increase
Haryana	55.62	–65.32	30.22	Decreased	Increased	Did not increase
Himachal Pradesh	–0.59	0.88	–10.17	Did not decrease	Did not increase	Increased
Jammu and Kashmir	93.62	–16.41	–59.05	Decreased	Increased	Increased
Jharkhand	74.58	17.05	–98.50	Decreased	Did not increase	Increased
Karnataka	1.66	6.87	–8.19	Decreased	Did not increase	Increased
Kerala	0.73	–13.03	2.97	Did not decrease	Increased	Did not increase
Maharashtra	6.21	–8.93	6.33	Decreased	Increased	Did not increase
Madhya Pradesh	12.22	–7.46	2.38	Decreased	Increased	Did not increase
North East	–4.56	3.47	–18.30	Did not decrease	Did not increase	Increased
Orissa	–15.31	–1.20	18.33	Did not decrease	Did not increase	Did not increase
Punjab	17.78	–23.27	21.84	Decreased	Increased	Did not increase
Rajasthan	–4.92	–5.64	3.80	Did not decrease	Increased	Did not increase
Sikkim	–6.33	1.30	–0.67	Did not decrease	Did not increase	Did not increase
Tamil Nadu	218.42	–101.34	–78.82	Decreased	Increased	Increased
Uttar Pradesh	–1.97	–9.03	–16.37	Did not decrease	Increased	Increased
Uttaranchal	22.64	19.90	–29.64	Decreased	Did not increase	Increased
West Bengal	–3.03	–60.62	78.92	Did not decrease	Increased	Did not increase

So far, we see that deprived districts remained poor and the richer districts strengthened their affluence from 2004/2005 to 2009/2010. In this situation, it is not easy to determine a state's performance in terms of both the proportion of people in the bottom quartile and their mean consumption. In order to reach a statistically tested conclusion, we apply the τ test. Table 4 presents the changes in the proportions of people in the bottom quartile by state from the 61st to the 66th Round of the HIES. According to the figures, Gujarat, Haryana, Jammu and Kashmir, Jharkhand, Karnataka, Maharashtra, Madhya Pradesh, Punjab, Tamil Nadu, and Uttaranchal successfully reduced their lowest 25% people from 2004/2005 to 2009/2010. In a democratic governance system with all forms of social backwardness, this achievement in 10 states cannot be underestimated in the high growth period since 2004/2005. On the other hand, the proportions of the lowest 25% did not decline statistically in Andhra Pradesh, Assam, Bihar, Chhattisgarh, Himachal Pradesh, Kerala, North East, Orissa, Rajasthan, Sikkim, Uttar Pradesh, and West Bengal. Note that this second group of states, where state governments failed to make any positive progress in the proportions of deprived people, is largely composed of the poorer states of India except Kerala and Himachal. Therefore, the advantage of the high growth period was harvested by the relatively people in backward states. Finally, the results for the middle 50% and top 25% are highly mixed, yet the more prosperous states like Gujarat, Haryana, Kerala, Maharashtra, Punjab, and Tamil Nadu were clearly successful in dispersing the 'fruits' of development among a larger population rather than enriching the top 25% group only.

Table 5 presents the trend in the average MPCE of the districts from 2004/2005 to 2009/2010. MPCE is doubtless a better indicator of the economic well-being of the general population than per capita income estimates because our study focuses more on the poorer sections. MPCE aggregates the monetary value of all goods and services actually consumed during a particular reference period. To analyse the change in MPCE over this period we use the equality test of MPCE for the bottom 25%, middle 50%, and top 25%.

Our null (H_0) and alternative hypotheses (H_1) are as follows:

$$H_0: p_1 = p_2;$$

$$H_1: p_1 > p_2,$$

where p_1 and p_2 are two population proportions of 61R and 66R, respectively. If H_0 is true, then population proportions of 61R and 66R are equal.

Table 5

Trend in the average MPCE by state, 2004/2005–2009/2010

State	Value of *Z* for the lowest 25% of the population	Value of *Z* for the middle 50% of the population	Value of *Z* for the top 25% of the population	Mean MPCE of the lowest 25% of the population in 66R compared with 61R	Mean MPCE of the middle 50% of the population in 66R compared with 61R	Mean MPCE of the top 25% of the population in 66R compared with 61R
Andhra Pradesh	–3,243,274.51	–10,601,618.92	–33,393.52	Improved	Improved	Improved
Assam	–4,326.87	793,989.29	–833,646.15	Improved	Did not improve	Improved
Bihar	396,382.78	–4,962,444.58	–1,916,679.80	Did not improve	Improved	Improved
Chhattisgarh	–254,890.93	–1,821,709.74	7,394.57	Improved	Improved	Did not improve
Gujarat	–1,661,390.43	–3,078,946.88	–232,826.02	Improved	Improved	Improved
Haryana	57,651.68	–2,146,151.49	188,842.30	Did not improve	Improved	Did not improve
Himachal Pradesh	–1,008,623.53	–1,656,732.38	–424,370.99	Improved	Improved	Improved
Jharkhand	–478,158.91	–4,724,281.58	–899,796.80	Improved	Improved	Improved
Jammu and Kashmir	–131,983.96	–871,142.99	–59,922.268	Improved	Improved	Improved
Karnataka	–84,986.00	–2,640,356.37	302,603.30	Improved	Improved	Did not improve
Kerala	–2,969,770.12	–8,925,438.93	–1,354,128.50	Improved	Improved	Improved
Maharashtra	–3,995,306.72	–10,029,492.34	–419,771.43	Improved	Improved	Improved
Madhya Pradesh	–703,660.13	–3,731,269.42	–1,056,726.80	Improved	Improved	Improved
North East	–63,267.37	–142,442.51	–81,502.37	Improved	Improved	Improved
Orissa	–1,188,903.70	–5,208,672.80	–338,459.31	Improved	Improved	Improved
Punjab	–397,353.55	–2,220,854.35	–234,909.59	Improved	Improved	Improved
Rajasthan	–1,317,783.52	–4,258,749.68	113,973.53	Improved	Improved	Did not improve
Sikkim	–19,695.24	–34,977.21	–16,741.10	Improved	Improved	Improved
Tamil Nadu	–2,233,198.10	–9,869,311.32	30,347.06	Improved	Improved	Did not improve
Uttaranchal	–680,810.04	–1,335,937.17	–441,819.11	Improved	Improved	Improved
Uttar Pradesh	–1,712,625.18	–2,782,689.23	414,083.27	Improved	Improved	Did not improve
West Bengal	–2,259,549.07	–6,076,443.96	1,327,926.70	Improved	Improved	Did not improve

Note: Inequality test of the proportions: 1. for the bottom 25% of the population: If $\tau > \tau_{(\infty = 0.05)} = 1.645$, then we reject H_0, otherwise we accept H_0; 2. for the middle 50% of the population: If $\tau > -\tau_{(\infty = 0.05)} = -1.645$, then we reject H_0, otherwise we accept H_0; 3. for the top 25% of the population: If $\tau > -\tau_{(\infty = 0.05)} = -1.645$, then we reject H_0, otherwise we accept H_0.

According to Table 5, the mean MPCE of the lowest 25% people improved between 2004/2005 and 2009/2010 compared to 2004/2005 for all states except Bihar and Haryana. In the poorest state of India, Bihar, the middle and top classes improved their MPCE over the period. This means that the poor are getting poorer and the rich and the middle are getting richer in Bihar. If we observe the status of Haryana, there is no improvement in the poor or rich class, but a statistically significant improvement for the middle class. Thus, the 50% of people living at the two extreme ends remained unchanged by the growth momentum in the new millennium in Haryana. The little MPCE improvement for the lowest 25% in Andhra Pradesh, Assam, Chhattisgarh, Himachal Pradesh, Kerala, North East, Orissa, Rajasthan, Sikkim, Uttar Pradesh, and West Bengal is nullified by the deterioration in the proportion of the population of the lowest 25% for these states. This suggests that the gap between rich and poor is not decreasing in rural districts at large. Thus, economic inequality measured from real consumption expenditure did not decline in these states.

Having performed the statistical tests across the states in terms of the people in the bottom quartile and their mean consumption, we attempt to determine how many districts improved their economic positions measured by the reduction in the bottom quartile population. Table 6 presents the mobility by district in terms of bottom 25% people from 2004/2005 to 2009/2010. The major observations are as follows. First, out of 22 states, only 11 reduced successfully (in at least one additional district) the proportion of its bottom 25% population in 2009/2010. Out of these 11 states, two states, Maharashtra and Orissa, have moved four additional districts out of dire deprivation, and three districts of Jharkhand and Madhya Pradesh have moved out of extreme deprivation. Second, in the examined period, the number of districts having more than 25% of its people in the bottom quartile increased in nine states: Bihar, Chhattisgarh, Himachal Pradesh, Karnataka, Kerala, Punjab, Sikkim, Uttaranchal, and Uttar Pradesh. The worst performing state in this regard is Uttar Pradesh, where 46 districts had less than 25% of their population in the bottom quartile in 2004/2005, which fell to 37 districts in 2009/2010. That is, nine additional districts became deprived in 2009/2010. These bad performing states include both wealthier states like Himachal Pradesh, Karnataka, Kerala, Punjab, Sikkim, and Uttaranchal, and impoverished ones like Bihar, Chhattisgarh, and Uttar Pradesh.

Overall, duality between affluence and deprivation is primarily sweeping across the poorer states, where most people are unskilled, lack any dependable asset base, and could thus not 'reap' the benefits of globalisation.

Table 6

Mobility of districts in terms of the bottom 25% of the population, by state and the total number of districts, 2004/2005–2009/2010

State/Total number of districts	61R	66R		61R	66R	
	Below 25% of the population	Below 25% of the population		Above 25% of the population	Above 25% of the population	
	G	G2G	B2G	B	G2B	B2B
Andhra Pradesh	11	7	5	11	4	6
Total number of districts	11	12		11	10	
Assam	10	7	5	12	3	7
Total number of districts	10	12		12	10	
Bihar	20	8	9	17	12	8
Total number of districts	20	17		17	20	
Chhattisgarh	10	6	2	6	4	4
Total number of districts	10	8		6	8	
Gujarat	13	5	9	12	8	3
Total number of districts	13	14		12	11	
Haryana	10	5	5	9	5	4
Total number of districts	10	10		9	9	
Himachal Pradesh	9	6	2	3	3	1
Total number of districts	9	8		3	4	
Jharkhand	7	4	6	11	3	5
Total number of districts	7	10		11	8	
Jammu and Kashmir	5	3	3	4	2	1
Total number of districts	5	6		4	3	
Karnataka	18	11	4	9	7	5
Total number of districts	18	15		9	12	
Kerala	9	5	3	5	4	2
Total number of districts	9	8		5	6	
Madhya Pradesh	25	15	13	20	10	7
Total number of districts	25	28		20	17	
Maharashtra	15	10	9	18	5	9

(Continued on the next page.)

(Continued)

State/Total number of districts	61R Below 25% of the population	66R Below 25% of the population		61R Above 25% of the population	66R Above 25% of the population	
	G	G2G	B2G	B	G2B	B2B
Total number of districts	15	19		18	14	
North East	24	10	15	24	14	9
Total number of districts	24	25		24	23	
Orissa	13	4	13	17	9	4
Total number of districts	13	17		17	13	
Punjab	10	6	2	7	4	5
Total number of districts	10	8		7	9	
Rajasthan	18	12	6	14	6	8
Total number of districts	18	18		14	14	
Sikkim	2	1	0	2	1	2
Total number of districts	2	1		2	3	
Tamil Nadu	20	16	6	9	4	3
Total number of districts	20	22		9	7	
Uttaranchal	9	6	1	4	3	3
Total number of districts	9	7		4	6	
Uttar Pradesh	46	20	17	24	26	7
Total number of districts	46	37		24	33	
West Bengal	9	5	5	8	4	3
Total number of districts	9	10		8	7	

Note: If a district has more than 25% of its population in the bottom 25% class, it is called 'bad' (B), and if it has less than 25% of its population in the bottom 25% class, it is called 'good' (G). G2G – the district was in the good category in both rounds of the Household Income Expenditure Survey (HIES); B2B – the district was in the bad category in both rounds of the HIES; B2G – the district moved from the bad category to the good category in the examined period; G2B – the district moved from the good category to the bad category in the examined period.

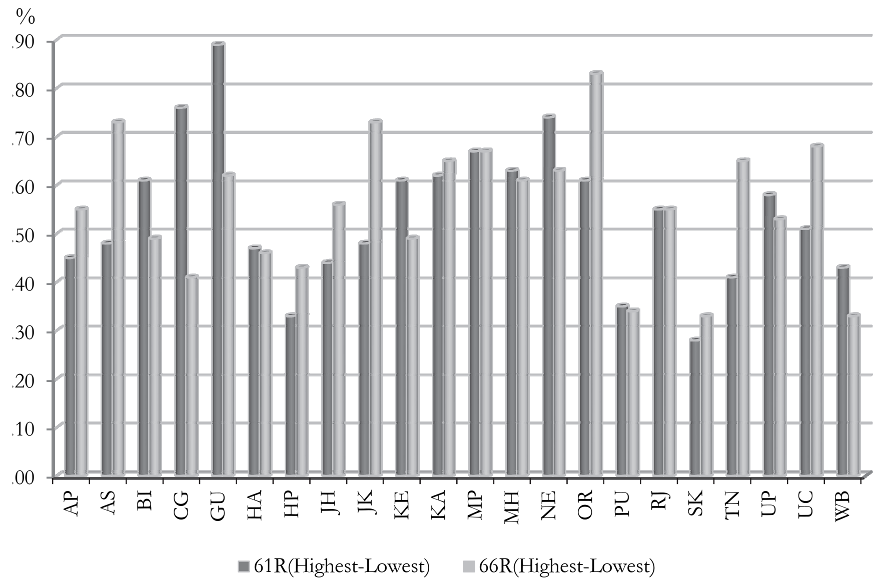
Inequality test of the MPCE: 1. for the bottom 25% of the population: We reject H_0 if $z_1 > z_{(\infty = 0.05)} = 1.645$, otherwise we accept H_0; 2. for the middle 50% of the population: We reject H_0 if $z_1 < z_{(\infty=0.05)} = -1.645$, otherwise we accept H_0; 3. for the top 25% of the population: We reject H_0 if $z_1 < z_{(\infty=0.05)} = -1.645$, otherwise we accept H_0.

Figure 1 captures the percentage gaps between the highest and lowest bottom quartiles of population across the districts of the Indian states for 2004/2005 and 2009/2010. The gaps remained significantly high in almost all states, ranging from 33% in Sikkim to 83% in Orissa in 2009/2010. In addition, it increased in Andhra

Pradesh, Assam, Himachal Pradesh, Jammu and Kashmir, Kerala, Orissa, Sikkim, Tamil Nadu, and Uttaranchal, while there is a significant fall in Chhattisgarh and Gujarat, and a moderate fall in Haryana, Punjab, and West Bengal.

Figure 1

Percentage gaps between the highest and lowest bottom quartiles of population across districts by state, 2004/2005, 2009/2010



Note: AP – Andhra Pradesh; AS – Assam; BI – Bihar; CG – Chhattisgarh; GU – Gujarat; HA – Haryana; HP – Himachal Pradesh; JK – Jammu and Kashmir; JH – Jharkhand; KA – Karnataka; KE – Kerala; MH – Maharashtra; MP – Madhya Pradesh; NE – North East; OR – Orissa; PU – Punjab; RJ – Rajasthan; SK – Sikkim; TN – Tamil Nadu; UP – Uttar Pradesh; UC – Uttaranchal; WB – West Bengal.

The analysis above shows that there is no improvement in rural deprivation for the bottom quartile of people across Indian districts. In fact, in 2004/2005 there were 246 districts in which more than 25% of the population was in the bottom quartile, and by 2009/2010 this number increased to 247. Furthermore, in 2004/2005, the highest percentage of the population in the bottom quartile was 89% (in Dang District in the state of Gujarat, where the total population was as low as 0.25 million – the lowest among the Indian districts), whereas in 2009/2010 it was 86% (in Koraput District in the state of Orissa). There is, therefore, hardly any doubt that even the nationwide implementation of the MGNREGA, which targeted the bottom 25% of unskilled rural people, could not change relative deprivation at the sub-state level beyond the state capitals.

Figure 2

Rural district map of India representing the percentage of population in the lowest economic class, 2004/2005

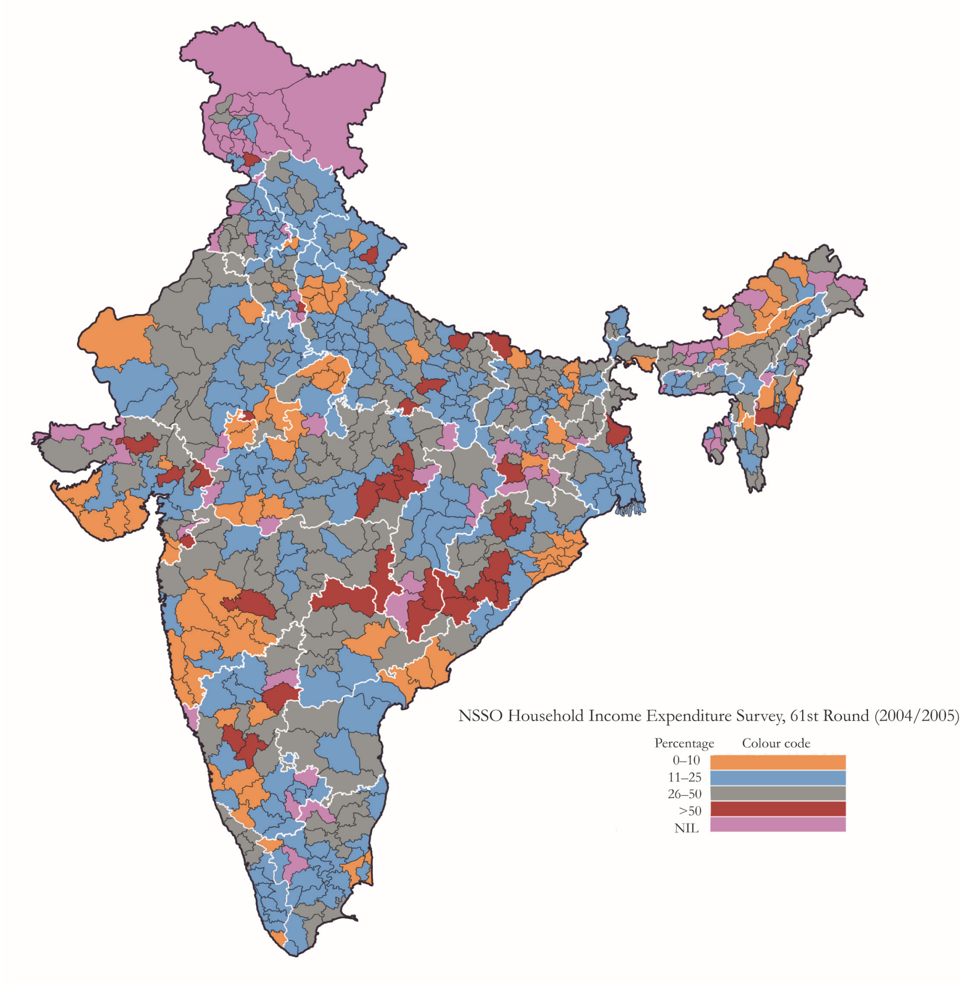
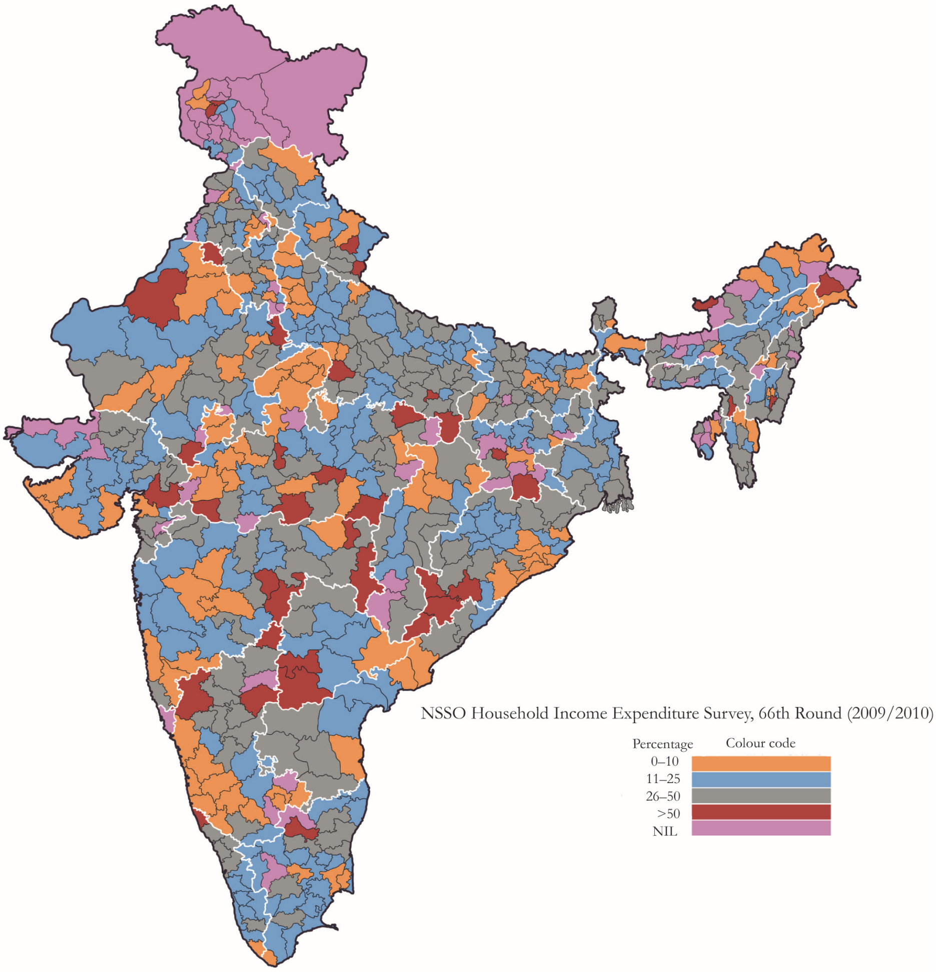


Figure 3

Rural district map of India representing the percentage of population in the lowest economic class (2009/2010)



Figures 2 and 3 show the maps of the districts of India for 2004/2005 and 2009/2010, respectively. The districts are distinguished by four different colours: districts with 0–10% deprived people in the total population appear in orange, those with 11–25% appear in blue, those with 26–50% appear in grey, and those with 'more than 50%' appear in red. The most interesting observation is that the number of districts in the lowest (0–10% deprived people in the population) and highest groups (more than 50% deprived people in the population) increased between 2004/2005 and 2009/2010. That is, there were 87 districts in 2004/2005 with 0–10% population in the bottom quartile, which increased to 114 by 2009/2010. This is certainly a major achievement. On the contrary, the number of districts with more than 50% of the population in the bottom quartile increased from 35 to 39. This suggests that even with rich-friendly consumption data, there is a tendency towards higher inequality between 2004/2005 and 2009/2010. Moreover, the number of districts in the middle fell from 437 to 406. Therefore, the maps reveal, overall, that

- the poorest districts (coloured to red) with more than 50% of their population in the bottom quartile have remained poor; these districts are mainly located in the central and Eastern regions along the imaginary vertical line from Kashmir to Rameswaram (Tamil Nadu);
- the number of the 'best' districts (orange) with 0–10% of their population in the bottom quartile has increased; these districts are located west of the formerly mentioned imaginary line;
- in the 'middle' districts (grey), 26–50% of the population is in the bottom quartile; these districts are evenly distributed across India, with a preponderance in Eastern India in 2009/2010; and
- Eastern India, excluding the northeast districts, has lagged behind the rest part of the country in terms of deprivation.

Relationship between social infrastructure facilities and deprivation

The immediate question is why deprivation persists at sub-state levels (i.e. in districts) for such a long period in the largest democracy, where the bulk of the development fund is allocated for removing rural backwardness, poverty, and deprivation.

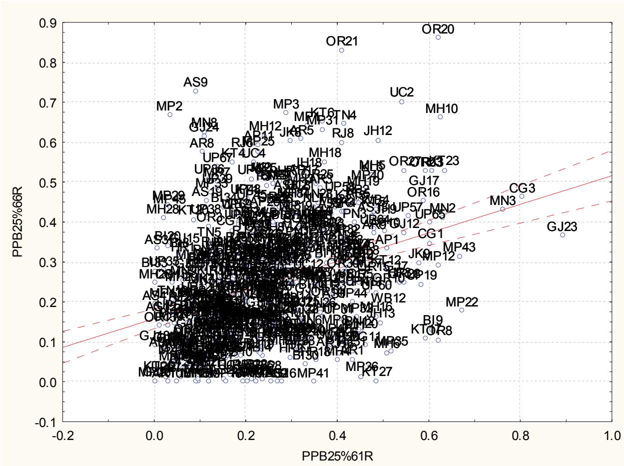
This section explores the relationship between economic indicators and various SDIs in districts as the units of analysis.

Factors determining districts level development

As it was already mentioned, we estimate four types of indices following the UNDP method. The overall SDI (SDIR2001 for 2001 and SDIR11 for 2011) consists of the WPI, the HHI, the HCI, and the TCI for the same set of districts (559) in comparable away. We present these estimates in Figures 4–10 at the district level. They prove that our hypothesis on the immobility of the districts during the highest growth phase is empirically true. The figures show each observation by district with the first two digits of the state names followed by a numerical serial number corresponding to the number of districts from the same state. For example, OR21 is the 21st district of state Orissa in alphabetical order; AS9 is the 9th district of Assam, and so on.

Figure 4

Scatterplot of the proportions of people in the bottom quartile in the 61st and 66th rounds of the Household Income Expenditure Survey (r = 0.34)



Note: PPB25%66R – proportion of people in the bottom quartile in the 66th Round; PPB25%61R – proportion of people in the bottom quartile in the 61st Round.

The seven figures capture the following issues. Figure 4, linking the proportions of people in the bottom quartile in the 61st Round of the HIES (PPB25%61R) and those in the 66th Round (PPB25%66R), suggests that there is no perceptible relationship between them (r = 0.34). Figure 5 illustrates the MPCE of the bottom quartile between the 61st (MCB25%61R) and 66th rounds (MCB25%66R). It is obvious that the poorest districts remained poor, while the richest districts remained rich between 2004/2005 and 2009/2010. The correlation coefficient is 0.85. Figure 6 presents the relationship between MPCE for all districts (MDtMPCE61R) in the 61st Round and that (MDtMPCE66R) in the 66th Round. The results show that there is no change in the relative positions of districts, even after the complete implementation of the MGNREGA in 2009/2010 (r = 0.72). Figure 7 reports on the relationship between the SDI in 2001 (SDI2001) and that in 2011 (SDI11). According to the figures, there is not any change in the relative positions of districts in terms of the overall SDIs between 2001 and 2011. The correlation is as high as 0.81. Figures 8 and 9 capture the relation between the overall SDI and the mean MPCE of districts for the 61st and 66th rounds, respectively. It is interesting to note in the figures the overall shift in the district-level MPCE during the highest growth phase, which is perceptible from the scale in the y-axes. Yet, the impact of the SDI became weak after 2009/2010. Finally, Figure 10 illustrates the result of a simple 'test of convergence' in SDI between 2001 and 2011, with SDI2001 on the x-axis and the rate of growth of the SDI on the y-axis. The result is not very depressing: there was a 'mild' tendency towards district-level convergence in the overall SDI during the last decade. That is, districts with lower SDI values in 2001 recorded relatively higher growth rates in SDI in 2011 with a rush in middle. However, the trend towards convergence is so weak that there are no significant changes in the relative positions of the districts, as Figure 7 shows.[15] Overall, it is evident from Figures 4–10 that a new form of dualism (affluence and deprivation) emerged between the rulers and the ruled in rural areas, due to widespread governance failures across most states. The question is: What went wrong with the billions of INR spent to eradicate poverty and inequality in rural areas through the MGNREGA during the last 10 years?

This prompts us to investigate the classic relationship preached by the pioneers of development economics between various SDIs ('infrastructure') and the economic well-being of people at the lowest level of the economic ladder under a 'socialistic pattern of society'. We test this relationship using a linear equation. We present two sets of results: one for the 2001 Census indices as independent variables, with the 2004/2005 MPCE as the dependent variable, and a similar set for the 2011 Census indices and the 2009/2010 MPCE. The supply of various infrastructure facilities is a static stock available at certain time that helps people get involved in the development process through meaningful engagement while generating economic activities, which reflect in consumer expenditure or purchasing power at the district level. First, we use the overall district-level MPCE as the dependent variable for each round of the HIES. Then, we test the equation with the proportion of the population in the bottom 25% economic class and the MPCE of the bottom quartile only for 2009/2010 as the dependent variable after eliminating 30 outliers according to Cook's distance statistics. We report the results in Tables 7–11.

Thus,

$$y = a + bx_1 + cx_2 + dx_3 + ex_4 + e \,, \qquad (1)$$

where y denotes district-level average of the MPCE, a is a constant term, x_1 is the rural WPI (WPIR), x_2 is the rural HCI (HCIR), x_3 is the rural HHI (HHIR), x_4 is the rural TCI (TCIR), and e refers to the error term.

Ideally, the work participation, human capital, health and housing, and transport and communication indices should yield positive elasticity with respect to household expenditure. As Table 7 reveals, the adjusted $\overline{R}^2$ is 0.40 for the combination between 2001 and 2004/2005, while it is 0.31 for 2011 and 2009/2010 (see Table 8). This means that the factors selected to explain purchasing power across extremely heterogeneous rural districts yield a moderately good result. One interesting similarity between these two sets of results is that the WPI has a negatively significant coefficient in Table 8. It points to a peculiar work culture in India: on the one hand, the

[15] The test of convergence in SDIs between 2001 and 2011 among the common districts of India is as follows. Regression summary for the dependent variable (SDIR2001): Adjusted $\overline{R}^2$ = 0.12, $F(1,557)$ = 79.457, p < 0.00; Intercept: B = 0.39, Standard error of B = 0.00, $t(557)$ = 102.83, p-level = 0; AAGRSDI2001 and 2011: β = –0,35, Standard error of β = 0.040, B = –0.02, Standard error of B = 0.00, $t(557)$ = –8.91, p-level = 0. (See Barro–Sala-i-Martin 1995, Ghosh–Marjit–Neogi 1998).

backward districts with primarily hilly and forest terrain dominated by poor illiterate ST and SC people have higher work participation rates with lower purchasing power, while, on the other hand, work participation is relatively low in the relatively richer districts. Perhaps, for the first time, this unique feature of India is proven statistically. We are not sure whether the classic dichotomy between earning and leisure can be experienced in the richer rural districts, or the semi-feudal nature of Indian society, in which the rich should not work hard, while the poor must. Another peculiarity is that the HCI appeared to have no contribution to purchasing power in 2004/2005, whereas it yielded moderate statistical significance in 2009/2010. In both situations, the HHI and the TCI show a positive, significant contribution to the district-level purchasing power; and in each case, the *F*-statistic along with *df* and the *p*-value assure that the overall estimates are highly consistent with the required statistical properties.

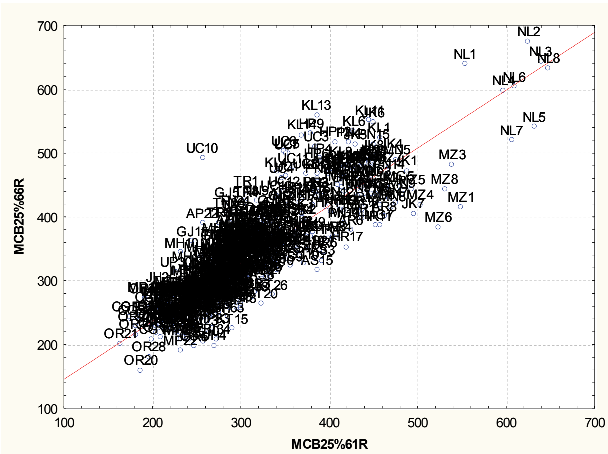
As the explained variation appears low in Table 8, we attempt to detect the extremely diverse districts as outliers using Cook's distance statistics. It has indeed yielded better results. The 29 outlier districts belong to only 13 states.[16] Tables 9, 10, and 11 provide three sets of results. In Table 9, where the overall district MPCE is the dependent variable, the adjusted $\bar{R}^2$ increased to 0.45. In Table 10, where the proportion of people in the bottom quartile is the regressand, it returns the oddest values of all the statistics, including the adjusted $\bar{R}^2$ (0.05). The results in Table 10 have tremendous implications for future policy issues. None of the four types of infrastructure facilities had any expected relation with the most deprived people across rural districts. In other words, even with the professed goals of removing poverty, a complete set of misplaced policy targets was pursued over the last five decades culminating in 2009/2010. Thus, work participation, health and housing, human capital, and transport and communication did not play their desired roles in helping the poor out of deprivation. It is indeed very unfortunate. Finally, Table 11 captures the results with the mean MPCE of the bottom quartile as the regressand. It has high significance in a democratic polity. The government's aim is to reduce the proportion of people in dire poverty (bottom 25%) by increasing their mean consumption. As a natural outcome of intense diversity across the districts, both the signs and values of the regressors differ for each of the three regressands. In Table 11, WPIR11 and TCIR11 have negative coefficients, thereby suggesting that 1. districts with a higher WPI dominated by poor ST and SC people have lower mean consumption in the bottom quartile and 2. the TCI does not have much significance in influencing the mean MPCE of the bottom quartile in the relatively backward districts. We may conclude that in order to increase district-level consumption expenditure, the government should place emphasise on human capital, health and

[16] The number of districts (in parentheses) from the respective states, deleted by Cook's distance statistics for the estimated results given in Tables 9–11 are: Kerala (5), Arunachal (4), Karnataka (4), Punjab (3) and Nagaland (3), Jammu and Kashmir (2), Uttaranchal (2), Assam (1), Himachal (1), Andhra Pradesh (1), Haryana (1), Tamil Nadu (1), and Uttar Pradesh (1). In

housing, and transport and communication infrastructure within the most backward districts in order to pull the most deprived people up from the bottom quartile.

Figure 5

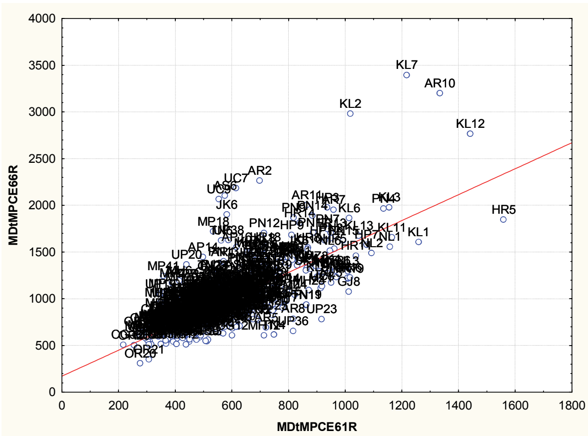
Scatterplot of the mean MPCEs of the bottom quartile in the 61st and 66th rounds of the Household Income Expenditure Survey
(r = 0.85)



Note: MCB25%66R – mean MPCE of the bottom quartile in the 66th Round; MCB25%61R – mean MPCE of the bottom quartile in the 61st Round.

Figure 6

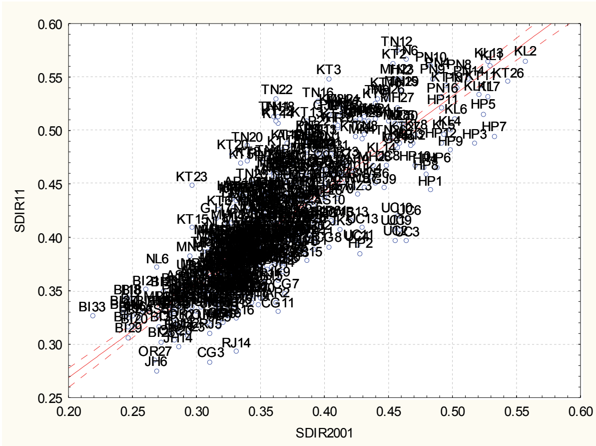
Scatterplot of the overall district mean MPCEs in the 61st and 66th rounds of the Household Income Expenditure Survey
(r = 0.72)



Note: MDtMPCE66R – overall district mean MPCE in the 66th Round; MDtMPCE61R – overall district mean MPCE in the 61st Round.

Figure 7

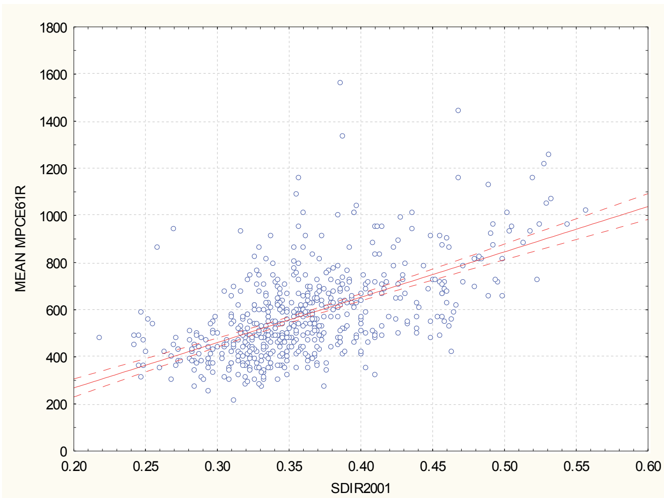
Scatterplot of the rural social development indices for 2001 and 2011
(*r* = 0.81)



Note: Here and in the following figures, SDIR2011 – rural social development index (SDI) for 2011; SDIR2001 – rural SDI for 2001.

Figure 8

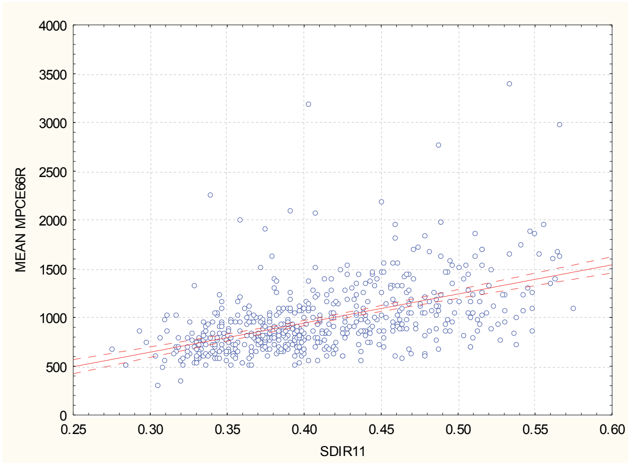
Scatterplot of the overall district mean MPCE in the 61st Round of the Household Income Expenditure Survey compared to the rural social development index for 2001
(*r* = 0.58)



Note: MDtMPCE61R – overall district mean MPCE in the 61st Round.

Figure 9

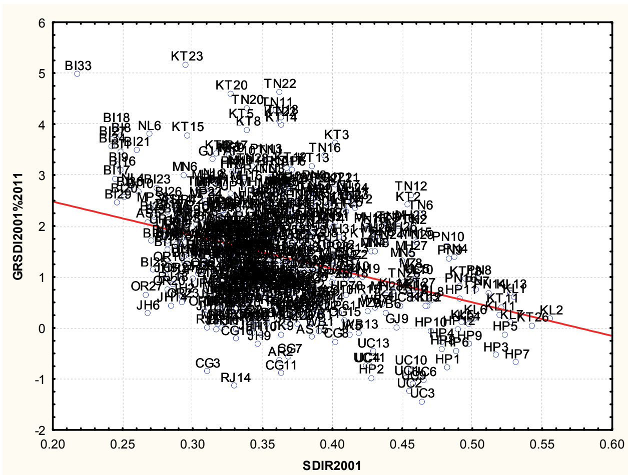
Scatterplot of the overall district mean MPCE in the 66[th] Round of the Household Income Expenditure Survey compared to the rural social development index for 2011
(r = 0.51)



Note: MDtMPCE66R – overall district mean MPCE in the 66th Round.

Figure 10

Scatterplot of the growth rate of SDIR compared to SDIR, 2001–2011
(r = –0.35)



Note: GRSDIR – growth rate of the rural SDI.

Table 7

Regression summary when the dependent variable is the district-level mean MPCE in the 61st Round of the Household Income Expenditure Survey

Denomination	β	Standard error of β	*B*	Standard error of *B*	*t*(554)	*p*-level
Intercept			–181.51	45.48	–3.99	0.000075
WPIR2001	0.24	0.04	693.13	106.20	6.53	0.000000
HCIR2001	–0.04	0.04	–72.44	80.88	–0.90	0.370857
HHIR2001	0.42	0.04	1,056.67	107.07	9.87	0.000000
TCIR2001	0.24	0.04	464.41	75.37	6.16	0.000000

Note: Here and in the following tables, WPIR – regional work participation index; HCIR – regional human capital index; HHIR – regional health and housing index; TCIR – regional transport and communication index. Adjusted $\bar{R}^2$ = 0.40, *F*(4,554) = 92.60, $p < 0.0000$, *SEE* (standard error of estimates) = 148.11.

Table 8

Regression summary when the dependent variable is the mean MPCE in the 61th Round of the Household Income Expenditure Survey

Denomination	β	Standard error of β	*B*	Standard error of *B*	*t*(554)	*p*-level
Intercept			–72.83	228.63	–0.32	0.750195
WPIR2001	–0.14	0.04	–770.66	236.79	–3.25	0.001205
HCIR2001	0.12	0.04	969.60	325.23	2.98	0.002996
HHIR2001	0.50	0.05	1,354.01	128.87	10.51	0.000000
TCIR2001	0.13	0.04	484.44	152.78	3.17	0.001604

Note: Adjusted $\bar{R}^2$ = 0.31, *F*(4,554) = 63.36, $p < 0.0000$, *SEE* = 303.79.

Table 9

Regression summary when the dependent variable is the mean MPCE in the 66th Round of the Household Income Expenditure Survey

Denomination	β	Standard error of β	*B*	Standard error of *B*	*t*(524)	*p*-level
Intercept			249.22	159.07	1.57	0.12
WPIR11	–0.20	0.04	–785.28	164.98	–4.76	0.00
HCIR11	0.08	0.04	509.97	227.52	2.24	0.03
HHIR11	0.67	0.04	1,371.01	90.42	15.16	0.00
TCIR11	0.13	0.04	400.92	107.32	3.74	0.00

Note: This regression was run after deleting 29 outliers according to Cook's distance statistics. Adjusted $\bar{R}^2$ = 0.45, F(4,524) = 108.39, $p < 0.0000$, *SEE* = 201.55.

Table 10

Regression summary when the dependent variable is the lowest 25% of the population in the 66th Round of the Household Income Expenditure Survey

Denomination	β	Standard error of β	B	Standard error of B	$t(524)$	p-level
Intercept			0.46	0.12	3.75	0.00
WPIR11	0.04	0.05	0.09	0.13	0.74	0.46
HCIR11	–0.04	0.05	–0.16	0.18	–0.89	0.38
HHIR11	–0.15	0.06	–0.18	0.07	–2.57	0.01
TCIR11	–0.14	0.05	–0.25	0.08	–2.99	0.00

Note: This regression was run after deleting 30 outliers according to Cook's distance statistics. Adjusted $\overline{R}^2$ = 0.05, $F(4,524) = 8.2356$, $p < 0.0000$, $SEE = 0.1556$.

Table 11

Regression summary when the dependent variable is the mean MPCE for the lowest 25% of the population in the 66th Round of the Household Income Expenditure Survey

Denomination	β	Standard error of β	B	Standard error of B	$t(524)$	p-level
Intercept			–41.58	49.20	–0.85	0.40
WPIR11	–0.17	0.04	–204.35	51.03	–4.00	0.00
HCIR11	0.26	0.04	467.24	70.38	6.64	0.00
HHIR11	0.60	0.05	362.63	27.97	12.96	0.00
TCIR11	–0.08	0.04	–71.74	33.20	–2.16	0.03

Note: This regression was run after deleting 30 outliers according to Cook's distance statistics. Adjusted $\overline{R}^2$ = 0.40, $F(4,524) = 88.163$, $p < 0.0000$, $SEE = 62.35$.

Conclusion

In this study sub-state level disparities were measured across the rural districts in terms of the bottom 25% of the population, their per capita monthly consumption expenditure, and the various SDIs widened during the highest growth in India, particularly since 2004/2005. This period coincides with the complete implementation of the MGNREGA across all districts. The rising horizontal rift among people was accompanied by stagnant average consumer expenditure among the destitute class. This clearly suggests that the advantage of the high growth phase was exploited mostly by the relatively rich people in the wealthier districts, even in backward states. The poor are getting poorer and upper and middle classes are getting richer in these districts. The small MPCE improvement for the lowest 25% of the population in Andhra Pradesh, Assam, Chhattisgarh, Himachal Pradesh, Kerala, North

East, Orissa, Rajasthan, Sikkim, Uttar Pradesh, and West Bengal is nullified by the increase in the proportion of people in this quartile. Economic inequality measured from real consumption expenditure did not decline in Indian districts at large during the last decade. Overall, the new duality between affluence and deprivation is more sweeping across the poorer states, where the bottom 25% people, being unskilled and without any dependable asset base, could not reap the benefits of the MGNREGA and globalisation as such. The poorest districts with more than 50% of their population in the bottom quartile are mainly located in the Central and Eastern regions. The number of the wealthiest districts with zero to 10% of their population in the bottom quartile has increased in the Western part of India. The mid-level districts with 26 to 50% of their population in the bottom quartile are spread across India, with preponderance in Eastern India. Eastern India, excluding the North East, is lagging behind the rest of India in terms of both consumption expenditure and SDIs.

The results of the multivariate linear regressions suggest that both textbook logic and the apparent relationship between various infrastructure facilities such as work participation, health and housing, human capital, and transport and communication on the one hand, and overall purchasing power at the district level, the proportion of poor in the bottom quartile, and their purchasing power on the other do not directly lead to any universally expected outcomes among diverse Indian districts. There is no change at all in the relative positions of the districts in terms of any SDI, overall district MPCE, the bottom quartile, or the mean MPCE of the bottom quartile. There is, if at all, a very weak link between SDIs and the bottom quartile.

There is hardly any consensus among economists and policy makers in India on the economics of development in post-independent India. What happened during last 70 years appears to be a 'default achievement' without any positive impact of the democratic governance system. The only consensus is that 'spatial disparity' is the main threat to the democratic governance system, leading to rising internal insurgency and external terrorist links through insider rift. Some hidden barriers crept into the governance system and 'delivery' mechanism, particularly during the high growth phase of the last two decades. In the case of funds disbursed for social capital formation in rural areas, including the ongoing MGNREGA, the middle to lower strata of the public administration system may be blamed for diverting and misappropriating funds in a situation in which the incentives and penalties are equally messy'. It is believed that a wide range of fraudulent methods are in use to create and maintain a below poverty line list that is dominated by party members, underage people, duplicate names, and deceased people instead of the genuinely poor, needy, and deprived. There are many such examples of discrepancies between the allocation of development funds and the target population intended for overall development spread across the districts.[17]

It took almost seven decades for Indian policymakers to understand that the issues related to poverty, inequality, and growth are not linearly linked to the process of infrastructure creation through decentralisation, democracy, and development. Targeting one in isolation from the others has become the 'conventional wisdom' in India. This has made the role of policy darker than ever before, even with sincere efforts from the top planning authorities. The widespread horizontal rift and limited vertical mobility by the educated and already rich represent intense and complex asymmetries at sub-state levels. The 'state' as a unit of analysis is largely responsible for ignoring the rising hidden rifts. Strict enquiries are required to diagnose where the largest rural development fund in human history (MGNREGA) went. It has tremendous lessons for many developing countries in various continents to learn from this programme. Intensive research with extensive ground-level data is needed to further probe the long-standing persistence of poverty and rising inequality, and a different indigenous type of development economics is required to understand Indian society, polity, and economy. Unless this is done, policy ineffectiveness will continue to baffle the honest thinkers in the Indian democracy.

Acknowledgement

We specially acknowledge Ananta K. Das Graphic Designer (Kharagpur, India) for his help and suggestions.

REFERENCES

ALKIRE, S.–ROCHE, J.–SETH, S.–SUMMER, V. (2014): Identifying the Poorest People and Groups: Strategies Using the Global Multidimensional Poverty Index *OPHI Working Paper* No. 78., Oxford University, Oxford.

ALKIRE, S.–SETH, S. (2015): Multidimensional Poverty Reduction in India between 1999 and 2006: Where and How? *World Development* 72: 93–108. https://doi.org/10.1016/j.worlddev.2015.02.009

ANNAN, K. (2016): The World Has a Leadership Crisis, Not a Knowledge Crisis https://www.linkedin.com/pulse/world-has-leadership-crisis-knowledge-kofi-annan.

[17] As Shankar and Gaiha (2013) point out, despite an ambitious workfare scheme through the MGNREGA, the statistics reveal a bleak picture. It was observed that money has continually been siphoned off from the scheme. Ghosh and Roy (2015) use primary data from the MGNREGA official evaluation to show that there are numerous ways money has stolen, including false documentation, false worker lists, and a significant proportion of missing assets. (See also Ghosh 2016a, 2016b.) Undoubtedly, the demand-driven nature of the MGNREGA and the 2005 enactment of the Right to Information Act replacing the Freedom of Information Act of 2002 have played an enormous role in curbing the age-old and established system of corruption in the name of decentralisation and democracy. However, their successes depend upon stringent punishment and the security of applicants. This led Sainath (1996) to write that everybody loves a good drought almost 20 years ago.

AZIS, I. J. (2013): Complex Interplay of Factors in Institutional Model of Decentralization: Theory and Application *Regional Statistics* 3(1): 3–29 https://doi.org/10.15196/RS03101

BARRO, R. J.–SALA-I-MARTIN, X. (1995): *Economic Growth* McGraw Hill, New York.

BASU, K. (2006): Globalization, Poverty, and Inequality: What is the Relationship? What Can Be Done? *World Development* 34(8): 1361–1373. https://doi.org/10.1016/j.worlddev.2005.10.009

BASU, K. (2010): *Beyond Invisible Hand: Groundwork for a New Economics* Penguin Books, New Delhi.

CHAUDHURY, S.–GUPTA, N. (2007): *Divergence in the Levels of Living in Indian Districts - A Profile* National Seminar on NSS 61st Round Survey Results, 29–30th October, New Delhi.

DAS, U. (2015): Does Political Activism and Affiliation Affect Allocation of Benefits in the Rural Employment Guarantee Programme: Evidence from West Bengal, India *World Development* 67: 201–217. https://doi.org/10.1016/j.worlddev.2014.10.009

DASGUPTA, A.–GAWANDE, K.–KAPUR, D. (2017) : (When) do antipoverty programmes reduce violence? India's rural employment Guarantee and Maoist conflict *International Organization* 71(3): 605–632. https://doi.org/10.1017/S0020818317000236

DEATON, A.–DREZE, J. (2002): Poverty and Inequality in India: A Re-examination *Economic and Political Weekly* 3 September. https://www.jstor.org/stable/pdf/4412578.pdf (downloaded: Sept. 2018)

DEININGER, K.–LIU, Y. (2013): *Welfare and Poverty Impacts of India's National Rural Employment Guarantee Scheme Evidence from Andhra Pradesh* (No. 1289) IFPRI Discussion Paper, Washington.

DREZE, J.–SEN, A. (1995) *India: Economic Development and Social Opportunity* OUP, New York.

DREZE, J.–SEN, A. (2002): *Development and Participation* OUP, New Delhi.

DREZE, J.–SEN, A. (2013): *An Uncertain Glory: India and Its Contradictions* Allen Lane, London.

DUBEY, A.–GANGOPADHYAY, S. (1998): *Counting the Poor? Where are the Poor in India? Sarvekshana Annual Report No. 1.* Dept. of Statistics, Govt. of India, New Delhi.

EGRI, Z.–TÁNCZOS, T. (2015): Spatial Layers and Spatial Structure in Central and Eastern Europe *Regional Statistics* 5(2): 34–61. https://doi.org/10.15196/RS05203.

ESWARAN, M.–KOTWAL, A. (1999): *Why Poverty Persists in India? An Analytical Framework for Understanding the Indian Economy* OUP, New Delhi.

GHOSH, B. (2010): *Construction and Analysis of Regional Variations of Social Development Indices in India* Ministry of Statistics & Programme Implementation, Department of Statistics, Govt. of India, New Delhi.

GHOSH, B. (2016a): *Evaluation Study on MGNREGA: All India Coordinated Report. Planning Commission (NITI Aayog)* Govt. of India, New Delhi.

GHOSH, B. (2016b): *Evaluation Study on MGNREGA in Himachal Pradesh* Jammu & Kashmir and Uttarakhand. Planning Commission (NITI Aayog). Govt. of India, New Delhi.

GHOSH, B.–DE, P. (2005): Investigating the linkage between infrastructure and regional development in India: era of planning to globalization *Journal of Asian Economics* 30(1): 1–28. https://doi.org/10.1016/j.asieco.2004.11.002

GHOSH, B.–MARJIT, S.–NEOGI, C. (1998): Economic Growth and Regional Divergence in India, 1960 to 1995 *Economic and Political Weekly* 33(26): 1623–1630.

GHOSH, B.–ROY, A. (2015): *Dream loot for powerful: For NREGA's success, Modi government needs to tackle pilfering and impart vocational skills DNA* 10 February, Mumbai.

GOI (2005): *The National Rural Employment Guarantee Act 2005 (NREGA)* New Delhi.

GOI (2007): *Eleventh Five Year Plan, 2007–2012* Planning Commission, Government of India, New Delhi.

GOI (2012): *MGNREGA Sameeksha* Ministry of Rural Development, New Delhi.

GUPTA, N.–GHOSH, B. (2007): *Horizontal Rift versus Vertical Mobility of People across Different Expenditure Classes in Indian States in Recent Period* National Seminar on NSS 61st Round Survey Results. MOSPI, Govt. of India: 29–30th October, New Delhi.

HIMANSHU, A. S. (2004): Poverty and Inequality – I and II, Widening Disparities during the 1990s *Economic and Political Weekly* 18 September.

JAKIMOW, T. (2014): "Breaking the backbone of farmers": contestations in a rural employment guarantee scheme. *Journal of Peasant Studies*, 41(2): 1–19.

MARCESSE, T. (2018): Public Policy Reform and Informal Institutions: The Political articulation of the Demand for Work in Rural India *World Development* 103: 284–296.

NAGY, E.–NAGY, G.–DUDÁS, G. (2016): The uneven transformation of consumption spaces and the rise of new marginalities in Hungary *Regional Statistics* 6(2): 149–172. https://doi.org/10.15196/RS06208

NEMES NAGY, J.–TAGAI, G. (2011): Regional inequalities and the determination of spatial structure *Regional Statistics* 1(1): 15–28.

ORTIZ, I.–CUMMINS, M. (2011): *Global Inequality: Beyond the Bottom Billion: A Rapid Review of Income Distribution in 141 Countries* UNICEF, New York.

PÁGER, B.–ZSIBÓK, ZS. (2014): The measurement of territorial differences in the information infrastructure in Hungary and the South Transdanubian Region *Regional Statistics* 4(2): 55–70. https://doi.org/110.15196/RS040204

PÉNZES, J. (2012): Changes in the spatial income structure of North-Eastern Hungary after the change of regime *Regional Statistics* 2(1): 90–107.
https://doi.org/10.15196/RS02107.

RAVALLION, M. (2012): Corruption in the MGNERGS: Assessing an Index *Economic & Political Weekly* 25 February.

SAINATH, P. (1996): *Everybody Loves a Good Drought: Stories from India's Poorest Districts* Penguin Books, New Delhi.

SHANKAR, S.–GAIHA, A. (2013): *Battling Corruption: Has NREGA Reached India's Rural Poor?* OUP, New Delhi.

SIPOSNÉ NÁNDORI, E. (2014): The role of economic growth and spatial effects in poverty in Northern Hungary *Regional Statistics* 4(1): 28–39.
https://doi.org/10.15196/RS04103.

SUNDARAM, K.–TENDULKAR, S. D. (2003): Poverty in India in the 1990s – An Analysis of changes in 15 major states *Economic and Political Weekly* 5 April.

THAROOR, S. (2007): *The Elephant, the Tiger & the Cell Phone* Penguin Viking, New Delhi.

THOMAS, S. K. (2007): *It's all about the lowest MPCE class* National Seminar on NSS 61st Round Survey Results. MOSPI, Govt. of India: 29–30th October, New Delhi.

VYAS V S. (2004): Agrarian Distress: Strategies to Protect Vulnerable Sections *Economic and Political Weekly* 25 December.

Determinants of electricity consumption based on the NUTS 2 regions of Turkey: A panel data approach

Semra Türkan
(corresponding author)
Hacettepe University
Faculty of Science,
Turkey
E-mail: sturkan@hacettepe.edu.tr

Gamze Ozel
Hacettepe University
Faculty of Science,
Turkey
E-mail:
gamzeozl@hacettepe.edu.tr

Previous studies have focused on the relationship in recent years between electricity consumption and socio-economic indicators, but little attention has been paid to regional differences in Turkey. This study determines the electricity consumption indicators of Turkey based on the Nomenclature of Territorial Units for Statistics (NUTS 2) regions during 2004–2011. To achieve this objective, panel data regression models, including socio-economic indicators (gross value added, and import and export amounts) were developed. The study represents the first instance of regional knowledge for electricity consumption using the panel data approach. The empirical results show that electricity consumption for the NUTS 2 regions can be modelled using fixed effects models with standard errors obtained by the White estimator. The study reveals that regional development leads to increases in electricity consumption. Furthermore, electricity demand has grown impressively in line with regional economic developments in Turkey.

Keywords:
electricity consumption,
panel regression,
fixed effects model,
Wald test

Introduction

Electricity is a key factor in improving quality of life and economic and social progress. Governments in developed and developing countries mainly pay attention to the determinants that increase electricity consumption. Hence, the policies that will increase electricity production or decrease electricity consumption are of prime importance (Kavaklioglu et al. 2009).

Electricity is among the most important energy resources in Turkey. Electricity is used in all areas of daily life, especially industrial production, the housing sector, agriculture, lighting, and heating. Although electricity consumption in the industry sectors has increased, its portion of total electricity consumption declined after the 1990s. The rate of electricity consumption in other sectors remained the same between 1970 and 2012. Following the restructuring of the electricity sector in the 1990s, both the consumption and the generation of electricity have expanded. Figure 1 shows the percentage of electricity consumption in each sector between 1970 and 2012. Figure 2 shows the sectoral distribution of the electricity consumption of Turkey in 2012.

Figure 1

Percentage of electricity consumption by sector

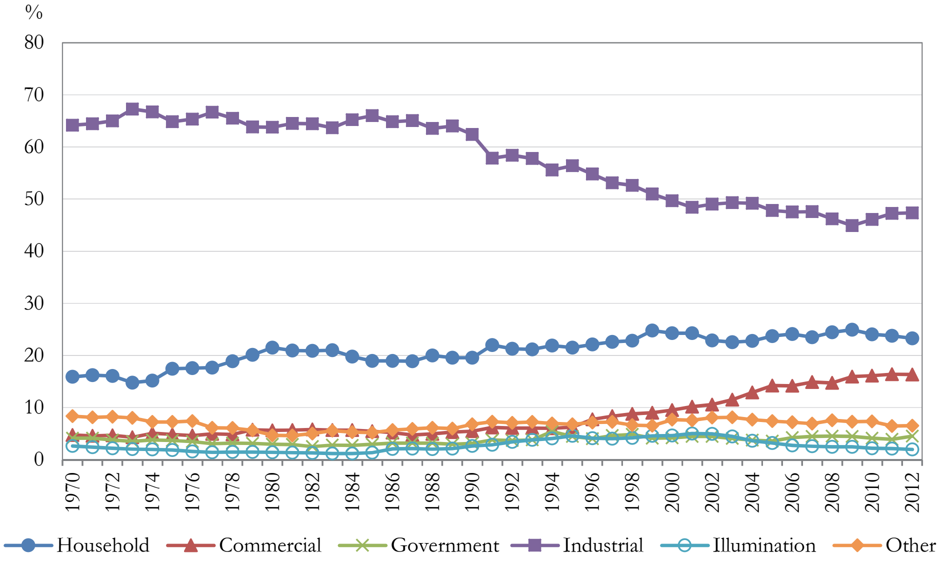
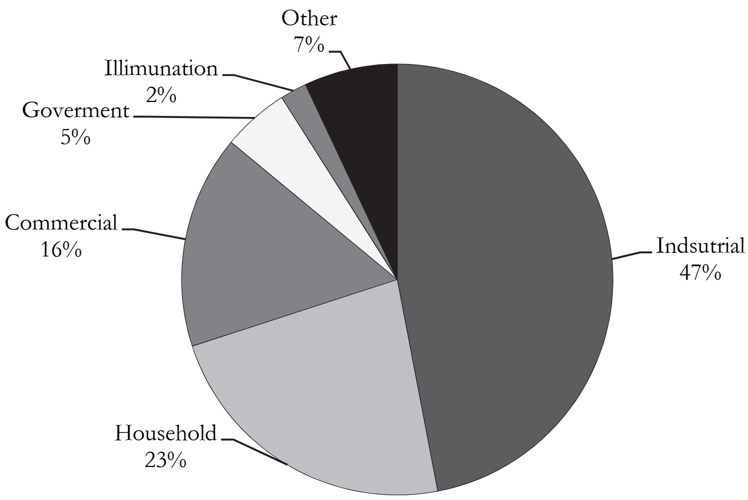


Figure 2

Distribution of electricity consumption by sector, 2012



Source: TEDAS (2012).

As observed in Figure 2, the industry sector has the largest portion of electricity consumption in Turkey, consuming 47% of the produced electricity. In Turkey, household consumption is second at 23%, reflecting the housing sector's second largest share of electricity consumption at 1,169 GWh in 1970 and 44,832 GWh in 2012. Lighting is the sector with the lowest electricity consumption at 2% (TEDAS 2012).

Turkey's electricity demand has rapidly increased during the past few years, and future increases are likely to continue given the country's rapid economic growth, young population, and other worldwide energy developments. According to the International Energy Agency (IEA), energy consumption will continue to increase at an approximate annual growth rate of 4.5% from 2015 to 2030. Given Turkey's limited energy resources, the country depends on imported energy sources for electricity production.

Turkey's energy sources include hydropower, geothermal, lignite, hard coal, oil, natural gas, biomass (wood, animal, and plant wastes), solar, and wind energy. Nuclear energy has not yet been included in the country's energy sources (Erdogdu 2010), but Turkey will begin to utilize nuclear energy to produce electricity in the next few decades. Most of the electricity is generated using fossil fuel sources. Fossil fuels contributed approximately 60% of Turkey's total energy consumption in the 1970s (Kentel–Alp 2013). This ratio reached 88% in 2011, with natural gas (32%) in first place, followed by coal (29%) and oil (27%) (Uzlu et al. 2009). However, fossil fuel reserves in Turkey are quite limited. Turkey's energy policy mainly focuses on improvements in sufficient and secure energy sources. To address the dependence on imported fossil fuels and to take the necessary precautions, Turkey's future electricity consumption should be known realistically. Thus, a critical factor for Turkey's energy policy is an accurate assessment of its future energy needs.

The relationship between energy consumption and economic indicators is analysed in several studies. In these studies, energy or electricity consumption was generally used as a dependent variable. We focus on investigating the relationship between regional economic indicators and electricity consumption for the Nomenclature of Territorial Units for Statistics (NUTS 2) regions of Turkey (Akay–Atak 2007, Jebaraj–Iniyan 2006). For regional policy, identifying the key industries within a given region, analysing their spatial and network dependencies, measuring their dependencies from weaker to stronger and knowing how to influence them to improve their economic performance are the essential steps (Jarosi 2017). Regression models, time series, neural networks, and econometric models are the most used statistical methods in modelling and forecasting of energy or electricity research. Regression models have been the most popular modelling method to determine the factors that affect and predict electricity consumption (Yumurtaci–Asmaz 2004). In particular, Zhang–Zhao (2014) used the panel data approach to estimate CO_2 emissions in China. However, panel data analysis has never been used for either electricity consumption or Turkey.

Research on predicting electricity consumption in Turkey has been conducted by the Ministry of Energy and Natural Resources (MENR) and the State Planning Organization (SPO). The MENR uses the regression model to analyse energy demand that historically has not produced reliable intermediate- to long-term results for Turkey (MENR 2014). In recent years, the Regional Competitiveness Operational Programme (RCOP) has been a basic document for the implementation of the Instrument for Pre-Accession in Turkey. The goal of the programme is to promote Turkey in its preparation for membership to the European Union (EU); this is part of the Regional Competitiveness Operational Programme framework to increase the competitiveness of the Turkish economy when converging with that of the EU and to reduce regional socio-economic differences. To achieve these aims, the RCOP focuses on resources in a limited number of sectors, regions, and priorities for which the impact of programmes and contributions will be greatest (Republic of Turkey Ministry of Industry and Trade Report 2007). Hence, the regional statistics and analyses are important steps in this programme. This statement has motivated the present research to study a different modelling technique for Turkey's electricity consumption on the basis of regional economic indicators, such as gross value added (GVA) per capita, imports (IM), and exports (EX). Our study focuses on modelling electricity consumption for NUTS 2 regions of Turkey, whereas most previous studies and analyses focused on modelling energy or electricity consumption for Turkey as a whole and not at the regional level. Furthermore, no study on the regional analysis of electricity consumption for Turkey used a panel data regression analysis.

Literature review

Prominent studies on electricity demand/consumption from literature search are listed in Table 1. These studies generally utilized diverse co-integration and causality tests to examine the relationship between electricity consumption and economic indicators. In these studies, separate findings based on the country/country group are obtained (Glasure–Lee 1997). In our study, we investigate the determinants of electricity consumption at the regional level. For this reason, socio-economic indicators, such as GVA per capita, imports, and exports, are determined by considering previous studies.

Table 1
Prominent research on electricity demand/consumption for different countries

Research	Method	Indicator	Dependent variable	Country
Glasure–Lee (1997)	Co-integration	GDP	Electricity consumption	South Korea
Shiu–Lam (2004)	Co-integration	GDP	Electricity consumption	China
Bianco et al. (2009)	Regression	Population, GDP per capita	Electricity consumption	Italy
Mohamed–Bodger (2005)	Regression	GDP, average price of electricity, population	Electricity consumption	New Zealand
Pao (2006)	Artificial neural network	National income, GDP, consumer price index	Electricity consumption	Taiwan
Zhou et al. (2006)	Trigonometric grey prediction	–	Electricity demand	China
Bianco et al. (2010)	Holt–Winters exponential smoothing, Trigonometric grey model with rolling mechanism	–	Electricity consumption	Romania
Zhang–Zhao (2014)	Panel regression	GDP per capita, Gini-coefficient, energy intensity, share of industry sector, urbanization	CO_2 emissions	China

As observed in Table 1, seven of these studies are related to electricity consumption. Only one study on electricity consumption used the panel data approach. The use of the panel data approach for modelling has attracted attention in recent decades. Hence, this study purposes to determine the indicators of regional electricity consumption for Turkey based on NUTS 2 regions through a panel data regression. In addition, few studies examined the relationship between economic indicators and electricity consumption in Turkey using different methods and approaches. A summary of the methods and variables used in these studies is presented in Table 2.

Research on forecasting energy demand or consumption for Turkey began in the 1960s. Regression analysis is used by the SPO for energy forecasting. Since 1984, several econometric methods have been used for energy demand or consumption forecasts (Kankal et al. 2011). Some studies on energy demand or consumption for Turkey are presented in Table 2.

Table 2

Literature search on electricity consumption or demand in Turkey

Study	Methodology	Independent variable	Dependent variable
Yumurtaci–Asmaz (2004)	Regression	Population, energy consumption, increase rates per capita	Electricity demand
Ozturk et al. (2005)	Genetic algorithm approach	–	Electricity demand
Hamzaçebi (2007)	Artificial neural networks	Transportation, agriculture, residence, industry sector	Electricity consumption
Kavaklioglu et al. (2009)	Artificial neural networks	Population, GNP, import, export	Electricity consumption
Erdogdu (2007)	Co-integration analysis, ARIMA	–	Electricity demand
Aslan (2014)	Granger causality test	–	Electricity consumption
Nazlioglu et al. (2014)	Co-integration, linear and nonlinear Granger causality test	–	Electricity consumption

As observed in Table 2, the panel data approach has never been used to analyse electricity consumption in Turkey.

Methodology

Panel data, also called cross-sectional time series data, are repeated observations of the same set of cross-section units. A panel dataset should have data on *n* cases over *t* periods for a total of ($n \times t$) observations.

Pooled ordinary least square

The simplest case of using longitudinal data arises from ignoring the panel structure of the data. The model can be written as

$$y = X\beta + e, \tag{1}$$

where $e \sim N(0, \sigma^2)$ is assumed to be identically independent distributed (iid) and, for a given X, the observations have no serial correlation and the errors are not heteroskedastic. The pooled ordinary least square (OLS) approach might be reasonable when the cross-sectional sample sizes are too small (Cancado 2005).

The pooled OLS estimator is considered a starting point for the empirical analyses. Two other models exist for an analysis of panel data in addition to the pooled OLS: the fixed effects and random effects models.

Fixed effects and random effects models

Consider the general panel data model

$$y_{it} = X_{it}\beta + \varepsilon_{it}, \quad (2)$$

where the error term $\varepsilon_{it}=a_i+\eta_{it}$ consists of individual specific effect a_i and common stochastic error term η_{it}, and η_{it} is assumed to be uncorrelated with X_{it}.

The assumption that distinguishes the fixed effects model from the random effects model is that a_i may or may not be correlated with the set of explanatory variables, X_{it}. a_i is uncorrelated with X_{it} in the random effects model and is correlated with X_{it} in the fixed effects model.

Two basic tests can determine which panel analysis should be used: the Lagrange multiplier (LM) and Hausman specification test. The LM, proposed by Breusch–Pagan (1980), is calculated to determine whether a pooled OLS regression or random effects model should be used. The LM test is easy to compute because it only requires pooled OLS residuals and is given by

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}(\sum_{t=1}^{T}\widehat{\varepsilon}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}^2} - 1\right]^2, \quad (3)$$

where N is the number of cases and T is the number of periods.

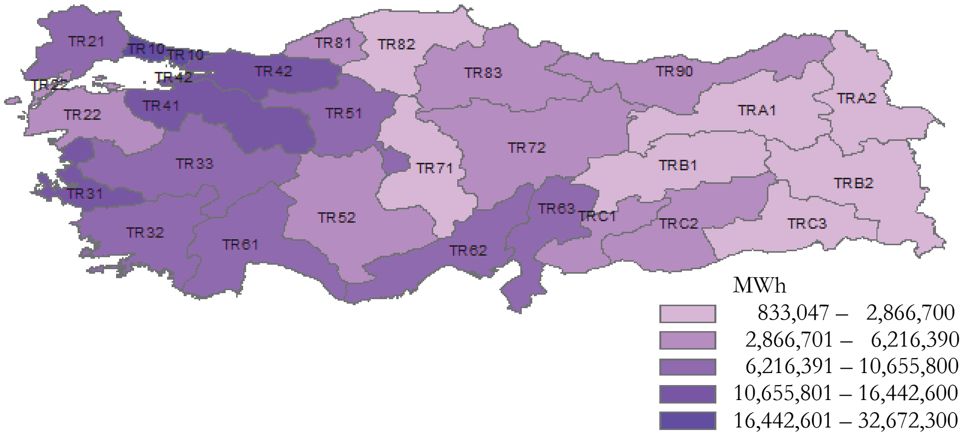
The LM is distributed as X^2 with one degree of freedom. If the LM exceeds the critical value, we conclude that the pooled OLS method is inappropriate and the random effects model is preferable. The Hausman specification test selects between the fixed effects and random effects models under the null hypothesis that the individual effects are uncorrelated with the other regressors in the model. If they are correlated (H_0 is rejected), a random effects model produces biased estimators, violating one of the Gauss–Markov assumptions; hence, a fixed effects model is preferred. Note that under the null hypothesis, the Hausman statistic is asymptotically distributed as X^2 with k degrees of freedom (Srivasan 2012).

Modelling of determinants of electricity consumption for NUTS 2 regions

Data description and modelling

In this study, electricity consumption for the NUTS 2 regions of Turkey is modelled as a function of socio-economic indicators. All of the data are obtained from TURKSTAT (http://www.turkstat.gov.tr) under titled statistics by theme. For this reason, we used a balanced panel dataset, which means that the number of periods T is the same for all individuals i of the 26 NUTS 2 regions of Turkey during 2004–2011. The list of NUTS 2 regions is provided in the Appendix. Figure 3 indicates the regional distribution of electricity consumption.

Figure 3
Map of electricity consumption by the NUTS 2 regions of Turkey, 2011



Note: For the list of the NUTS 2 regions of Turkey, see the Appendix.

Regarding the regional distribution of electricity consumption, Figure 3 indicates that the first five regions with the highest electricity consumption (the darkest regions) are TR10, TR41, TR31, TR42, and TR63. TRA2, TR82, and TRA1 are regions with the lowest electricity consumption. As is known, energy provides input for each sector of the economy. As electricity is an indispensable input in industrial production processes, energy consumption has increased. For this reason, the import and export amounts have a significant effect on electricity consumption. Turkish manufacturers are mainly located in TR10, TR41, and TR31. In recent years, TR63 has become a main new industrial district in Turkey given that electricity consumption has increased in this region.

The data used in this study are obtained from the Turkish Statistical Institute. Socio-economic indicators, such as the GVA/pop, imports, and exports – similar to precedent studies – are considered to determine electricity consumption. In this study, we examine electricity consumption for the NUTS 2 regions because the variables are determined to reflect the regional structure. Hence, we have used GVA/pop and not GDP as the regional development indicator. GVA/pop is a measure in economics of the value of goods and services produced in an area, industry, or sector of the economy. GVA/pop, one of the major macroeconomic indices in regional development, is an indicator of economic growth. An increased GVA/pop means improved living standards in a region and, thus, increased electricity consumption. Import and export amounts for Turkey are related to manufacturing processes; therefore, they strongly affect the amount of electricity consumption (Kankal et al. 2011). Electricity consumption is considered to be a dependent variable, whereas GVA/pop, import amount (IM), and export amount (EX) are considered to be independent variables. Table 3 lists the definition of the variables.

Table 3

Definition of variables and acronyms

Variable	Units of measure	Definition
Total electricity consumption	MWh	Total electricity consumption
Gross value added per capita	1,000 Turkish Lira	GVA divided by population
Imports	1,000 USD	Imports by economic activities
Exports	1,000 USD	Exports by economic activities

Figure 4 reflects the electricity consumption, GVA per capita, import amount, and export amount of Turkey from 2004 to 2011.

Figure 4

Electricity consumption and corresponding indicators in Turkey

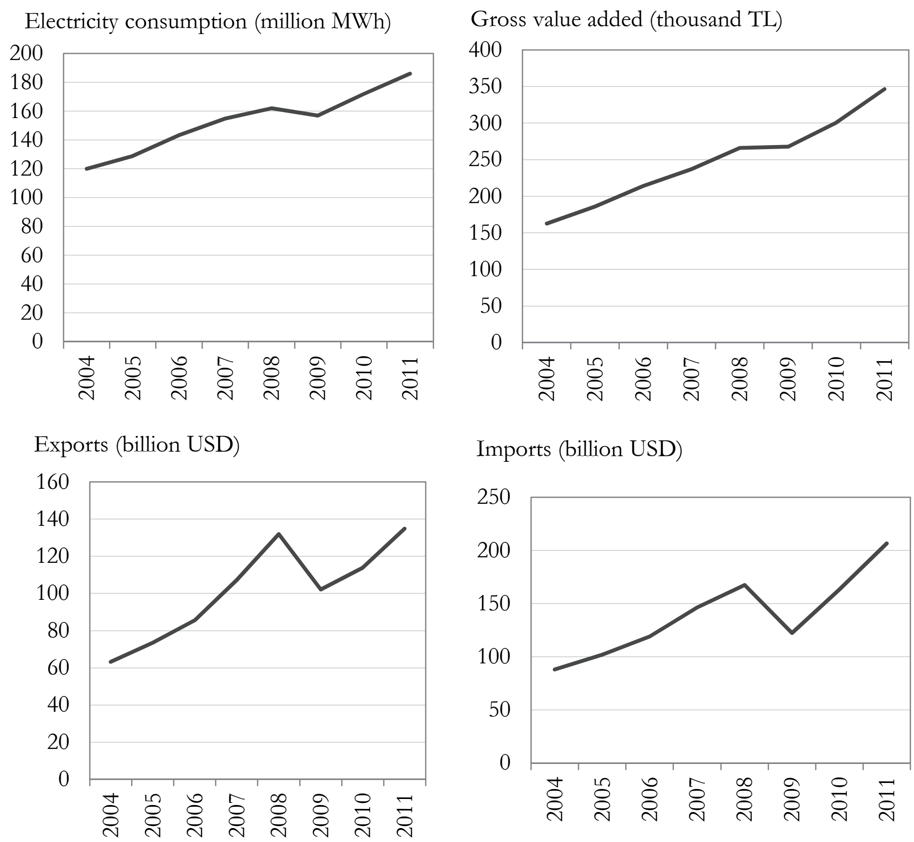


Figure 4 indicates that all of these variables in Turkey increased gradually until 2008. Because of the global crisis in 2009, imports and exports decreased aggressively, whereas electricity consumption and GVA per capita decreased slightly. After 2009, all variables steadily increased. Electricity consumption and GVA per capita in all regions indicated similar tendencies, varying slightly during 2004 and 2011.

To examine the determinants of electricity consumption in Turkey, the panel data approach has been applied. The panel log–log form of the estimable regression model can be written as follows:

$$lnEC_{it} = \beta_0 + \beta_1 \ln(GVA/pop)_{it} + \beta_2 lnIM_{it} + \beta_3 lnEX_{it} + \varepsilon_{it}.$$

The natural log form measures the GVA/pop elasticity, import elasticity, and export elasticity of electricity consumption. ε_{it} is an error term that is assumed to be independently and normally distributed.

Empirical results

First, we need to determine the panel data techniques that should be used. Before the analysis, multicollinearity is controlled by variance inflation factor (VIF) values. The VIF values of GVA/pop, import, and export are found to be 2.0946, 5.7096, and 6.6574, respectively. No multicollinearity existed because all VIF values are less than 10. For this reason, the LM test is used to determine whether the pooled OLS regression or the random effects model should be selected. The calculated LM test value (LM test value = 480.61) exceeds the tabulated chi-squared value ($X^2_{1,\alpha=0.05}$=3.841), leading us to conclude that the random effects model is more appropriate than pooled OLS. Next, we need to choose between the fixed and random effects models. For this reason, we conducted the Hausman specification test, whose null hypothesis is that the preferred model is the random effects model. The Hausman test statistic is defined as:

$$HS = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'(\mathrm{Var}(\hat{\beta}_{FE} - \hat{\beta}_{RE}))^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}),$$

where β_{SE} is the estimated parameter vector of the fixed effects model and β_{RE} is the estimated parameter vector of the random effects model. The results of the Hausman test revealed that the random effects model is not appropriate and that the fixed effects specification is preferred (p = 0.000 < 0.050).

Before estimating the fixed effects model, group-wise heteroskedasticity, autocorrelation, and cross-sectional independence are investigated (Zhang–Zhao 2014). The modified Wald test in the fixed effects regression model is used to test group-wise heteroskedasticity. A Wooldridge test is used to test serial correlation. The Pesaran cross-sectional dependence (CD) test is used to determine whether the residuals are correlated across entities. The results of the tests are presented in Table 4.

Table 4

Test results

	Test value	p-value
Modified Wald test	1537.25	0.000*
Wooldridge test	0.255	0.618
Pesaran CD test	0.783	0.434

* Means significant at the 5% confidence level.

As Table 4 indicates, the null hypotheses 'There is no serial correlation' and 'Residuals are not correlated' are not rejected, respectively ($p = 0.618 > 0.050$) and $p = 0.434 > 0.050$ Therefore, no autocorrelation and cross-sectional dependence exist in the data. However, the null hypothesis 'There is no heteroskedasticity' is rejected $p = 0.000 < 0.050$), indicating that heteroskedasticity exists.

Relying on 'robust' standard errors is common to ensure a valid statistical inference when some of the underlying regression model's assumptions are violated. One of the most popular of these alternative covariance matrix estimators was developed by Huber (1967), Eicker (1967), and White (1980). Standard errors, which are obtained with the assistance of the White estimator, are consistent even if the residuals have heteroskedasticity. As group-wise heteroskedasticity may exist within the dataset, the White estimator is used to determine more reliable estimations given violations of the classic hypothesis (Hoechle 2007). The results of the fixed effects model (FE) and the White estimator are provided in Table 5.

Table 5

Panel estimation results

	Fixed effect	White estimator
GVA/pop	0.5324*	0.5324*
	(0.0431)	(0.00746)
IM	0.0136	0.0136
	(0.0214)	(0.0311)
EX	0.035	0.035
	(0.0236)	(0.0324)
Constant	9.724*	9.724*
	(0.1823)	(0.3499)
R^2	0.73	0.73
Hausman test 92.93*		
(p-value) (0.000)		
LM test 480.61*		
(p-value) (0.000)		

*Means significant at the 5% confidence level.
Note: Figures in parentheses are the standard errors.

Table 5 indicates that the coefficient of GVA/pop is statistically significant at the 5% level. The coefficient of GVA/pop (0.5324) is significantly positive, implying that electricity consumption will increase when regional development is improved.

Conclusion

Turkey is strategically located at the crossroads of the world's largest oil and natural gas routes. The Turkish economy is currently among the fastest growing in the OECD and has become the 18th largest economy in the world based on purchasing power parity. Turkey's economy is driven by strong productivity gains and robust growing private consumption, investments, and exports, and has not been hindered by reductions in government consumption and investment. Given the country's growing population and ongoing industrialization, Turkey's electricity consumption has been increasing. For these reasons, Turkey is a source of interest in the energy economics literature and has therefore been examined by numerous studies analysing the relationship between electricity demand or consumption and national income. In contrast, however, the findings are still indecisive. The recently introduced panel data approach provides a flexible and powerful tool for electricity consumption. This approach has never been used for Turkey. Therefore, the contribution of this study is to model the socio-economic indicators of electricity consumption for the NUTS 2 regions of Turkey using the panel data approach. Regional data are used because of the differences among the regions in terms of economic, social, and cultural aspects. The major conclusion of this research is that electricity consumption in the NUTS 2 regions can be modelled as a function of economic indicators using panel data regression models. This study reveals that regional economic growth affects electricity consumption, indicating that increases in economic growth increase electricity consumption. As a result of the empirical findings, we determine that electricity consumption in regions is positively affected by the region's development level. At this point, energy policies aimed at reducing electricity consumption may negatively affect the growth in the regions. This study represents an initial study that determines the economic indicators of electricity consumption at the regional level. This study has been constrained to the years from 2004 to 2011 given a lack of availability statistics.

Appendix

List of NUTS 2 regions of Turkey:

TRA1: Bayburt, Erzincan, Erzurum
TRA2: Ağrı, Ardahan, Iğdır, Kars
TRB1: Bingöl, Elazığ, Malatya, Tunceli
TRB2: Bitlis, Hakkâri, Muş, Van
TRC1: Adıyaman, Gaziantep, Kilis
TRC2: Diyarbakır, Şanlıurfa
TRC3: Batman, Mardin, Siirt, Şırnak
TR10: İstanbul
TR21: Edirne, Kırklareli, Tekirdağ
TR22: Balıkesir, Çanakkale
TR31: İzmir
TR32: Aydın, Denizli, Muğla
TR33: Afyonkarahisar, Kütahya, Manisa, Uşak
TR41: Bilecik, Bursa, Eskişehir
TR42: Bolu, Düzce, Kocaeli, Sakarya, Yalova
TR51: Ankara
TR52: Karaman, Konya
TR61: Antalya, Burdur, Isparta
TR62: Adana, Mersin
TR63: Hatay, Kahramanmaraş, Osmaniye
TR71: Aksaray, Kırıkkale, Kırşehir, Nevşehir, Niğde
TR72: Kayseri, Sivas, Yozgat
TR81: Bartın, Karabük, Zonguldak
TR82: Çankırı, Kastamonu, Sinop
TR83: Amasya, Çorum, Samsun, Tokat
TR90: Artvin, Giresun, Gümüşhane, Ordu, Rize, Trabzon

REFERENCES

AKAY, D.–ATAK, M. (2007): Grey prediction with rolling mechanism for electricity demand forecasting of Turkey *Energy* 32(9): 1670–1675. https://doi.org/10.1016/j.energy.2006.11.014

ASLAN, A. (2014): Causality between electricity consumption and economic growth in Turkey: An ARDL bounds testing approach *Energy Sources, Part B: Economics, Planning, and Policy* 9(1): 25–31. https://doi.org/10.1080/15567241003681882

BIANCO, V.–MANCA, O.–NARDINI, S. (2009): Electricity consumption forecasting in Italy using linear regression models *Energy* 34(9): 1413–1421. https://doi.org/10.1016/j.energy.2009.06.034

BIANCO, V.–MANCA, O.–NARDINI, S.–MINEA, A. A. (2010): Analysis and forecasting of nonresidential electricity consumption in Romania *Applied Energy* 87(11): 3584–3590. https://doi.org/10.1016/j.apenergy.2010.05.018

BREUSCH, T. S.–PAGAN, A. R. (1980): The Lagrange multiplier test and its applications to model specification in econometrics *The Review of Economic Studies* 47(1): 239–253. https://doi.org/10.2307/2297111

CANCADO, L. P. (2005): *Economic Growth: Panel Data Evidence from Latin America* Master of Arts Thesis, Faculty of Center for International Studies of Ohio University.

ERDOGDU, E. (2007): Electricity demand analysis using co-integration and ARIMA modelling: a case study of Turkey *Energy Policy* 35(2): 1129–1146. https://doi.org/10.1016/j.enpol.2006.02.013

ERDOGDU, E. (2010): Natural gas demand in Turkey *Applied Energy* 87(1): 211–219. https://doi.org/10.1016/j.apenergy.2009.07.006

EICKER, F. (1967): Limit theorems for regressions with unequal and dependent errors In: LE CAM, L.–NEYMAN, J. (eds). *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* pp. 59–82. University of California Press, Berkeley.

GLASURE, Y. U.–LEE, A-R. (1998): Cointegration, error-correction, and the relationship between GDP and electricity: the case of South Korea and Singapore *Resource and Energy Economics* 20(1): 17–25. https://doi.org/10.1016/s0928-7655(96)00016-4

HAMZAÇEBI, C. (2007): Forecasting of Turkey's net electricity energy consumption on sectoral bases *Energy Policy* 35(3): 2009–2016. https://doi.org/10.1016/j.enpol.2006.03.014

HOECHLE, D. (2007): Robust standard errors for panel regressions with cross-sectional dependence *The Stata Journal: Promoting communications on statistics and Stata* 7(3): 281–312. https://doi.org/10.1177/1536867x0700700301

HUBER, P. J. (1967): *The behavior of maximum likelihood estimates under non-standard conditions. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1: 221–233. University of California Press, Berkeley.

JAROSI, P. (2017): Modelling Network Interdependencies of Regional Economies Using Spatial Econometric Techniques *Regional Statistics* 7(1):3–16. https://doi.org/10.15196/RS07101

JEBARAJ, S.–INIYAN, S. (2006): A review of energy models *Renewable and Sustainable Energy Reviews* 10(4): 281–311. https://doi.org/10.1016/j.rser.2004.09.004

KANKAL, M.–AKPINAR, A.–KÖMÜRCÜ, M. I.–ÖZŞAHIN, T. Ş. (2011): Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables *Applied Energy* 88(5): 1927–1939. https://doi.org/10.1016/j.apenergy.2010.12.005

KAVAKLIOGLU, K.–CEYLAN, H.–OZTURK, H. K.–CANYURT, O. E. (2009): Modeling and prediction of Turkey's electricity consumption using artificial neural networks *Energy Conversion and Management* 50(11): 2719–2727. https://doi.org/10.1016/j.enconman.2009.06.016

KENTEL, E.–ALP, E. (2013): Hydropower in Turkey: Economical, social and environmental aspects and legal challenges *Environmental Science & Policy* 31: 34–43. https://doi.org/10.1016/j.envsci.2013.02.008

MOHAMED, Z.–BODGER, P. (2005): Forecasting electricity consumption in New Zealand using economic and demographic variables *Energy* 30(10): 1833–1843. https://doi.org/10.1016/j.energy.2004.08.012

NAZLIOGLU, S.–KAYHAN, S.–ADIGUZEL, U. (2014): Electricity consumption and economic growth in Turkey: co-integration, linear and nonlinear granger causality *Energy Sources, Part B: Economics, Planning, and Policy* 9(4): 315–324. https://doi.org/10.1080/15567249.2010.495970

OZTURK, H. K.–CEYLAN, H.–CANYURT, O. E.–HEPBASLI, A. (2005): Electricity estimation using genetic algorithm approach: a case study of Turkey *Energy* 30(7): 1003–1012. https://doi.org/10.1016/j.energy.2004.08.008

PAO, H-T. (2006): Comparing linear and nonlinear forecasts for Taiwan's electricity consumption *Energy* 31(12): 2129–2141. https://doi.org/10.1016/j.energy.2005.08.010

SHIU, A.–LAM, P-L. (2004): Electricity consumption and economic growth in China *Energy Policy* 32(1): 47–54. https://doi.org/10.1016/s0301-4215(02)00250-1

SRIVASAN, P. (2012): Determinants of equity share prices in India: a panel data approach *The Romanian* Journal *of* Economic 46: 205–228.

UZLU, E.–KANKAL, M.–AKPINAR, A.–DEDE, T. (2009): Estimates of energy consumption in Turkey using neural network s with the teaching e learning-based optimization algorithm *Energy* 75: 295–303 https://doi.org/10.1016/j.energy.2014.07.078

WHITE, H. (1980): Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity *Econometrica* 48(4): 817–838. https://doi.org/10.2307/1912934

YUMURTACI, Z.–ASMAZ, E. (2004): Electric energy demand of Turkey for the year 2050 *Energy Source* 26(12): 1157–1164. https://doi.org/10.1080/00908310490441520

ZHANG, CH.–ZHAO, W. (2014): Panel estimation for income inequality and CO_2 emissions: A regional analysis in China *Applied Energy* 136: 382–392. https://doi.org/10.1016/j.apenergy.2014.09.048

ZHOU, P.–ANG, B. W.–POH, K. L. (2006): A trigonometric grey prediction approach to forecasting electricity demand *Energy* 31(14): 2839–2847. https://doi.org/10.1016/j.energy.2005.12.002

INTERNET REFERENCES

MINISTRY OF ENERGY AND NATURAL RESOURCES (MENR 2014): Energy sources, Ankara, Turkey. http://www.enerji.gov.tr/tr/inc_enerji.php (2014). Accessed 20 April 2014.

REPUBLIC OF TURKEY MINISTRY OF INDUSTRY AND TRADE (2017): The Regional Competitiveness Operational Programme Report, Ankara, Turkey. http://ec.europa.eu/enlargement/pdf/turkey/ipa/tk3_regionalcompetitivenessop_2007_en.pdf. Accessed 26 May 2014.

TURKISH ELECTRICITY DISTRIBUTION COMPANY (TEDAS (2012): Energy Statistics, Ankara, Turkey. http://www.tedas.gov.tr (2012). Accessed 11 January 2014.

Changes in the retail sector in Budapest, 1989–2017

Tamás Sikos T.
University of Miskolc
Institute of Management Sciences,
Hungary
E-mail:
sikos.t.tamas@uni-miskolc.hu
National University of
Public Service
Faculty of Political Sciences and
Public Administration,
Hungary
E-mail: sikos.t.tamas@uni-nke.hu

Keywords:
retail trade of Budapest,
shopping centres,
consumer behaviour,
geomarketing,
typology of shopping centres,
urban structure,
CBD

The retail network of in the Hungarian capital has gone through a significant change in the last 30 years. This fundamental change can be explained partly by global processes and is, to some extent, the result of the unique Hungarian or Central European situation – that is, conditions of the Socialist era that were affected by contemporary European tendencies, but the system still had its own peculiarities – and of the regime change. The changes in the last four decades have significantly altered the retail sector in Budapest and the shopping habits of its inhabitants. Quality of life has also fundamentally changed, as have living standards. In addition to the spread of mobilisation, new satellite technologies have brought considerable changes in the field of trade as well. Today, the question is not how close we are to the developed world but how fast we can gain access to suitable services and basic supplies and how these supply chains are organised. In our globalised world, a new form of harmony must be created between globality and locality.

Introduction

Although this paper describes the changes in the retail network of Budapest, some of the findings are also relevant at the national level. Shopping centres are still the 'new cathedrals' of consumption – even given their sizes – and have a stronger impact on the city's spatial structure than previous department stores did. (In 2017, the total floor area of shopping centres in Budapest exceeded 1.2 million m^2). Today, shopping centres do not only formulate the spatial structure of commercial zones but in many cases endanger the existence of department stores and small shops. The emerging conglomerates affect not only the structure of shop chains but also the constantly changing spatial structure of the city and consumer behaviour (Sikos T.–Hoffmann 2004a, 2012).

Thus, this study addresses not merely the structural changes in the Hungarian capital but also shopping centre locations and their typologies. In our previous studies, similar analyses were done on Bratislava, Slovakia, Prague, the Czech Republic, and Vienna, Austria (Sikos T. 2012, 2013), followed by a detailed analysis of Vienna and Bratislava (Mitrikova et al. 2016, Kita–Grossmanová 2014, Krizan. et al. 2018). Our studies are based on relevant theoretical and methodological foundations. Our previous typology of shopping centres, similar to this study, is based on the model developed by Dawson (1983) (Dawson–Mukoyama 2013). The strategies and trends used to select the locations that we describe in this study are also in accordance with the works of Brown (1991) and Borchert (1994, 1998). *The theoretical and methodological research on the subject examined is outstandingly relevant in light of the change in the regime* because economic processes in a planned economy operate differently from those under free market conditions. At the same time, satisfying consumer needs at the highest possible level in both quality and quantity became a priority. In my research, I concentrated not only on the processes and shopping centres at the top of the pyramid but also attempted to follow how traditional retail trade lost space (Sikos T. 2000, 2004). This study provides a good opportunity for a detailed review of retail processes.

Methodology

The method used in this research was based on processing significant quantities of data collected from a large number of questionnaires. Here, we emphasise that all of the volunteers participating in the survey were selected randomly; however, despite this fact, the sample cannot be considered representative. Data were processed using the mathematical and statistical software package SPSS 20. In addition to traditional fieldwork methods, GIS technology (Tóth–Csomós 2016) and digital analysis methods offered by Google Earth were used to analyse changes in businesses.

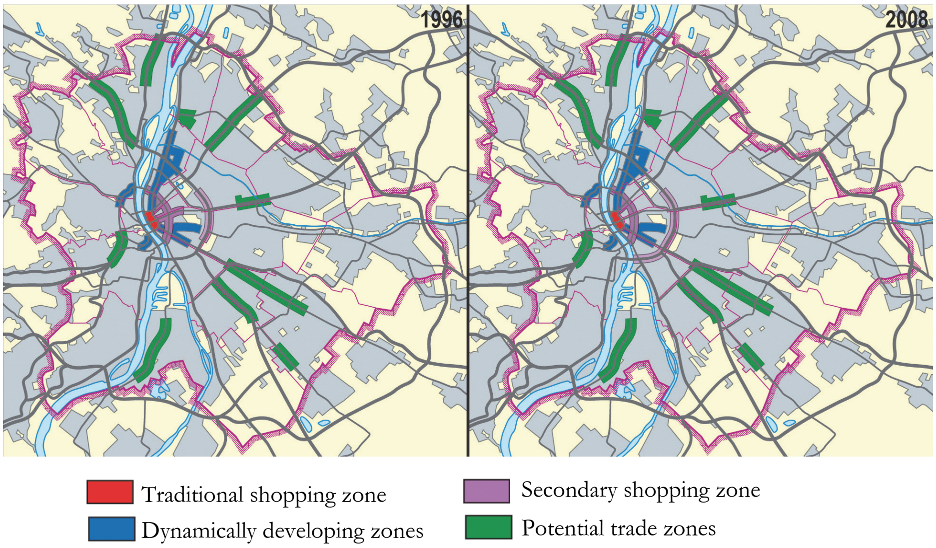
Structural characteristics of retail zones in Budapest

The evolvement of the most important retail areas in the Hungarian capital is closely connected to the development history of Budapest. In the topographic development of Budapest, among the urban development and urban regulation measures, the most important one was probably the reshaping of the downtown area, which became necessary when the Erzsébet híd (Elizabeth bridge) was built. Construction was completed during World War I; however, the modern city centre of the metropolis was only completed after the war. During this period, *the traditional shopping zone of the capital* was developed in the Vörösmarty tér–Károly körút–Kossuth Lajos utca–Danube area. This zone is still one of the most elegant and reputable business zones in the capital and, simultaneously the most expensive shopping area. Regarding its function, the zone fulfils the same role as Kärtner Strasse, the shopping street of Vienna.

In reshaping the downtown area, banning traffic from the central business district and making the northern part of Váci utca a pedestrian street were significant steps. This process started at the end of the 1970s, but the construction of the pedestrian area in the southern section of Váci utca took place only in the second part of the 1990s. The retail characteristics of the northern and southern parts of Váci utca also differ: the northern part is home to trading luxury items, whereas the southern part can be characterised as traditional retail trade. Presently, the northern and southern parts of Váci utca are clearly separated. Szabadsajtó út and Kossuth Lajos utca cross the area, which increases the feeling of separation between the two halves of the street. In the long run, however, the luxury zone will probably spread to the south and, over time, a bipolar commercial core might evolve the downtown area (see Figure 1).

Figure 1

Commercial zones of Budapest, 1996 and 2008



Source: Sikos T. (2009).

The commercial role of Rákóczi út became more articulated after the Erzsébet híd (Elizabeth bridge) was built but remained secondary relative to the northern part of Váci utca until 1996. The larger department stores of Rákóczi út (Corvin, Verseny, Csillag, and others) were closed down (2008), the range of goods they sold used to belong to the low-end market anyway, and items sold in Váci utca and its neighbourhood were high quality, up-market products.

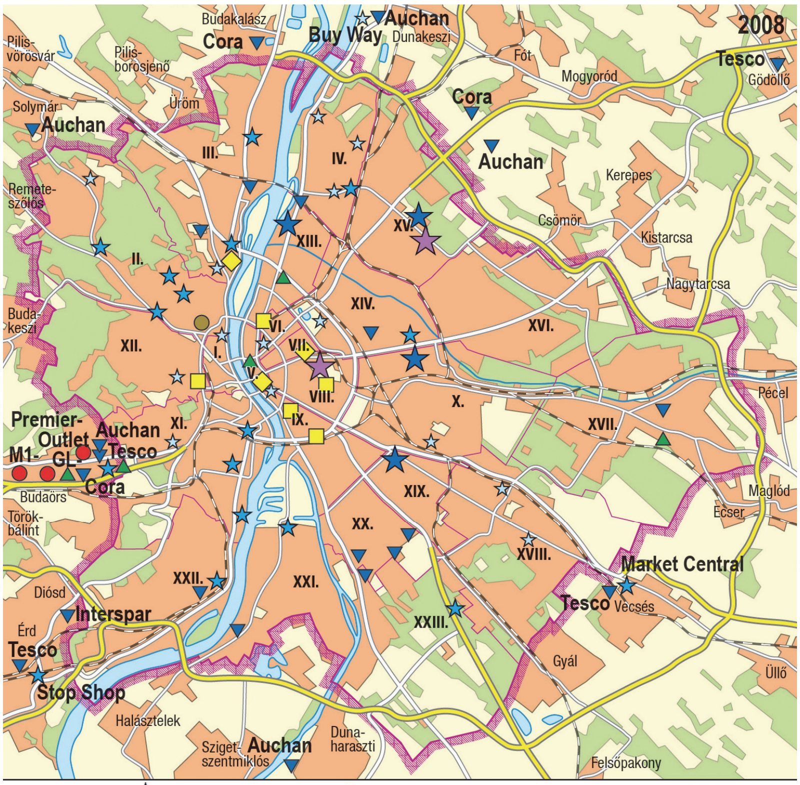
Nagykörút – located between Margit híd and Petőfi híd – is an organic part of the *secondary shopping zone*. This area is mainly characterised by small shops that had, in many cases, not more than 20–50 m^2 of floor size, sometimes even smaller. In Nagykörút, the total size of the shop floor is approximately 150,000 m^2. Recently, retail trade in Váci út has been dynamically developing and shops with large floor area that had to leave their previous locations because of high rental fees resettled here from the downtown area and the secondary shopping zone.

Next to roads leading out of the city, *another commercial zone* appeared in which mainly car trader companies, used car parts yards, solid fuel merchants, and building material and supplies yards, requiring large spaces, settled down. An analysis of the structure of Budapest's commercial zones indicates that the development trend is similar to that of large south European cities decades ago (see Figure 1).

In Budapest, shopping malls were first built in the 1970s (Flórián, Skála),[1] but their spectacular and explosive development started only in the 1990s. The 80 smaller retail units that we view as units belonging to a shopping centre (2008) have a significant impact on Budapest's spatial structure and its formulation (see Figure 2).

Figure 2

Types of shopping centres in Budapest, 2008



☆Neighbourhood (15) ★Regional (4) ◇Infill (3) ●Extension (1) ▼Focused (24)
★Community (16) ★Super-regional (2) □ Multi-use (5) ▲ Specialist (5) ● Outlet centres (3)

Source: Sikos T. (2009).

[1] At the location of the former Skála department store (opened in 1976), the Allee shopping centre has been operating since 1996.

In the capital, 16 shopping centres play a regional role, and the number of hypermarkets in Budapest and its agglomeration is 25. The development of the transport network in the capital could hardly keep pace with the rapid appearance of new retail units, and the hindering effect of this situation can still be sensed in some shopping centres. Until 2008, during the economic downturn, the structure of retail trade in Budapest and its agglomeration zone was developing dynamically but then came to a halt. A significant portion of the Hungarian population took out Swiss franc-denominated loans and became indebted,[2] causing purchasing power in commercial centres to decline, which negatively affected further investments. During the crisis, households started to elaborate on special strategies and ways to minimise their losses.

Effect of the crisis on retail trade

Because of the global economic crisis, the dynamism of retail trade declined (Kocziszky et. al. 2018). There was a significant break in the expansion of retail trade complexes. In this situation, even ongoing investments halted (see the Tó-park project), and new ones were not launched in the market after 2011. Among projects in progress before the crisis, the second phase of Allee (2009), Corvin Plaza (2010), Europeum (2011), Hegyvidék Bevásárlóközpont (2012), and Árkád (2013) was completed.

After 2008, the retail network in the capital was mainly expanded with 20 hypermarkets built as of 2017. However, after the completion of the second phase of Árkád (2013), no new shopping centre was built in the capital. Presently, a few new projects are being prepared, but they will not be implemented until the economic situation stabilises. The preparation process for building a shopping centre generally takes four or five years. For the Árkád shopping centre, the process took five years.

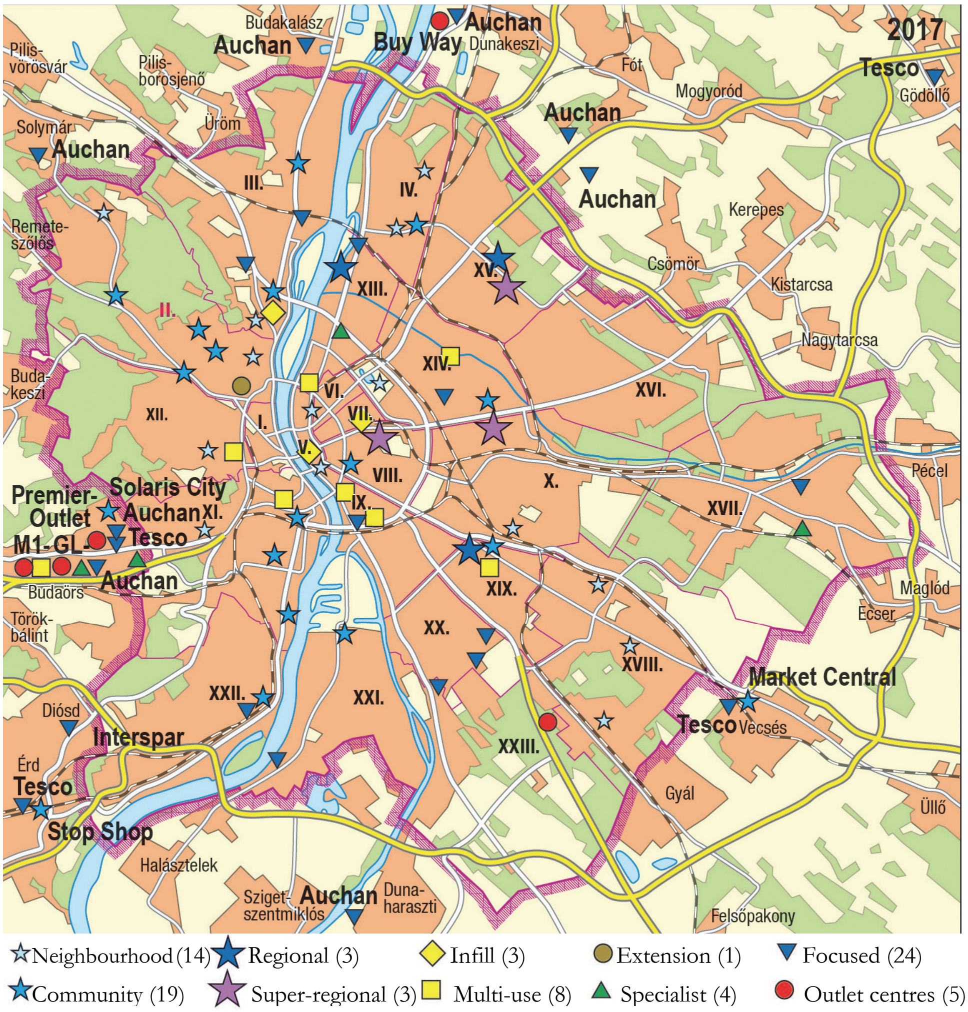
Retail trade was further burdened in this challenging economic period. First, companies with a retail profile were progressively taxed (2011), and, second, the 'Plaza stop' law (a ban on building shopping centres)[3] came into effect (2011), which stipulated that permission is needed to establish retail units.

[2] Many real estate and car buyers became indebted given changes in the exchange rate of the Swiss franc.

[3]The amendment bans building retail outlets larger than 300 m^2 (shops and shopping centres) and calls for expanding existing outlets to larger than 300 m^2 until the end of 2014 in Hungary. Therefore, until then, no shopping centres, supermarkets, hypermarkets, or discount stores were to be built. However, the amendment left a loophole: the minister of trade – based on the opinion of the committee, including the ministers of trade, environmental protection, and rural development – is authorized to make exceptions to the law. (http://koos.hu/2011/12/13/plazastop-300-m2/)

Figure 3

Types of shopping centres in Budapest, 2017



The 'Plaza stop law 2' that came into force that year[4] was extended, and it regulated the proposed changes to shopping centres. The re-division and regulation of the national tobacco market was also a sensitive issue for the operation of commercial centres.

Under these conditions, a government measure was born that banned retail units from opening on Sundays. This central regulation came into force on 15 March, 2015. Relative to previous periods, the new commercial law considerably limited the Sunday and night opening hours of retail outlets with more than 200 m^2 of floor space. Act CII of 2014, 'on prohibition of work on Sundays in the retail sector', prohibits retail units from opening between 10:00 p.m. and 6:00 a.m. on business days and Sundays, except for small retail stores operated by the owner. The decree aimed to enhance the competitive position of small shops and businesses.

This measure contradicted the liberal practices of the previous decades, strongly influenced the shopping habits of the population accustomed to continuous opening hours, and had a negative effect on people's feelings of comfort. Therefore, the measure had significant publicity among both its supporters and opponents. In addition to the social impacts that could be almost immediately sensed, the act indicated a significant rearrangement in the retail sector by changing buyers' time dimension.

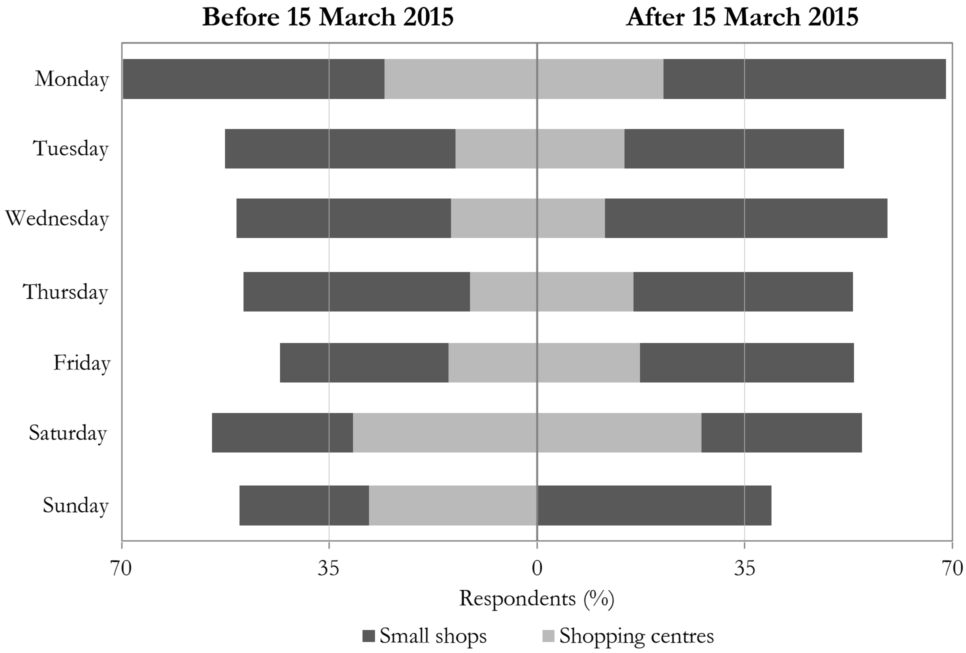
Our research also highlights this point. Data collection was carried out in two phases: before the introduction in January and February 2015 of the Sunday closure and after the act came into force in April and May. The sample size (*N*) was 151 in phase one and 121 in the second phase of the research. Respondents were selected by using a simple random sampling method (Kovács–Sikos T. 2015). The sample consisted of respondents living in the capital, large cities, small towns, and villages. However, choosing a territorially representative sample was not possible because of the sample size.

The compulsory Sunday closing affected mainly shopping centres. These retail stores with large floor sizes can be regarded as a sort of new 'city centre' and new 'cathedrals of consumption' offering not only commercial but – in a broader sense – service, cultural, and community functions as well. More than one-quarter of the people asked in our survey visited one of the shopping centres on Sundays. This group of potential buyers disappeared from the shops after 15 March, 2015 (see Figure 4).

[4] The legislator ordered that a specialized authority shall participate in the evaluation of tenders to build a shopping mall. The authority gave its resolution concerning environmental, traffic, and urban development effects on the proposed location and gravity zone of the planned investment. They participated in issuing building permissions for building or altering retail units larger than 400 m^2 of gross floor space and for expanding existing units to this size – these activities required building permission. http://www.kormanyhivatal.hu/hu/zala/hirek/februartol-hatalyos-a-plazastop-2

Figure 4

Time dimension of shopping before and after 15 March, 2015 in shopping centres and small shops



Source: Kovács–Sikos T. (2015).

After the act came into force, Sunday shopping – even though to a smaller extent – was channelled to the other days of the week; however, at the same time, turnover in shopping centres in general declined. Businesses attempted to protect themselves against Act CII of 2014 in two ways: they extended the opening hours and reduced their prices. Given these processes, the impacts meant barriers for the companies – they generated uncertainty and unpredictability in business – and resulted in tightening instead of expansionary measures.

The significant insecurity (in business and in life) to which we referred earlier had its impact on individual lives and households as well. The changes can be detected in rationalised spending. The previously flourishing, successful commercial areas and zones were rearranged, and the losers and winners could be clearly separated in the shopping centre market. The real winners in the market were Duna Plaza, WestEnd City Center, Árkád, Aréna, Europark, and Mammut, for which the secret of their success was their thorough site selection policy.

Obviously, losers can explain their failures using the same reasons; moreover, a badly chosen location in an unfortunate business environment worsened their situation. Csepel Pláza, a member of the Plaza Center group, is an example of this situation: here, the purchasing power potential of the area was incorrectly estimated. Investors opened Lurdy Ház in an area that lacked businesses, but the shopping centre is difficult to reach on public transport, and the purchasing power in the area is rather low. In the case of Új Udvar and Rózsadomb Center, the impact of incorrect site selection is worsened by the unsuitable building, unfortunate shop mix, and unattractive morphologic features.

Some shopping centres fell back from their former leading positions (e.g. Pólus Center) mainly because their role became secondary. The so-called domino principle is determinative, which indicates that new entrants continuously rewrite the rank of shopping centres in the market.

Campona is a good example of this phenomenon, with a market position that significantly improved by building a tropicarium. The shopping centre, also built in an area that had an undersupply of shops and stores, partly owes its favourable position to this undersupply and partly to the fact that it stands alone, away from its competitors and with its own, well-defined gravity zone. Opening the tropicarium resulted in a 20% increase in the number of visitors, and its previous 80% saturation index jumped to 100%. Eleven Center was not as fortunate regarding its utilisation (35%). These unfavourable numbers are the result of incorrect site selection. Although Gazdagréti lakótelep (housing estate) would mean enough purchasing power for the centre, investors failed to calculate the extraction effect of the shopping centres in the Budaörs–Biatorbágy–Törökbálint triangle.

Shopping centres in the western gate of the capital have a significant impact on the region, even though their shop network remarkably weakened after the crisis, and many stores in the area went bankrupt or were sold. It is very likely that, in the future, building of shopping centres will take a turn and hypermarkets will enjoy priority (see Figure 4). We will also have to count on a new, upcoming direction – the expansion of e-commerce – which will act increasingly as a supplement to shopping centres and hypermarkets.

Customer opinions on shopping centres

Hungarian buyers accepted, learned to like, and visited shopping centres in a relatively short period. However, acceptance does not equal a long-lasting positive attitude. For businesses, loyal and satisfied customers who return are a valuable asset that can be relied on in the long run. It was also important to examine what customers think about shopping centres when Act CII of 2014 – regulating the opening hours in the retail sector – came into force and when it was eventually withdrawn.[5]

[5] We must note that in a 2016 study of ours, when 472 people were surveyed on Facebook, two-thirds of them rejected the Sunday closure and voted for the stores to re-open (Sikos T.–Kovács 2017).

Research was carried out using 163 respondents in Budapest shopping centres who were selected through a simple random sampling method. We aim to indicate that the emotion-driven opinions of customers can be favourable if they are positive or disadvantageous if they are negative.

The opinions of the surveyed individuals on shopping centres was more positive than negative, and 54.6% of them stated that people either like very much or like these establishments, 42.9% both like and dislike them, and only 0.6% claimed to dislike them. According to this response, on a scale of 1 to 5, the attitude index is 3.62, indicating that the 'like' rating was predominant. Obviously, this score does not indicate loyalty, support, or returning because we discuss subjective feelings and generalisations. In these cases, people seemingly make abstractions because researchers ask for general opinions, but respondents always respond with what they think about the object, place, or concept in question. Shopping centres had positive ratings, and answers regarding the reasons also reflected these ratings (see Table 1).

Table 1

Attitude indices on Budapest shopping centres*, 2016

Statement	Index value
I can shop on weekends	4.79
Wide selection	4.17
Encourage wasteful expenses	3.99
Offer temptation	3.68
Increases prices of goods/services	3.67
Everything under one roof	3.65
Good experience	3.64
Planned	3.54
Comfortable	3.43
Negatively affects children	3.30
Small shops are more favourable	2.87
Helpful service	2.75
No crowd	2.64

* To calculate the attitude index, we multiply the distribution ratios by the weights of 1–5, add them up, and divide by 100. The resulting value may range from 1 to 5.

Attitude indices also underpin the previous findings that showed that weekend shopping possibilities (4.79) and wide selection (4.17) are the most determinative in accepting shopping centres; therefore, being open on Sundays is important to customers. We must also emphasise that customers do not condemn temptation (3.68) but consider it an acceptable feature of shopping centres. Similarly, putting everything under one roof is also considered a positive feature (3.65). At the same time, as another study revealed (Sikos T.–Hoffmann 2004b), service is regarded as poor (2.75), and the index hardly reaches the average.

Buyers find shopping centres crowded (2.64), which is not surprising because it is difficult to move around in most of them, especially in the early evenings and on the weekends. In vain, businesses attempt to help the situation because the width of the corridors, the size of the shops, and other passageways are set, even though they may differ for each shopping centre depending on the planned number of visitors and the expected intensity of shopping.

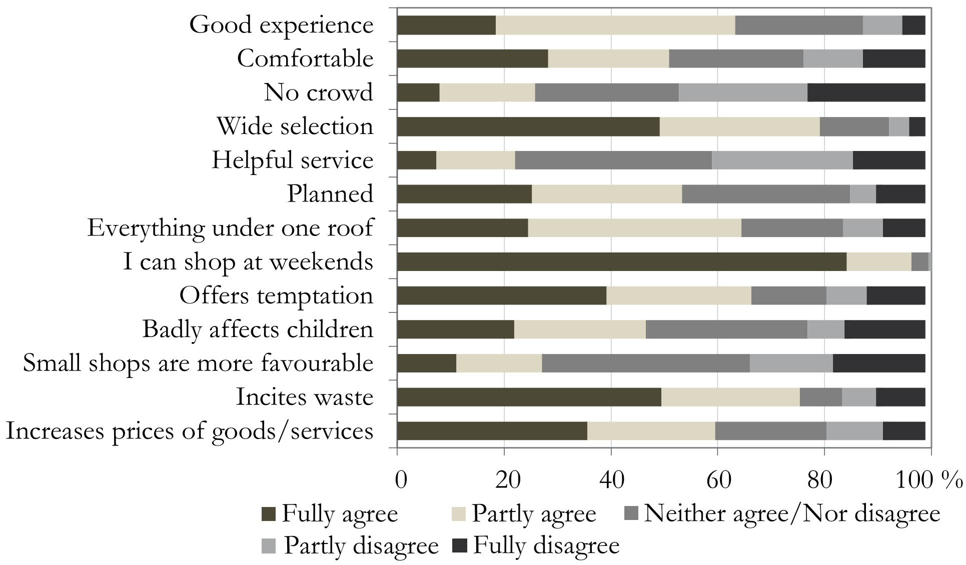
Adults are unsure when deciding on whether or not shopping centres negatively affect children. Because the relevant index (3.3) is comprised of 'both yes and no' answers, it is not negative. This is a good reason against the opponents of shopping centres, many of whom based their negative opinions on the negative impact that shopping centres have on children. Customers agree only partly with the statement that shopping centres encourage wasteful expenses (3.99) and offer temptation (3.68). The index for temptation could be higher from a marketing point of view, even if respondents partly agree with the statement (approximately a value of 4).

Also favourable is that the majority of the people surveyed stated that shopping centres do not incite waste, namely, make us buy products we do not need. People are not tempted by the large volume of products and can resist buying them, which enables them to avoid the uncomfortable feeling of cognitive dissonance.

Figure 5

Reasons for choosing shopping centres, 2016

Do you agree with the following statements? (Budapest inhabitans)



We state that buyers leaving shopping centres are mostly satisfied, and feeling regret after buying something is rare. This feeling might also play a role in why Hungarian buyers grew to like retail units with large floor sizes (see Figure 5).

The survey indicated that the majority (55%) of those surveyed have the opinion that prices in shopping centres are fairly high, and that quality is regarded as average by the overwhelming majority (72%). This situation cannot be viewed as favourable because it reflects a shift in the price–quality ratio and that the probability that buyers will return decreases. In other words, according to customers, stores in shopping centres charge higher prices than they should considering the quality of the products, as indicated by the attitude indices (see Table 2).

Table 2

Opinion of customers on prices applied by shopping centres and the quality of goods, 2016

Buyer sample	Price	Quality
Total sample	3.53	3.21
Intellectual worker	3.48	4.12
Holding academic degree	2.75	2.62
Inhabitant of Budapest	4.61	4.61
21–30 age group	2.75	4.61

Note: Maximum value = 5.

The opinion of individuals holding academic degrees and the 21–30 age group is significantly different from the average. People living in the capital consider prices to be high, and youngsters judge them favourably. The same refers to quality. Among customers in shopping centres, we often find that young, white-collar workers – especially women – apparently consider prices close to the average and quality higher than the average. This finding is in contrast to the entire sample, which judges the quality to be the same as anywhere else. People with academic degrees view both prices and quality as average and only buy products in shopping centres in exceptional cases. In this segment, the negative attitude that characterises environmentally conscious social groups can be detected.

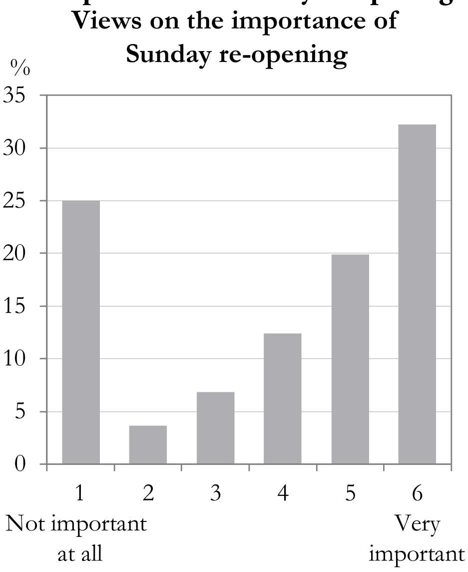
Because young, intellectual workers are overrepresented in the sample (43–44%), the findings of this research apply to this segment. This situation is favourable for stores in shopping malls because the most frequent buyers view prices as reasonable and regard the quality as good. They also consider the operation of shopping centres as important.

The Hungarian government – in reaction to pressure from shopping centres and buyers – withdrew Act CII of 2014, a step that was received positively. Of the people asked on Facebook through the assistance of a Google online questionnaire, more than 65% voted for Sunday openings and 35% preferred Sunday closures (see

Figure 6). Our analysis also justified that shopping habits had changed. The spread of e-commerce must be increasingly considered, whereas comfortable shopping is becoming a priority for consumers.

Figure 6

Importance of Sunday re-opening and changes in shopping habits, 2016



Conclusion

Based on the described and multilaterally evaluated results of our research, the following important conclusions can be drawn about the changes in Budapest's retail sector.

- In the last 40 years, the retail network in Budapest was built, and its further development and renewal depend on future market conditions.
- The global economic downturn set back the expansion of shopping centres in Budapest, and the last significant investment was completed in 2013 when the number of shopping malls was 101. Presently, two malls are under construction in the Hungarian capital (Etele Plaza and Óbudai Plaza).
- Shopping centres were received positively, and buyers remained committed to their re-opening on Sundays even when the government regulated opening hours in the retail sector through Act CII of 2014. The majority of Hungarian society (two-thirds of them) did not accept the Sunday closure because it significantly hindered their schedules and costs (travel costs, waiting, and crowds).

- Between 2014 and 2016, consumer behaviour changed remarkably (in shop preferences and spending). The division in purchase costs also changed; in particular, the share of government-supported retail units – small shops – increased in the case of FMCG products, and the turnover of other market players did not considerably change.
- Regulating the opening hours also affected consumer habits. The changes resulted in an increase in e-commerce, a comfortable shopping alternative for buyers.

REFERENCES

BORCHERT, J. G. (1994): Urban marketing. A Review In: BRAUN, G. O.–KORN, H.–LINEMANN, A.–SCHULTZ, G.–WOOSNAME, J.–WOOSNAM *Managing Marketing of Urban Development and Urban Life*. pp. 415–428., Deitrich Reimer, Berlin.

BORCHERT, J. G. (1998): Spatial dynamics of retail structure and the venerable retail hierarchy *GeoJournal* 45(4): 327–336. https://doi.org/10.1023/a:1006976407047

BROWN, S. (1991): Tenant placement in planned shopping centres: implications of an observation survey *Journal of Property Research* (8): 179–187. https://doi.org/10.1080/09599919108724032

DAWSON, J. A (1983): *Shopping Centre Development* Longman, London.

DAWSON, J. A.–MUKOYAMA, M. (2013): Global Strategies in Retailing: Asian and European Experiences. Routledge, London.

KITA, P.–GROSSMANOVÁ, M. (2014): Reflection of Bratislava Retail Network in Selected Aspects of Consumer Behaviour *Verslas: Teorija ir Praktika/Business: Theory and Practice* 15(3): 279–284. http://dx.doi.org/10.3846/btp.2014.28

KRIZAN, F.–BILKOVA, K.–BARLIK, P.–KITA, P. (2018): Spatial distribution of consumer preferences: Case study of shopping malls in Bratislava *Theoretical and Empirical Researches in Urban Management* 13(1): 13–21.

KOCZISZKY, GY.–BENEDEK, J.–SZENDI, D. (2018): The impact of the 2008 financial crisis on household income and wealth in Visegrad countries *Regional Statistics* 8(1): 141–167. https://doi.org/10.15196/RS080102

KOVÁCS, A.–SIKOS T., T. (2015): Effects of Sunday closure on Hungarian shopping behaviour EUGEO Budapest *Congress programme and abstracts. Hungarian Geographical Society* Budapest.

MITRIKOVA, J.–SENKOVÁ, A.–ANTOLIKOVA, S.–PAROVA, V. (2016): The analysis of shopping behaviour in the context of spending leisure time activities of consumers in the chosen shopping centres in Vienna (Austria) *Economic Annals-XXI* 161(9–10): 71–74. https://doi.org/10.21003/ea.V161-16

SIKOS T., T. (2004): Siker vagy kudarc? A barnaövezet hatása a budapesti bevásárlóközpontok fejlődésére In: BARTA, GY. (szerk.): *A budapesti barnaövezet megújulási esélyei* pp. 181–191., MTA Társadalomkutató Központ, Budapest.

SIKOS T., T. (2013): Wien – die Transformation einer traditionsbewussten Stadt und das Erscheinen der Einkaufszentren *Klagenfurter Geographische Schriften* 29(1): 15–31. Paper 6.

SIKOS T., T.–KOVÁCS, A. (2017): A vásárnapi zárvatartás megszűnése Magyarországon – vásárlói vélemények és várakozások *Földrajzi Közlemények* 141(2): 139–151.

SIKOS T., T.–HOFFMANN, I.-NÉ (2004a): *A fogyasztás új katedrálisai.* Magyarország az ezredfordulón: stratégiai tanulmányok a Magyar Tudományos Akadémián, MTA Társadalomkutató Központ, Budapest.

SIKOS T., T.–HOFFMANN, I.-NÉ (2004b): Budapesti bevásárlóközpontok tipológiája *Földrajzi Értesítő* 53(1–2): 111–127.

SIKOS T., T. (2009): A marketingföldrajz kialakulásáról és feladatairól. *Magyar Tudomány* (6): 642– 650.

SIKOS T., T. (2012): Budapesti bevásárlóközpontok *Területi Statisztika* 52(6): 583–591.

SIKOS T., T.–HOFFMANN, I.-NÉ (2012): *A kiskereskedelem új kihívói: Bevásárlóközpontok Budapesttől Prágáig* Akadémiai Kiadó, Budapest.

SIKOS T., T. (2000): *Marketingföldrajz* VÁTI, Budapest.

TÓTH, G.–CSOMÓS, GY. (2016): Mapping the position of cities in corporate research and development through a gravity model-based bidimensional regression analysis *Regional Statistics* 6(1): 217–220. https://doi.org/10.15196/RS06111

INTERNET REFERENCES

http://koos.hu/2011/12/13/plazastop-300-m2

http://www.kormanyhivatal.hu/hu/zala/hirek/februartol-hatalyos-a-plazastop-2

Return migration and identity change: A Hungarian case study

Gábor Lados
(corresponding author)
Department of Economic and Social Geography
University of Szeged, Hungary
E-mail: ladosg@geo.u-szeged.hu

Gábor Hegedűs
Department of Economic and Social Geography
University of Szeged, Hungary
E-mail: hegedusg@geo.u-szeged.hu

This study analyses the identity changes experienced by migrants returning to Hungary before, during, and after their stay abroad. Sussman's (2011) cultural identity change model is used as a theoretical framework to which two factors, 'skills' and 'family status' are added. Although identity change plays an important role in the success or failure of return, reintegration, and reemployment of returnees, to date this has not been studied in Hungary in detail. According to the research results, Hungarian returnees can be categorised into the same four groups as distinguished by Sussman (2011). Returnees in the affirmative identity group are found to be mostly low skilled migrants unable to integrate into the host society; as such, reintegration after their return is relatively easy. The members of the subtractive and additive identity groups can better integrate into the host country as they feel less attached to their native country than the affirmative group, but this causes various reintegration issues. The group of global identity shifters has the most natural reintegration process. Based on the research results, it is recommended that Hungarian return migration policies and initiatives should give greater consideration to these reintegration differences.

Keywords:
emigration,
return migration,
identity change,
remigration initiatives,
Hungary

Introduction

The emigration of manpower from both East-Central Europe and Hungary has exacerbated what is commonly known as the brain drain problem (Glaser–Habers 1974, Nadler et al. 2016, Siska-Szilasi et al. 2016). Return migration is not unprecedented in Europe; there have been many waves of migration and remigration throughout history, but this process reached significant numbers after the EU-accession of East-Central European states and after the more developed member states of the EU opened labour markets (Fassmann et al. 2014). Since the enlargement of the EU in 2004, the free movement of workers has caused a significant number of emigrants (in millions) to leave East-Central Europe, creating complex problems related to the emigration of highly skilled workers such as physicians (Lados–Hegedűs 2016, Boros–Pál 2016, Glorius 2018). Evidence of re-emigration consists of different emigration or immigration flows, interpreted as circular migration (Illés 2018). The mobility of the various actors taking part in voluntary migration flows is increasing because of international socioeconomic transformation, through the increasing embeddedness into transnational networks or the growing number of migrants with multiple citizenships. For example, professionals belonging to the management team of transnational companies, or various migrations for transnational family reunification, also contribute to the mobility of migrants (Conway–Potter 2009, Sussman 2011, Illés–Kincses 2012).

The main objective of this study is to explore the characteristics of identity change of Hungarian emigrants with different work experiences, skills and knowledge before emigration, during and after their return. The study not only touches upon changes in identity that can be observed among returnees as a result of the different migration phases but also how the identity change process and the return are influenced by marital status, the type of work abroad and level of education.

Available relevant international and Hungarian literature was duly taken into account in this study. We used data on the Hungarian migration in the statistical analysis. Emigrants' return and the related identity changes were explored through an interview survey. A total of 48 in-depth interviews were carried out with highly and partially lower skilled individuals who returned to Hungary between 2012 and 2015 after a 1–10 year-long stay abroad. The snowball sampling technique was used for the selection of interviewees. However, given the sample size of our qualitative research, only approximate conclusions can be drawn.

The definition of return migration and its theoretical approaches

Since the 1980s, the scientific research on return has become increasingly significant in migration studies (Glaser–Habers 1974, Gmelch 1980, Cassarino 2004, van Houte–Davids 2008, de Haas 2010, Nadler et al. 2016). To date, the primary concern of Hungarian social geography has mainly focused on emigration in terms of several different migration processes, such as migration of refugees to Europe (Kocsis et al. 2016, Farkas–Dövényi 2018). Studies on return migration, its individual and social aspects, and the practical, economic-development utilisation of this phenomenon (Kovács et al. 2013, Lados–Hegedűs 2016, Csányi 2018) are even rarer. However, similarly to emigration and remigration, labour mobility has also increased in Hungary since 1990, including inland or cross-border commuting on a daily basis or less often (Illés–Kincses 2012, Kovács et al. 2015).

The phenomenon of remigration is diversely defined and interpreted in several different theoretical frameworks (Gmelch 1980, Cassarino 2004, OECD 2008, van Houte–Davids 2008, Sussman 2011, Lados–Hegedűs 2016). The concept of returnees can also be interpreted in many ways. In our research, re-migrants are members of the economically active population aged 15 years and older who voluntarily returned to their country of origin after spending a certain amount of time as migrants in another country. Some researchers highlight the inaccuracy of placing returnees in a homogenous social group (van Houte–Davids 2008), as in the case of emigrants and their diasporas (Sinatti–Horst 2014, Sik–Szeitl 2016).

The effects of the return migration, the relationship between returnees, and the development of the home country are judged differently in the literature. The related countries' policies are also assessed differently (Cassarino 2004, Rédei 2007, van Houte–Davids 2008, de Haas 2010). Return migration may cause multiple effects on different scales, such as a macro-level geographical scale (nation, region, urban or rural area) and a micro-level scale (individuals, family). We identified some positive effects such as the labour market use of material, intellectual, and relational capital obtained abroad; remittances; contribution to modernisation in the form of technological and other innovations; and confidence building between members of the society (Rédei 2007, OECD 2008, Klein-Hitpaß 2016). Our findings indicate that returnees may also face negative effects: for example, they may find it difficult to reintegrate into their homeland's economy and society, and their material prosperity and 'success stories' may give rise to envy. These factors can trigger circular migration and the emigration of newer groups (Gmelch 1980, OECD 2008, van Houte–Davids 2008).

The generally interdisciplinary theories related to remigration (Cassarino 2004, Smoliner et al. 2013, Kunuroglu et al. 2016) can be grouped in several ways. Most migration theories seek to answer questions such as the emigration decision, its motivations, the development of social networks, the existence or absence of links between the receiving and the home country, and the mobility of capital obtained abroad. In a foreign environment, many different influences affect individuals' decisions, habits, and identity. However, migratory theories focus less on the personality of migrants, their identity, and their identity change (Sinatti–Horst 2014). Nevertheless, identity and identity change can have a significant impact on one's future migration decisions (Berry 1997, van Houte–Davids 2008, Sussman 2011). Individual resource mobilisation and preparedness are inherent factors that influence the success of emigration and remigration (Cassarino 2004). In our empirical research, we found that most of the emigrants and returnees tried to prepare for and plan their migration decisions, albeit with differing levels of awareness.

Analysing the identity change of returnees is not a mainstream theory in return migration research, but there are several cases in the literature (Gullahorn–Gullahorn 1963, Berry 1997, Sussman 2011). Sussman's (2011) cultural identity model underlines the shifts in the personal attributes of migrating individuals from a socio-psychological approach. The model examines the strengths and changes of one's attachment to the home and host countries during the whole migration process (before and during emigration, after return), focusing on migrants' cultural adaptation in the host country while maintaining their home identity. This theoretical construct helps us understand the identity changes of returnees, categorise them, and hypothesise the future migration decisions of returnees, that is, the probability of their settlement in the home country or the possibility of re-emigration. Sussman created the model based on a group of Hong Kong returnees. However, groups of returnees from America and Europe were also evaluated. We note that the homeland's historical and social context plays an important role in returnees' experiences of cultural transition, which must be observed in the model's application. According to our research, there is generally no significant negative prejudice against returnees in Hungary.

According to Sussman (2011), the adaptation to the host cultural values takes place in different ways while abroad. The Cultural Identity Model defines four different returnee identity group types (and shifts), which express a relationship to the identity of the country the returnees were originally from. It is difficult to distinguish between these group types, and the spatial comparisons based on the model are difficult to achieve with only a few identity change studies performed globally to date. Both macro- and micro-level factors determine each type: the characteristics of the sending and receiving countries, the extent to which emigrants should be attached to the home country's values, the generational differences among returnees, and the family roles. All these factors can have considerable influence on which group returnees from the same period and country may belong.

The *affirmative identity shifters* are firmly attached to their home culture identity while abroad, as they feel much better in their native country and are not so adaptive towards the host country's social values and culture. However, members of other two model groups may experience significant stress upon return. The return of the *additive* and *subtractive groups* occurs differently. Subtractive returnees are not intimately tied to the culture of either the home or the host country, and they try to learn as many new things as possible during their work abroad. However, due to stress after return, they often become alienated from their host and home countries. In several respects, the additive returnees resemble the subtractive group; the main difference between the two groups is the relationship to their home country cultures. The additive returnees retain and practice their own culture and traditions, but they are much more open to new influences experienced in the host country and usually adopt certain elements (e.g. attitude, work ethics, lifestyle). After returning, despite the stress involved, their connections with the host country will be retained. If changing their cultural identity after returning is difficult, then the re-emigration of the additive group is also possible (Sussman 2011).

The author refers to the fourth, less common group as *global identity returnees*. Such individuals can carry multiple identities. They can follow and adapt individual cultural samples according to their current working and living conditions. However, this is neither facilitated by the mixing of cultural values from the home and the host country, nor by the development of a bicultural strategy. Members of this group consider themselves world citizens, so they can flexibly and quickly adapt to social requirements anywhere abroad. They are moderately positive about their return and ready to move abroad in the future for a shorter or longer period.

In our study, we held interviews with 48 returning Hungarian migrants. We also analysed the type of foreign employment held abroad, and the returning individuals were classified as lower and highly skilled. In our grouping, we assumed, partly because of their qualifications, migrants in lower-level jobs abroad are usually at a lower level of language skills than those working in higher positions.

The European and Hungarian characteristics of return migration

The following findings relate to the territorial characteristics of the migration. In Europe, the brain drain process is a migration from the East to the West. Unfortunately, there are no reliable statistical data on the level of emigration and return migration, as with other migration processes (Blaskó–Gödri 2016, Horváth 2016, Bálint et al. 2017), partly because those who actually reside abroad also include those who have left the country for less than a year and will therefore not appear in the statistical data. In many cases, the European Union Labour Force Survey data are used to interpret return migration. However, this survey examines the labour market situation of the population, and while it comprises data on previous employment abroad (more precisely a year ago), its applicability is very limited in the study of return migration. To date, the characteristics of returnees have not been studied in great detail. In the case of people returning to their post-socialist states, work experience in Western Europe, new skills acquired, and technical knowledge learnt abroad are appreciated in the homeland's labour market (Smoliner et al. 2013). Opinions differ regarding the financial profitability of returning migrants upon returning to their home country. Revenues of migrants returning to the homeland are usually higher compared to their revenues before emigration (Lang–Nadler 2014).

Following the enlargement of the European Union that began in 2004, millions of East-Central Europeans left their homeland given the possibility of the free movement of labour (see Table 1). In 2017, more than 8.5 million people from East-Central European region countries lived in other EU countries. Most of the EU regions affected by emigration are in the former socialist countries (Eurostat 2019). In absolute numbers, more people have migrated from Romania, Poland, and Bulgaria, and fewer people have migrated from relatively advanced countries such as the Czech Republic, Estonia, and Slovenia. In terms of the number of emigrants, Hungary ranks fifth among post-socialist countries. In addition, the proportion of emigrants in the entire home country's population is also an important indicator. The number of emigrants per ten thousand home country citizens is the highest in Romania, Lithuania, and Latvia, while it is the lowest in Hungary, Slovenia, and the Czech Republic. Irrespective of its extent, emigration can cause serious problems in many countries, such as shortage of skilled workers.

Table 1

The number of East-Central European emigrants living in the European Union*

Country/Region	2004–2007	2008–2011	2012–2015	2016–2018
Romania	389.2	859.8	1,198.7	1,622.9
Lithuania	172.4	499.7	819.0	1,179.1
Croatia	765.5	564.5	758.8	1,077.8
Bulgaria	231.4	375.5	596.2	1,068.6
Latvia	90.8	194.7	652.7	941.6
Poland	158.1	313.2	472.4	638.9
Estonia	189.5	280.9	436.9	628.3
Slovakia	180.3	251.7	407.2	621.5
Hungary	82.7	104.8	256.5	433.0
Slovenia	161.6	126.6	197.1	298.7
Czech Republic	55.2	60.6	81.0	149.5
East-Central Europe	218.1	385.3	576.5	817.0

* The number of citizens living abroad per 10 thousand inhabitants.
Source: Own calculation based on the Eurostat (2019) database.

The proportion of people returning to East-Central Europe has increased significantly since 2004 (see Table 2). During the 2008 global economic crisis, especially in 2008–2009, a mass return migration occurred in most post-socialist countries, and since 2012, the number of returnees has begun to increase along with the increase in emigration. We found the highest rates are among Romanian and Lithuanian returnees.

Table 2

The share and number of returnees to East-Central Europe

Country/Region	Proportion of returnees, %[a]			Number of returnees per 10 thousand inhabitants		
	2004–2007	2008–2011	2012–2017	2004–2007	2008–2011	2012–2017
Romania	no data	92.1	88.4	no data	64.6	67.0
Lithuania	68.8	79.9	76.6	14.9	23.3	56.1
Estonia	39.1	50.0	51.8	6.1	13.2	38.3
Poland	87.0	90.8	55.4	2.5	30.8	31.5
Latvia	no data	72.0	58.5	no data	34.5	29.2
Hungary	8.1	9.7	50.0	2.1	2.9	26.0
Croatia	92.8	60.7	50.9	34.4	16.7	14.3
Slovenia	9.1	12.8	17.6	8.5	14.2	13.3
Bulgaria	96.0	no data	39.7	2.0	0.0	12.0
Slovakia	no data	18.8	52.0	no data	2.2	6.0
Czech Republic	2.9	28.3	13.1	2.0	17.6	5.0
East-Central Europe	20.3	61.8	59.8	3.1	25.1	32.1

[a] In the given period, the proportion of citizens returning home compared to the average number of immigrants in a year.

Source: Own calculation based on the Eurostat (2019) database.

In 2016, more than 43% of Hungarians in EU countries lived in Germany, 21% in the United Kingdom, and 16% in Austria. The proportion of Hungarian citizens working abroad is not high compared to the other countries in East-Central Europe. Nevertheless, it is important to note that the rate of Hungarian emigration has increased three-and-a-half times since 2001 (Gödri 2015).

Greater human and relational capital provision will further increase the potential for emigration. Hungary is at the forefront of European countries with medium-level migratory tendencies. The three most significant potential destinations are Germany, Austria, and the United Kingdom (Sik–Szeitl 2016). The majority of emigrants are young males, generally aged 20–39 years. Furthermore, the proportion of people with higher education and the proportion of unmarried people is greater than that of the remaining population (Blaskó–Gödri 2016). Similar to emigrants, the aforementioned groups are in the majority among the returnees (Horváth 2016). Migration intention is the highest among young people, particularly the 21–30 age group (Siskáné Szilassi–Halász 2018).

Returnees originally from Central and Eastern European countries are more likely to start their own business than those who remain, but unemployment rates may be higher in their group (Martin–Radu 2012, Horváth 2016). To date, only a few studies have been conducted to analyse the regional-local spatial characteristics of return within the home country (Kincses 2014, Apsite-Berina et al. 2018). According to the 2011 census (Kincses 2014) and the 2016 micro-census data (HCSO 2018), returning Hungarians adopt certain attitudes after their return when choosing their place of residence. They settle down mainly in the agglomeration of Budapest, in the vicinity of Lake Balaton and in more populous cities and towns outside Central Hungary. According to Kincses (2014), only 31% of homebound migrants returned to their previous residences, which were presumably important to them. This phenomenon can be observed in other East-Central European states, including Poland, especially among highly qualified returnees (Klein-Hitpaß 2016).

In parallel with the migratory wave, the number of returning Hungarians has continuously increased. For instance, according to the Eurostat (2019) data, 32,557 people returned in 2015, which accounts for 56% of all immigrants (excluding refugees) in Hungary. This rate has just been approaching the East-Central European average in the recent years. However, considering the statistical errors, we can conclude that the actual figures may be considerably higher.

Several programmes and initiatives tackling brain drain were implemented to minimise the negative effects of emigration and support return migration at various geographical scales (e.g. at the national, regional or local level) (Kovács et al. 2013, Lados–Hegedűs 2016). Currently no comprehensive migration policy dealing with out and return migration in the European Union exists (Kálmán 2016, Nadler et al. 2016). The majority of relevant short- and medium-term national policies and initiatives supporting return migration were developed and implemented in the 2000s. Evaluating the relatively new and delayed East-Central European initiatives is complicated (Kovács et al. 2013, Boros–Hegedűs 2016, Kálmán 2016).

A complex and comprehensive migration strategy has not yet been developed in Hungary; however, some initiatives focusing on specific target groups or areas of migration have been implemented. One of the most well-known and successful initiatives is the 'Lendület' ('Momentum') program founded by the Hungarian Academy of Sciences in 2009, which aims to re-attract young Hungarian researchers who have left the country. Another initiative is an NGO called 'Gyere Haza Alapítvány' ('Come Back Home Foundation'), which was founded in 2010. In 2013, it expanded its services to offer online and personal support to those out-migrants wishing to return. We can identify several retain-type programmes in the health care system or the Act CCIV of 2011 on National Higher Education. In 2015, the so-called 'Gyere Haza' ('Come Back Home') program was launched to re-attract young, higher educated Hungarians living in London and offered re-employment after return; the program ended in 2016 (Lados–Hegedűs 2016).

Characteristics of Hungarian returnees and their identity change

Main characteristics of interviewees

For the interview survey, returnees were divided into two groups according to their foreign work experience: lower and higher skilled returnees (see Table 3). The grouping was based on the hypothesis that emigrants working in jobs beneath their qualifications may return with less useful foreign work experience. Most of them, due to incomplete foreign language knowledge, were not or only partially integrated into the host society, and therefore their cultural identity and emotional attachment to their homeland remained relatively strong (Sussman 2011). Other influential factors included the length of stay abroad and marital status. The interview survey was based on Sussman's (2011) cultural identity change model, which compared foreign work experience and marital status. This method, along with a complex analysis of return migration, projects future migration strategies.

Similar patterns highlighted by international and national research were extracted from the interviews (Martin–Radu 2012, Horváth 2016). The majority of the sample comprised of men aged 20–34 years with a high level of education. The United Kingdom and the United States were overrepresented among the target countries, while Germany and Austria were underrepresented. Diverse migration paths illustrate the differences between returnees. Highly skilled returnees were more likely to have migrated greater geographical distances to achieve their goals and were likely to live in two or more countries.

Table 3

The personal data of highly skilled interviewees, 2012–2015

Inter-view num-ber	Gender	Age[a)]	Highest level of education	ISCO category[b)]	Target country/countries
I1	male	25–29	PhD	ISCO-21	USA
I2	female	25–29	PhD	ISCO-21	USA
I3	male	30–34	PhD	ISCO-21	Germany
I4	male	30–34	PhD	ISCO-21	USA
I5	male	30–34	PhD	ISCO-21	USA, Netherlands
I6	male	25–29	university	ISCO-36	United Kingdom
I7	male	30–34	PhD	ISCO-21	Belgium, United Kingdom
I8	male	25–29	PhD	ISCO-21	United Kingdom
I9	male	30–34	PhD	ISCO-21	Australia
I10	female	20–24	college	ISCO-36	United Kingdom
I11	male	25–29	PhD	ISCO-21	United Kingdom, Italy
I12	male	25–29	PhD	ISCO-21	United Kingdom, Switzerland, Germany
I13	male	30–34	university	ISCO-21	Kazakhstan
I14	male	25–29	university	ISCO-21	Ireland
I15	female	20–24	university	ISCO-24	United Kingdom
I16	male	30–34	PhD	ISCO-21	Denmark, Japan
I17	female	15–19	university	ISCO-25	France
I18	female	30–34	PhD	ISCO-22	USA
I19	male	25–29	PhD	ISCO-21	United Kingdom
I20	male	25–29	PhD	ISCO-21	Czech Republic
I21	male	40–44	PhD	ISCO-21	USA
I22	male	20–24	PhD	ISCO-24	USA
I23	male	25–29	university	ISCO-13	Luxembourg
I24	male	25–29	PhD	ISCO-21	USA
I25	male	25–29	university	ISCO-26	Luxembourg
I26	male	25–29	high-school graduation	ISCO-21	Belgium, United Kingdom, Germany, Spain
I27	female	25–29	university	ISCO-36	Luxembourg
I28	male	20–24	college	ISCO-24	United Kingdom

(Continued on the next page.)

a) At the time of emigration.

b) Foreign work experience by International Standard Classification of Occupations' (ISCO) categories: ISCO-1 = Legislators, senior officials, and managers; ISCO-2 = Professionals; ISCO-3 = Technicians and associate professionals; ISCO-4 = Clerks; ISCO-5 = Service workers and shop and market sales workers; ISCO-6 = Skilled agricultural and fishery workers; ISCO-7 = Craft and related trades workers; ISCO-8 = Plant and machine operators and assemblers; ISCO-9 = Elementary occupations. For the detailed classification structure, see https://www.ilo.org/public/english/bureau/stat/isco/

(Continued)

Interview number	Gender	Age[a)]	Highest level of education	ISCO category[b)]	Target country/countries
I29	male	20–24	professional training	ISCO-51	United Kingdom
I30	male	25–29	technical school	ISCO-71	Germany, Ireland
I31	male	25–29	technical school	ISCO-71	Kazakhstan
I32	female	20–24	college	ISCO-41	United Kingdom
I33	male	40–44	university	ISCO-93	Germany
I34	male	50–54	college	ISCO-93	Germany
I35	male	20–24	college	ISCO-93	Ireland
I36	male	20–24	technical school	ISCO-93	USA, United Kingdom, Sweden
I37	female	25–29	college	ISCO-42	United Kingdom, Ireland
I38	male	30–34	technical school	ISCO-93	Netherlands, Germany
I39	female	20–24	university	ISCO-51	Spain, United Kingdom
I40	female	25–29	professional training	ISCO-51	United Kingdom
I41	female	25–29	college	ISCO-51	United Kingdom
I42	female	45–49	technical school	ISCO-51	USA
I43	male	25–29	technical school	ISCO-51	United Kingdom
I44	male	30–34	university	ISCO-51	Estonia
I45	male	40–44	technical school	ISCO-75	Germany
I46	male	25–29	university	ISCO-51	United Kingdom
I47	male	25–29	university	ISCO-51	United Kingdom
I48	male	30–34	technical school	ISCO-93	Austria

a) At the time of emigration.

b) Foreign work experience by International Standard Classification of Occupations' (ISCO) categories: ISCO-1 = Legislators, senior officials, and managers; ISCO-2 = Professionals; ISCO-3 = Technicians and associate professionals; ISCO-4 = Clerks; ISCO-5 = Service workers and shop and market sales workers; ISCO-6 = Skilled agricultural and fishery workers; ISCO-7 = Craft and related trades workers; ISCO-8 = Plant and machine operators and assemblers; ISCO-9 = Elementary occupations. For the detailed classification structure, see https://www.ilo.org/public/english/bureau/stat/isco/

In our opinion, the possible benefits of foreign work experience, that is, whether returnees are able to profit from emigration after return, depend on the type of foreign work. On one hand, those who worked in hospitality, construction, and food industry or had other manual jobs abroad (ISCO 4–9) were lower skilled migrants. On the other hand, those who worked as researchers, IT professionals, engineers, or doctors (ISCO 1–3) were identified as highly skilled migrants.

The sample's gender distribution indicates that the majority are male, which correlates with the literature (Horváth 2016). Those who emigrated with their family or as singles had different migration paths.

Lower skilled migrants returned with less useful work experience. Foreign language knowledge was typically not a primary requirement for those who worked in trained jobs, which means they worked with other Hungarians or immigrants from other countries who did not speak a foreign language.

Lower skilled migrants had more problems integrating in the host country than their highly skilled counterparts due to lack of language skills and unfamiliar working conditions. Their social relationships were superficial and fragile, and they broke after return. The main motivation for their emigration was a higher salary and accumulation of financial reserves abroad, as shown in previous studies (Martin–Radu 2012, Lang–Nadler 2014). They mainly held trained jobs abroad and were overqualified for their foreign jobs: *'Together with my partner, we planned to earn enough money for a house during our time abroad. It seemed completely absurd that we could have achieved this if we had stayed home (in Hungary)'* (Mária[1], female, unskilled worker). Another interviewee said the emigration plan was *'very easy. I could barely afford to pay for my home. It was much easier and more convenient to support them from abroad (…) though, we had a hard time living apart. Fortunately, it only worked for a while, after a while, my family moved'* (Zoltán, male, butcher).

Those who had a higher level of education were categorised as a special subgroup; they did not work in their profession abroad and were overqualified for their foreign jobs. This may be regarded as brain waste (Személyi–Csanádi 2011). They could not return with newly acquired professional skills, but improved their language skills, which seemed to be beneficial after return.

Preparing for return was more common for highly skilled returnees, but not typical for all of them. The majority regarded emigration as temporary. Their main motivation was to improve their standard of living, financially (Horváth 2016) or professionally (e.g. career development). According to a young researcher who worked abroad for five years: *'In my profession, especially in the institute where I worked before, it was an unspoken expectation that a young researcher would go abroad to learn new methods for some years. It is thought to be a good way to start your career, and after the return, it might be also beneficial for your institution'* (Zsolt, male, researcher). Highly skilled returnees worked in multicultural environments where they had colleagues from all over the world. Job advancement in their foreign workplace was achievable, depending on their knowledge and previous work experience. However, despite prosperous career

[1] The names of interviewees have been changed.

paths, several interviewees stated they could only reach a certain position as immigrants; the host country's citizens occupied senior positions.

On the other hand, lower skilled migrants did not prepare as intentionally for emigration as highly skilled migrants. They usually made a quick decision to move abroad, even though many of them did not speak the language of the host country.

Our study showed that family ties were one of the most important motivations for Hungarians to return, in the same way as in previous studies (Martin–Radu 2012, Lang–Nadler 2014). According to the interviewees, family could be regarded as positive not just as negative motivation (e.g. return because of the death of a close relative). The fear of their children's assimilation led some interviewees to return home: *'I loved my foreign, well-paid job, but I always planned that my children would grow up in Hungary as Hungarians'* (Gergő, male, translator). Another returnee had a similar response: *'We always spoke in Hungarian to each other when we were at home. But later, my children started to answer in English and refused to switch to Hungarian. Then I realised we spent enough time abroad (…) The environment was very inspiring to my children, but I did not want to see them losing their Hungarian identity'* (Péter, male, researcher). This statement was reflected in other cases of families moving abroad.

Family did not influence emigration or return for single people because they stayed in contact with their relatives: *'Interestingly, there were friends and other relatives whom I spoke with more frequently, via Skype, than before my emigration (…) Of course, after some time, it was not regular'* (János, male, waiter).

Identity change of returnees

Among the four groups of Sussman's (2011) identity change model, affirmative identity shifters formed the smallest group, similar to other international results. Those who had the worst working conditions abroad, the so-called 3D jobs (dirty, dangerous, difficult) (Gregory et al. 2009), were lower skilled migrants who left their families at home: *'Hungarians did not have a good reputation at my foreign workplace. We received the hardest tasks. Our superiors treated us in this way'* (Géza, male, construction worker). They experienced return as a real relief and did not mention significant obstacles during re-integration. Their return may be regarded as a permanent migration decision.

Subtractive identity shifters were migrants who had lower skilled jobs abroad and emigrated with their spouse, children, or other family members. The expectation of higher wages and potential savings motivated them to leave the country. Their limited language skills did not particularly improve during the emigration. In the sample, subtractive and additive migrants were one of the most common identity types, somewhat different from Sussman's (2011) results which stated that subtractive migrants are the most common.

The meagre language skills of subtractive migrants determined their foreign working conditions. On one hand, they worked with other immigrants with basic foreign language skills, so they could not properly learn the host country's language. On the other hand, there was a native Hungarian atmosphere (mainly in trained jobs). Their families were generally closer to each other, which hindered integration in the host country. There were several obstacles following their return, such as re-employment in the Hungarian labour market. They could not utilise their foreign work experience after their return. Major barriers were migration plans made before the return that subtractive migrants were unable to fulfil, such as starting a family and having a stress-free, calm life without financial problems.

For subtractive and additive migrants, family was the most important motivation for return. Hence, the decision to move back was against their intentions and caused negative feelings about returning. Some who returned for family reasons mentioned clear migration plans for the future: *'Our children wanted to come back. We did not want to be separated from each other, so we followed them. When they finish secondary school, we [the parents] are sure to go abroad again because I cannot find any suitable job here'* (Zoltán, male butcher).

According to these negative factors, lower skilled returnees are more likely to leave the country again, except for affirmative identity shifters; this suggests that subtractive, and to some extent, additive interview partners have found their reintegration as difficult as other European and non-European returnees (Sussman 2011). Hence, they could be regarded as potential circular migrants (Illés–Kincses 2012, Martin–Radu 2012).

Additive identity shifters fell into two sub-groups: lower skilled single people and highly skilled married people. The former sub-group included those with university degree who worked in underqualified jobs abroad. They tried to acquire as many new skills as they could during emigration in order to be more successful in the labour market after return. They mentioned their improved language skills (usually English or German) and multinational work experience. The latter sub-group returned with more skills and knowledge that could be beneficial in the labour market, but like with the subtractive returnees, they returned for external reasons (e.g. family): *'I would have gladly stayed abroad, but my child did not feel comfortable there (…) I chose my family over career (…) But on my own, I would not have returned. Nevertheless, as for my current workplace, I do not have anything to complain about'* (Tamás, male, researcher).

According to the Sussman's (2011) identity change model, the most successful returnees are the global identity shifters, who changed the most during the migration process. Similarly, our study found that returnees with a global identity were the minority. They are mainly highly skilled single migrants and could improve both their professional and personal life abroad. During their emigration, they were able to acquire a new identity. One of them commented, *'First, I would position myself as a European, second, a Budapestan, and third, a Hungarian'* (Szabolcs, male, translator). Highly skilled returnees experienced smoother re-employment in the Hungarian labour market than their lower skilled counterparts did due to their newly acquired skills (e.g. management skills, technological know-how) and because they did not cut ties with former employers. Furthermore, as a result of emigration, almost all of the interviewees with global identity were able to achieve career development. Their decision to return was not affected by family reasons as in the case of the members of the subtractive or additive groups.

Conclusion

In this paper we analysed the identity change of Hungarian returnees based on Sussman's identity change model. Statistical data on emigration and return migration are very limited, but according to these databases, the importance of both processes in Hungary is increasing. Our interview survey showed that the identity change model and classification could be adapted to Hungarian returnees. The original model did not include family status and foreign work experience in the evaluation of return migration, but we added these factors to our model. According to the results, highly skilled returnees were able to utilise their foreign work experience, and their broadened social and professional networks may be beneficial in the labour market. They faced fewer obstacles during the return process and self-assessed a more successful return than lower skilled returnees, who returned with less useful work experience and who could not develop skills useful for the labour market. They face more difficulties during reintegration in the home country and regard their return as less successful. Family status could affect the return of both lower and highly skilled migrants. Having a child or the fear of children's assimilation may motivate people to return. We can conclude that identity change should be considered during decision making and implementation of return initiatives. Moreover, foreign work experience, newly acquired skills and knowledge, and various family-related aspects (e.g. adequate social services) should be considered and adopted in the context of initiatives supporting return migration.

REFERENCES

ABBA APSITE-BERINA, E.–KRISJANE, Z.–SECHI, G.–BERZINS, M. (2018): Regional patterns of belonging among young Latvian returnees *Economic Science for Rural Development Conference Proceedings* 48: 77–84. https://doi.org/10.22616/esrd.2018.071

BÁLINT, L.–CSÁNYI, Z.–FARKAS, M.–HLUCHÁNY, H.–KINCSES, Á. (2017): International migration and official migration statistics in Hungary *Regional Statistics* 7(2): 101–123. https://doi.org/10.15196/rs070203

BERRY, J. W. (1997): Immigration, acculturation, and adaptation *Applied psychology* 46(1): 5–34. https://doi.org/10.1080/026999497378467

BLASKÓ, ZS.–GÖDRI, I. (2016): The social and demographic composition of emigrants from Hungary. In: BLASKÓ, ZS.–FAZEKAS, K. 2016: *The Hungarian Labour Market 2016* pp. 60–68., Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budapest.

BOROS, L.–HEGEDŰS, G. (2016): European national policies aimed at stimulating return migration. In: NADLER, R.–KOVÁCS, Z.–GLORIUS, B.–LANG, T. (eds.): *Return migration and regional development in Europe: mobility against the stream* pp. 333–357., Palgrave Macmillan, Basingstoke.

BOROS, L.–PÁL, V. (2016): A magyarországi orvosmigráció néhány jellemzője *Észak-Magyarországi Stratégiai Füzetek* 13(1): 64–72.

CASSARINO, J-P. (2004): Theorising return migration: The conceptual approach to return migrants revisited *International Journal on Multicultural Societies* 6(2): 253–279.

CSÁNYI, Z. (2018): Exploring the behavioural aspects of migration decisions in the biography of a returner *Análise Europeia* 3(5): 58–90.

CONWAY, D.–POTTER, R. B. (2009): Return of the 'Next Generations': Transnational mobilities, family demographics and experiences, multi-local spaces. In: CONWAY, D.–POTTER, R. B. (eds.): *Return Migration of the Next Generations: 21st Century Transnational Mobility* pp. 236–237., Ashgate, Farnham.

DE HAAS, H. (2010): Migration and development: a theoretical perspective *International migration review* 44(1): 227–264. https://doi.org/10.1111/j.1747-7379.2009.00804.x

EUROSTAT (2019): http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_imm2ctz&lang=en (accessed 22.05.2019.)

FARKAS, M.–DÖVÉNYI, Z. (2018): Migration to Europe and its demographic background *Regional Statistics* 8(1): 29–48. https://doi.org/10.15196/RS080103

FASSMANN, H.–MUSIL, E.–BAUER, R.–MELEGH, A.–GRUBER, K. (2014): Longer-term demographic dynamics in South-East Europe: Convergent, divergent and delayed development paths *Central and Eastern European Migration Review* 3(2): 150–172.

GLASER, W. A.–HABERS, G. C. (1974): The migration and return of professionals *International Migration Review* 8(2): 227–244. https://doi.org/10.2307/3002782

GLORIUS, B. (2018): Migration to Germany: Structures, processes, and discourses *Regional Statistics* 8(1): 3–28. https://doi.org/10.15196/RS080101

GMELCH, G. (1980): Return migration *American Review of Anthropology* 9(1): 135–159. https://doi.org/10.1146/annurev.an.09.100180.001031

GÖDRI, I. (2015): Nemzetközi vándorlás. In: MONOSTORI, J.–ŐRI, P.–SPÉDER, ZS.: *Demográfiai portré 2015* pp. 187–211., KSH Népességtudományi Kutatóintézet, Budapest.

GREGORY, D.–JOHNSTON, R.–PRATT, G.–WATTS, M. J.–WHATMORE, S. (eds.) (2009): *The dictionary of human geography* Wiley-Blackwell, Chichester, 462–464.

GULLAHORN, J. T.–GULLAHORN, J. E. (1963): An extension of the U-Curve hypothesis *Journal of Social Issues*, 19(3): 33–47. https://doi.org/10.1111/j.1540-4560.1963.tb00447.x

HCSO (2018): *Mikrocenzus 2016. Nemzetközi vándorlás.* Hungarian Central Statistical Office, Budapest.

HORVÁTH, Á. (2016): Returning migrants. In: BLASKÓ, ZS.–FAZEKAS, K. 2016: *The Hungarian Labour Market 2016* pp. 110–116., Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budapest.

ILLÉS, S. (2018): Gazdasági válság és cirkuláció *Területi Statisztika* 58(1): 103–122. https://doi.org/10.15196/TS580105

ILLÉS, S.–KINCSES, Á. (2012): Hungary as receiving country for circulars *Hungarian Geographical Bulletin* 61(3): 197–218.

KÁLMÁN, J. (2016): Public policies encouraging return migration in Europe. In: BLASKÓ, ZS.–FAZEKAS, K. 2016: *The Hungarian Labour Market 2016* pp. 117–121., Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budapest.

KINCSES, Á. (2014): Nemzetközi migrációs körkép Magyarországról a 2011-es népszámlálási adatok alapján *Területi Statisztika* 54(6): 590–605.

KLEIN-HITPAß, K. (2016): Return migrants as knowledge brokers and institutional innovators: New theoretical conceptualisations and the example of Poland. In: NADLER, R.–KOVÁCS, Z.–GLORIUS, B.–LANG, T. (eds.) (2016): *Return migration and regional development in Europe: mobility against the stream* pp. 333–357., Palgrave Macmillan, Basingstoke.

KOCSIS, K.–MOLNÁR SANSUM, J.–KREININ, L.–MICHALKÓ, G.–BOTTLIK, ZS.–SZABÓ, B.–BALIZS, D.–VARGA, GY. (2016): Geographical characteristics of contemporary international migration in and into Europe *Hungarian Geographical Bulletin* 65(4): 369–390. https://doi.org/10.15201/hungeobull.65.4.6

KOVÁCS, Z.–BOROS, L.–HEGEDŰS, G.–LADOS, G. (2013): Returning people to the homeland: Tools and methods supporting remigrants in a European context In: LANG, T. (ed.): *Return Migration in Central Europe: Current trends and an analysis of policies supporting returning migrants* pp. 58–94., Forum ifl. 21. Leibniz-Institut für Länderkunde, Leipzig.

KOVÁCS, Z.–EGEDY, T.–SZABÓ, B. (2015): Az ingázás területi jellemzőinek változása Magyarországon a rendszerváltozás után *Területi Statisztika* 55(3): 233–253.

KUNUROGLU, F.–VAN DE VIJVER, F.–YAGMUR, K. (2016): Return migration *Online Readings in Psychology and Culture* 8(2): 1. https://doi.org/10.9707/2307-0919.1143

LADOS, G.–HEGEDŰS, G. (2016): An evaluation of Hungarian return migration *Hungarian Geographical Bulletin* 65(4): 321–330. https://doi.org/10.15201/hungeobull.65.4.2

LANG, T.–NADLER, R. (eds.) (2014): *Return migration to Central and Eastern Europe – transnational migrants' perspectives and local businesses' needs Forum ifl.23*. Leibniz-Institut für Länderkunde, Leipzig.

MARTIN, R.–RADU, D. (2012): Return migration: The experience of Eastern Europe *International Migration* 50(6): 109–128. https://doi.org/10.1111/j.1468-2435.2012.00762.x

NADLER, R.–KOVÁCS, Z.–GLORIUS, B.–LANG, T. (eds.) (2016): *Return migration and regional development in Europe: mobility against the stream* Palgrave Macmillan, Basingstoke.

OECD (2008): *International migration outlook* 2008, OECD Publishing, Paris.

RÉDEI, M. (2007): Hazautalások Kelet- és Közép-Európába *Statisztikai Szemle* 85(7): 581–601.

SIK, E.–SZEITL, B. (2016): Migráció a mai Magyarországról *Educatio* 25(4): 546–557.

SINATTI, G.–HORST, C. (2014): Migrants as agents of development *Ethnicities* 1–19. https://doi.org/10.1177%2F1468796814530120

SISKA-SZILASI, B.–KÓRÓDI, T.–VADNAI, P. (2016): Measuring and interpreting emigration intentions of Hungarians Hungarian *Hungarian Geographical Bulletin* 65(4): 361–368. https://doi.org/10.15201/hungeobull.65.4.5

SISKÁNÉ SZILASI, B.–HALÁSZ, L. (2018): Reasons and characteristics of the dynamizing emigration intention of the Hungarian youth *Eastern European Business and Economics Journal* 4(1): 79–96.

SMOLINER, S.–FÖRSCHNER, M.–HOCHGERNER, J.–NOVÁ, J. (2013): Comparative report on re-migration trends in Central and Eastern Europe. In: LANG, T. (ed.): *Return Migration in Central Europe: Current trends and an analysis of policies supporting returning migrants Forum ifl. 21*. pp. 11–57., Leibniz-Institut für Länderkunde, Leipzig.

SUSSMAN, N. M. (2011): *Return migration and identity. A global phenomenon, a Hong Kong case* Hong Kong University Press, Hong Kong.

SZEMÉLYI, L.–CSANÁDI, M. (2011): Some sociological aspects of skilled migration from Hungary *Acta Universitatis Sapientiae, Social Analysis* 1(1): 27–46.

VAN HOUTE, M.–DAVIDS, T. (2008): Development and return migration: from policy panacea to migrant perspective sustainability *Third World Quarterly* 29(7): 1411–1429. https://doi.org/10.1080/01436590802386658

High demand for local area level statistics – How do National Statistical Institutes respond?

Marie Haldorson
Statistics Sweden,
Sweden
E-mail: marie.haldorson@scb.se

Linking people to place is becoming increasingly important, not least because of the 2030 Agenda for sustainable development. Enhancing traditional statistics at the regional level with those at the local area level enriches the way National Statistical Institutes provide information to society. To provide such statistics at local level in a timely manner, the production of statistics needs to be modernised and based on various data sources, such as administrative registers and geospatial information.
The United Nations has recognised the importance of statistical-geospatial integration by creating an Expert Group under the UN Global Geospatial Information Management initiative. The European Statistical System has also paved the way for such an integration by funding the GEOSTAT projects.
For National Statistical Institutes looking to improve their capability to produce statistics at the local level, the Global Statistical Geospatial Framework is a starting point that describes the requirements that need to be fulfilled. For guidance on how to start with a more bottom-up approach, the GEOSTAT 2 project has developed key actions aligned with the Generic Statistical Business Process Model. Statistics Sweden's capability assessment might also provide some inspiration.

Keywords:
statistical-geospatial integration,
statistics at local area level,
UN 2030 Agenda for sustainable development,
Global Statistical Geospatial Framework (GSGF)

Introduction

National Statistical Institutes (NSIs) in Europe and across the globe are providing policymakers regional statistics that capture what is happening in the society, economy, and environment. However, how can this be done in a way that enables new policies to be effective and detailed enough to meet targets and improve the situation wherein they are most needed? Territorial policies at national and European levels, together with the UN 2030 Agenda for sustainable development, have put pressure on European NSIs to produce relevant and timely statistics (Haldorson et al. 2016). The 2030 Agenda aims to leave no one behind; from a geographical perspective, this means, for example, taking the urban–rural dimension into account, as well as issues of accessibility.

It is more important than ever to relate official statistics to the realities people are facing, for example, in deprived neighbourhoods. Furthermore, disaster risk management or measures to improve accessibility to public transport, schools, health clinics, or green areas require local-level statistics. The European Statistical System[1] has recognised this necessity by funding the GEOSTAT projects[2], such as GEOSTAT 2, thereby providing guidance to NSIs on the fundamentals of a more flexible statistical infrastructure that is point-based (Moström et al. 2017).

Regional statistics are regularly produced by all the countries in Europe, typically using the Nomenclature of Territorial Units for Statistics, NUTS, classification[3] of administrative units in all Member States. The Eurostat database provides harmonised regional statistics, but for users looking for more geographic detail, the situation is far from harmonised. An initial step to improve this has been the introduction of statistical grids, degree of urbanisation, and so on, together with classifying Local Administrative Units as the new geographic outputs as per the revised EU regulation on territorial typologies, Tercet.[4]

Statistical grids are set to become one of the dissemination areas in the next European round of population and housing censuses in 2021.[5] Production of grid-based statistics calls for the modernisation of the way countries conduct their censuses and leads the way to a register-based statistical system.

Linking people and businesses to place

Benefits of statistical geospatial integration

Statistics Sweden has used geospatial data and integrated it with statistical data sources, mainly administrative registers, for a long time. It started with census maps, continued with statistics on grids in the 1980s, and has, since then, moved on to using multiple geospatial data sources to produce land accounts; it has a dynamic web tool called the Statistical Atlas for performing advanced geospatial analysis.

[1] Eurostat: European Statistical System http://ec.europa.eu/eurostat/web/european-statistical-system/overview?locale=fr

[2] European Forum for Geography and Statistics GEOSTAT Project Descriptions http://www.efgs.info/geostat/

[3] NUTS Regulation Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003, on the establishment of a common classification of territorial units for statistics (NUTS), OJ L 154, 21.6.2003

[4] Tercet, Territorial Typologies of the European Union: http://ec.europa.eu/eurostat/web/nuts/tercet-territorial-typologies

[5] European Census: https://ec.europa.eu/eurostat/web/population-and-housing-census/overview

However, the Statistical Atlas is no longer supported; this means that the website lacks a visualisation tool to convert the statistics to maps. Another problem is that statistics at the local level are mostly sold as commissioned services. Statistics Sweden decided only recently to start publishing such statistics as official statistics accessible to everyone, something that other NSIs have been doing for a long time.

The benefits of statistical geospatial integration are starting to be widely recognised and the UN Committee on Global Geospatial Information Management has created an Expert Group on Integration of Statistical and Geospatial Information (EG–ISGI).[6] Statistics Sweden is taking an active part in the work of the expert group and is currently coordinating the ESSnet project GEOSTAT 3, which focuses on developing a European implementation of the Global Statistical Geospatial Framework (GSGF).[7] An ESSnet project is a network of several ESS organisations aimed at providing results that will be beneficial to the whole European Statistical System, ESS.

The GSGF was adopted in 2016 and directly supports the integration of statistical and geospatial information (UN EG-ISGI 2016). It builds on a national framework developed by the Australian Bureau of Statistics; this has been refined and globalised by the EG–ISGI. Statistics Sweden has started implementing the GSGF nationally and it has proved to be a useful tool for planning and identifying improvements in this aspect of statistics, as described later in this paper.

The GSGF is based on five principles:

1. Use of fundamental geospatial infrastructure and geocoding;
2. Geocoded unit record data in a data management environment;
3. Common geographies for dissemination of statistics;
4. Statistical and geospatial interoperability – data, standards and processes; and
5. Accessible and usable geospatially enabled statistics.

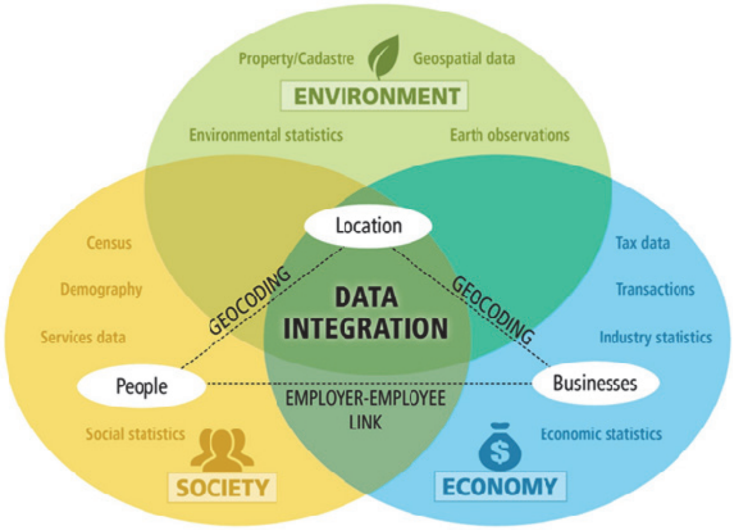
The GSGF helps adding location to statistical data that describes individuals and businesses. It also helps organising geospatial data management and the processes needed to get high-quality statistical output.

[6] UN-GGIM Expert Group on Integration of Statistical and Geospatial Information, http://ggim.un.org/UNGGIM-expert-group/

[7] The GEOSTAT 3 Project, http://www.efgs.info/geostat/geostat-3/

Figure 1

Linking society, economy and environment through data integration



Source: Statistics Australia and Statistics New Zealand.

Furthermore, the GSGF provides a bridge between the official statistical agency and the geospatial profession. Statistical units are geocoded using geospatial identifiers, such as cadastral parcel, address, or building. Data are then aggregated and released for the geographic boundaries in the boundary (or geography) layer. This supports integration of data across diverse sources and provides the bridge between statistical and more traditional geospatial datasets; through use of a common geography and greater flexibility in geographic outputs.

The five principles of the Global Statistical Geospatial Framework

The GSGF is a high-level framework, meaning that it is not intended to provide a detailed implementation plan or design; rather, it provides guidance on *what* should be available in countries, providing a lot of flexibility on *how* it is done.

To help countries implement the framework, supporting documents were developed in 2018, both at global level (by the EG–ISGI) and at European level (through the GEOSTAT 3 project). The draft descriptions from GEOSTAT 3 give a good overview of how to understand the principles from a European perspective. As this is still a work in progress, the latest available information can be found on the websites of UN–GGIM and GEOSTAT 3.

Principle 1: Use of fundamental geospatial infrastructure and geocoding

The goal of this principle is to obtain a high quality and standardised location attribute or reference. The aim is to agree on the protocol that regulates this attribution of location information to statistics. The framework recommends using existing geospatial information provided by specialised geospatial authorities (e.g. physical address, cadastral parcel, building, or some other location description) in order to assign accurate coordinates and/or a small geographic area or standard grid reference to each statistical unit (i.e. at the micro-data or unit record level).

Time- and date-stamping these locations will place the unit in both time and space. The process for obtaining locations and geocodes should use official and fundamental geospatial data from the National Spatial Data Infrastructures or other nationally agreed sources.

Principle 2: Geocoded unit record data in a data management environment

The GSGF recommends that the linking of a geocode to each statistical unit record in a data set (i.e. a person, household, business, building, or parcel/unit of land) occur within a data management environment. Persistent and version-controlled storage of high precision geocodes enables any geographic context to be applied when preparing the data to be released in the future (i.e. in aggregating data into a variety of larger geographic units or to adapt to changes in geographies over time). Moreover, geocodes can enable data linking processes that aim to integrate information of different types and from various sources by the so-called linked data techniques. An efficient data management environment should allow linking statistical and geospatial objects at unit record level without compromising privacy of micro data.

Principle 3: Common geographies for dissemination of statistics

To enable comparisons across data sets from different sources, the GSGF recommends that a common set of geographies be used for the display, reporting, and analysis of social, economic, and environmental information. While the EG–ISGI recognises the importance of traditional statistical and administrative geographies, it also recommends NSIs to consider the benefits of gridded data as grids provide consistent geography for disseminating and integrating information.

Principle 4: Statistical and geospatial interoperability – Data, standards, and processes

Both the statistical and geospatial data communities operate their own general data models and metadata capabilities; however, these are often not applied universally. The statistical community uses the Generic Statistical Information Model (GSIM), the Statistical Data and Metadata Exchange (SDMX), and the Data Documentation Initiative mechanisms. The geospatial community, meanwhile, makes use of the General Feature Model (GFM) and the ISO 19115 metadata standard, plus a number of application specific standards. A good overview of various standards, frameworks, and models can be found on the website of the joint UNECE/UN–GGIM: Europe Workshop on Standards in Stockholm, November 2017.[8]

Within the statistical community, there is a need to build geospatial processes and standards into statistical business processes in a more consistent manner. Thus, the EG–ISGI has recognised that a top-down approach is required to incorporate geospatial frameworks, standards, and processes more explicitly into the Common Statistical Production Architecture (CSPA) and its components.

UNECE has revised the Generic Statistical Business Process Model (GSBPM)[9] during 2018, including better descriptions of the use of geospatial data and methods in the statistical production process.

Principle 5: Accessible and usable geospatially enabled statistics

This principle of the GSGF emphasises the need to identify or, where required, develop policies, standards, and guidelines that support the release, access, analysis, and visualisation of geospatially enabled information. There is a wide range of legislative and operational issues about which organisations should be aware when releasing and analysing information about people and businesses in a spatial context. One important aspect of this principle is to ensure that data can be accessed using safe mechanisms that not only protect privacy and confidentiality, but also enable access to data to undertake various analyses that foster decision-making.

Including geospatial information in statistical production

The role of geospatial information

If the GSGF gives guidance on *what* is required to perform integration of statistical and geospatial information, results from the GEOSTAT 2 project are useful in understanding *how* to incorporate geospatial information in statistical processes.

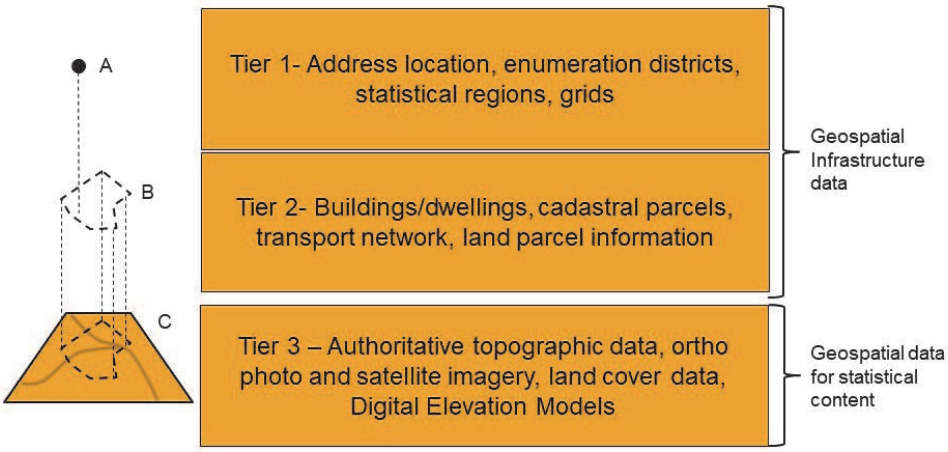
At first, it is important to establish the role of geospatial information in statistical production. It is helpful to distinguish between the information needed as the *infrastructure* for geocoding data and the geospatial information needed to *create statistical content*.

The different roles of geospatial information in statistical production are illustrated in figure 2. A workplace geocoded to an address location (A) can be linked to a cadastral parcel (B), from which land use can be assessed by combining the parcel with a land use map (C). The more consistent the system, the more opportunities for flexible linking of data.

[8] https://www.unece.org/index.php?id=45404
[9] http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0

Figure 2

Tiers of information in the generic statistical-geospatial data model for the production and dissemination of statistics



The following are the three different tiers of geospatial information, as identified in the GEOSTAT 2 project:

Tier 1 consists of geospatial information used exclusively for the purpose of geocoding, geographically representing, or disseminating statistical or administrative information. Examples of information in Tier 1 include address data, census enumeration districts, postal code areas, statistical grids, or other statistical or administrative geographies.

Tier 2 consists of geospatial information that is used to: 1. geocode, geographically represent, or disseminate statistical or administrative information, and 2. create statistical content. Typical information found in Tier 2 include not only those on buildings, cadastral parcels, and transport networks, but also new data sources, such as traffic sensor information.

Tier 3 consists of geospatial information that is used only to produce statistical content. Despite its geospatial component, this category of information cannot be used directly to geocode statistical or administrative data. Thus, information in Tier 3 can be regarded as complementary to, and independent from, information in Tiers 1 and 2.

Some examples of data found in Tier 3 include digital elevation models, land use or land cover maps, topographic data, orthophoto or satellite imagery, and other products derived from earth observation data. Typically, data in Tier 3 needs to be combined with data from Tier 1 or 2 to be transformed into statistical information.

The calculation of land area within a NUTS region can serve as an example: data on land mass and topographic maps data (Tier 3) is combined with a data set containing NUTS regions (Tier 1) and the land (and water) area for each NUTS region is calculated. NUTS is the Nomenclature of territorial units for statistics, a hierarchical system for dividing the economic territory of the EU.

A point-based geocoding infrastructure

The GEOSTAT 2 proposes a point-based approach to flexible statistical production, allowing for all kinds of statistics at the local level. In a fundamental sense, a point-based geocoding infrastructure for statistics can be understood as a production setting where a record holding the X, Y (and Z) coordinates of a location, along with a unique identifier (Id) and time stamp. This process is called the 'geocoding' of statistics or other data. The actual purpose of the point-based approach is to assign a single coordinate to each unit record. This point and the associated record can then be flexibly linked to a statistical or administrative data this point is located within.

Figure 3

The conceptual difference between point-based and area-based geocoding infrastructures

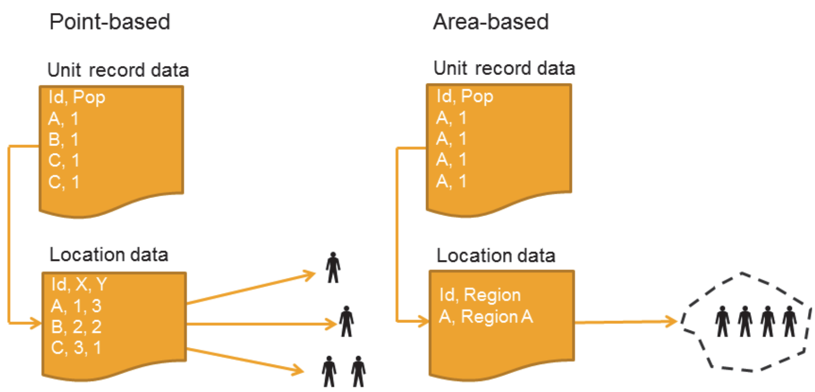


Figure 3 shows the conceptual differences between a point-based and an area-based geocoding infrastructure. In both cases, there is a record with statistical data consisting of four individuals, and a record corresponding to it that contains location data. In the point-based approach, shown on the left, each individual in the statistical data records are linked to a unique dwelling location which has not been aggregated, and is spatially represented by three different point locations. Two individuals have been assigned the same location as they are linked to the same dwelling. In the area-based approach on the right, all four individuals are linked to the same spatial object (Block A); this is because the area-based approach does not support spatial discrimination of their respective dwelling locations.

It should be stressed that the proposed shift from an area-based approach to a point-based one, as described here, only refers to the geocoding infrastructure itself and, hence, to the collection and processing of statistical information. The area-based approach is, and will continue to be, the primary method for the dissemination of statistics in order to protect privacy. The point-based approach can be implemented in the context of traditional census data collection, as well as in that of the administrative database.

Processes involving geospatial integration

GEOSTAT 2 identified thirteen tasks linked to the first six phases of the GSBPM; these phases are needed to set up and maintain a point-based geocoding infrastructure. Phases 7 and 8 of the GSBPM (Dissemination and Evaluation) are beyond the scope of the project. The tasks are as follows:

1. *Specify needs*: Find out what the users need; promote geospatial statistics and the potential of geospatial information.

2. *Design*: Recognise geospatial data sources; assess data processing capacity; specify geospatial statistics output.

3. *Build*: Create a flexible production set-up; build the geocoded survey frame.

4. *Collect*: Obtain and manage geospatial data.

5. *Process*: Conduct geospatial data quality assessment; assess identifiers to enable correct data linkage; geocode data; prepare geospatial statistics products.

6. *Analyse*: Assess data dissemination constraints.

The GEOSTAT 2 report gives detailed guidance for each of the thirteen activities, underpinned by national case studies from the countries participating in the project.

Statistics at the local level — Is the terminology clear enough?

Working with statistical-geospatial integration for a number of years, I have come across a number of different terms used to describe statistics at the local area level. Often, they relate to the fact that there is a strong geospatial component: location-based statistics, geographically referenced statistics, location-enabled statistics, or just geospatial statistics. They may also indicate that the NSI is using registers or censuses in the statistical production, linking statistical data to place with a bottom-up approach. In this paper, a number of these terms are used interchangeably.

There are also a number of terms relating to more model-based top-down approaches, such as small area statistics resulting from small area estimation[10], spatial statistics, or geostatistics. However, at times, one might find that these terms are not

[10] http://www.nss.gov.au/nss/home.NSF/pages/Small+Areas+Estimates?OpenDocument

used very consistently; thus, the EG–ISGI needs to keep working on terminology related to geospatial statistics and other related terms.

Let us assume that users, in general, just want to find relevant statistics for different levels of regional or local detail. The subject matter areas should display these levels as clearly as possible in the statistical databases and associated metadata. However, there should also be ways for users to start searching for statistics by providing the name of a place, which could be a municipality, neighbourhood, or locality. Some additional tools are required to help users work with statistics at the local area level related to maps.

The volume of statistics on various regional and local levels will vary, owing to the need for appropriate data sources and confidentiality concerns. Even if it is tempting to start using very specific terms for statistics based on geospatial information, in the long run, our users will benefit more with simple terms, that is, keeping the statistical content in focus, but adding user-friendly tools to better understand local dimension.

While keeping it simple for users, there are many considerations to be taken into account for NSIs embarking on increased production of statistics requiring statistical-geospatial integration. The complexity increases even more when data sources, such as earth observations, are being considered for use.

At Statistics Sweden, regional statistics can be produced simply by supplying the code or the name of the administrative region in one's register or survey. Regional statistics is produced by a large number of statisticians at Statistics Sweden, who have no need for specific geospatial skills to link people or businesses to place. Statistics at the local level, on the other hand, and statistics on geospatial phenomena, such as land use, are only produced by a limited number of experts in geospatial analysis and geospatial tools.

The capability to integrate statistical and geospatial information

Statistics Sweden's capability assessment

The developments and challenges lead to a question about the capability that an NSI needs to develop to increase the volume of statistics at the local level. A capability enables a statistical organisation to undertake one or more activities. The general elements that are required for a capability, according to the Common Statistical Production Architecture[11], are people, processes, methods, technology, standards, and frameworks.

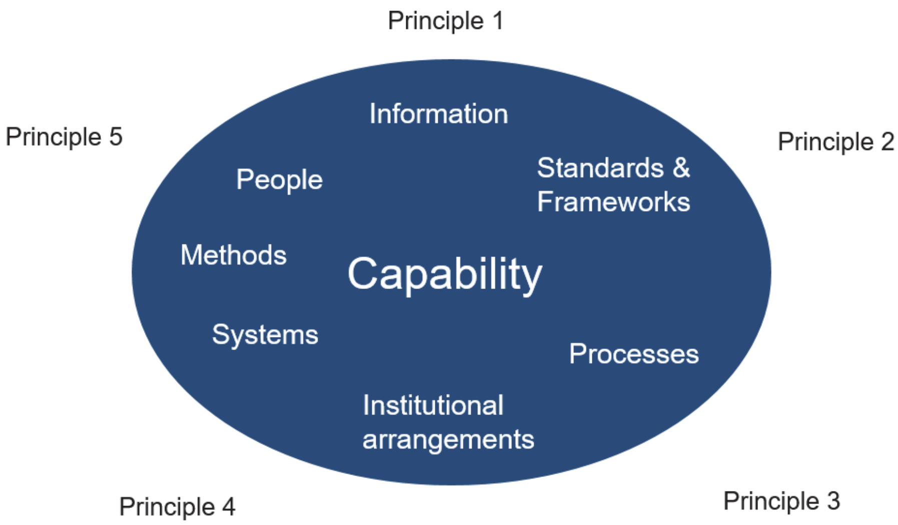
As part of the GEOSTAT 3 project, working on how to apply the GSGF in Europe, Statistics Sweden has performed a national assessment of each of the framework's principles. Looking at the framework from a capability point of view proved

[11] https://statswiki.unece.org/display/CSPA/Common+Statistical+Production+Architecture

helpful, as the framework principles need to be underpinned by certain skills, processes, methods, and so on. The assessment showed that in the case of statistical and geospatial integration the elements underpinning the capability may be expanded to include information and institutional arrangements, too (Haldorson–Moström 2018).

Figure 4

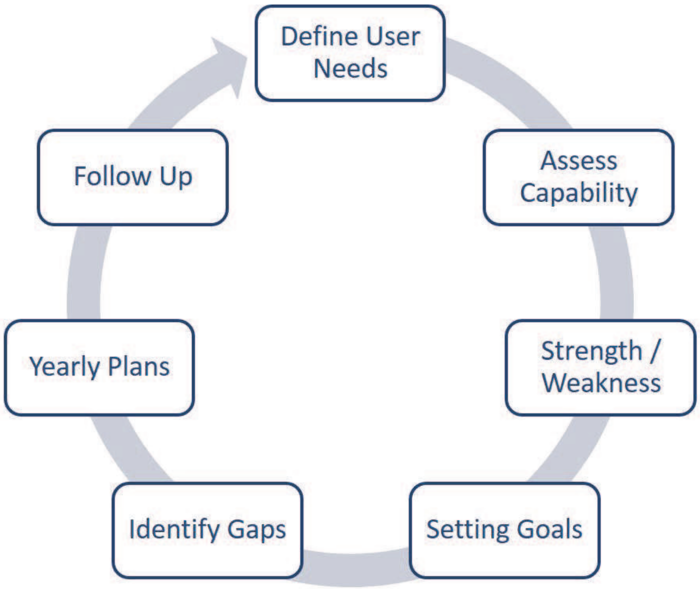
Elements needed for an NSI to have the capability to integrate statistical and geospatial information



The assessment exercise can easily be conducted by NSIs: those who have recently started to work on statistical-geospatial integration, as well as those looking for a way to improve it. The steps proposed in the Swedish report are simple and start with user needs.

Figure 5

The process of identifying how to improve statistical-geospatial integration at Statistics Sweden



Looking at each of the elements underpinning a capability for each principle in the framework revealed the strengths and weaknesses of Statistics Sweden. Setting goals and deciding which activities to prioritise makes it easier to include actions in yearly plans and to follow up. The actions can be categorised as either new activities that need to be initiated or as those that need to be improved. The assessment has also shown the activities that need to be maintained or reduced.

The following are the results of the assessment for each capability element; they cover the important aspects related to the five principles of the GSGF.

Information

This element covers the data needed to perform statistical and geospatial integration activities. Statistics Sweden has a register-based statistical system (Wallgren–Wallgren 2007) which is well suited for producing statistics at regional and local levels. All administrative registers used by Statistics Sweden include information on location, either directly (real estate, address, building) or indirectly (through personal identification number, work place id, etc.). They also include information on administrative geography, such as a municipality or a county.

Statistics Sweden maintains a national register on real estate, with weekly updates, from the Geospatial Agency. The register serves as a full-fledged location data framework, comprising a national address register, building register, and dwelling register. It includes coordinates and unique ids; this means that the information in the register can be linked to the corresponding cadastral parcels on the cadastral map. To facilitate statistical production, the register stores time series as well as yearly versions.

All references to location are stored in a separate Geography Database (GDB). The GDB is the central hub for all geocoding activities at Statistics Sweden. It is a key database used to connect (link) geospatial locations with data from different administrative registers using a set of identifiers. The general principle is to leave data at its source to avoid the necessity to update geospatial variables in more registers than the GDB.

The GDB currently lacks the geospatial link between real estate and environment and/or land use. However, it can be established by linking the geospatial features of cadastral parcels (polygons) on the cadastral map with the real estate information in the real estate register. The method has been used to produce land use accounts with good results. In the future, cadastral parcels should be included in the GDB.

Authoritative boundary data for administrative geographies are produced, maintained, and provided by the Geospatial Agency as part of the National Spatial Data Infrastructure (Lantmäteriet 2017): Codes are retrieved from Statistics Sweden's database. Boundaries and codes for other non-administrative geographies, such as localities and neighbourhood areas, are produced and maintained by Statistics Sweden.

People

This element covers the skills, knowledge, and aptitudes required in people undertaking the various statistical activities. Statistical-geospatial integration is mainly carried out by two groups of experts in Geographical Information Systems, GIS, and geospatial analysis at Statistics Sweden, both in the Regions and Environment Department. One group is responsible for land use and real estate statistics, and the other for statistics at the local level.

Statistics on administrative geographies are widely produced in many subject matter units, covering for example statistics on population by age, income, education or household status. As most registers contain references to administrative units (county or municipality), aggregating and disseminating statistics for these units typically does not require the participation of geospatial experts or special methods. For example, the metadata system of the Statistical Database automatically links the national coding system with the NUTS system.

Knowledge of statistical and geospatial models, frameworks, and standards lies with different people in the organisation; thus, greater collaboration is needed. The potential of geospatial analysis is not widespread within the organisation.

Methods

This element covers the set of specific techniques or algorithms required to undertake the activity. As explained earlier, Statistics Sweden has a register-based system, so the methods are mainly concerned with ways to integrate various data sources by using a bottom-up approach.

Statistics Sweden uses a cadastral parcel as a geocode type, together with addresses, buildings and dwellings. All objects are hierarchically linked. This creates a very flexible system that allows for all kinds of geographic classification and aggregation. All administrative geographies are hierarchical spatial units built on the cadastral parcels (property units) as the smallest spatial element. A change in property structures (property units being merged or split) can affect the administrative borders.

Owing to the comprehensive area coding in the GDB, standard aggregation of statistics into different statistical units is possible without using spatial operations. An even aggregation of grid statistics is based on Structured Query Language (SQL) statements.

There is a need to ensure the use of better integration methods based on the statistical production model. The knowledge of geospatial methods is limited to a few geospatial experts, this being the reason why the rest of the organisation might not recognise the potential of using more geospatial information and tools in statistical production – for example in sampling or editing

Among the new activities identified after the assessment is the need to provide guidance on the use of geospatial data and geocoding practices similarly to the Australian Bureau of Statistics documentation online[12]. This will help to streamline the internal processes at Statistics Sweden, besides helping other authorities responsible for official statistics.

Further, there is a need to find a balance between the obligation to protect confidential information and requests from users to access information with maximum detail. Disclosure control is applied according to a well-established method, but with more flexible output systems, this might need to be changed.

Standards and frameworks

This element covers the standards, frameworks, guides, and policies required to undertake the activities. At a high level, the enterprise architecture at Statistics Sweden needs to incorporate geospatial elements better.

Relating to more technical standards used for integration, Statistics Sweden uses INSPIRE standards and the Swedish Metadata Profile when providing search, view,

[12] http://www.abs.gov.au/geography

and download services through the Swedish Geoportal. INSPIRE is the Infrastructure for Spatial Information in Europe[13].

As part of the Open Data initiative, the World Wide Web Consortium has suggested the Data Catalogue Vocabulary, DCAT, as a standard that will support the discovery use case of all types of information. Specific application profiles of DCAT to geospatial information (GeoDCAT) and statistics (StatDCAT) are being developed that are interoperable with ISO 19115 and the Statistical Data and Metadata eXchange, SDMX, respectively. Statistics Sweden needs to explore these standards further.

Implementation of international web service standards also needs to improve.

Systems

This element covers the Information and Communications Technology, ICT, applications, hardware, and platforms required by the organisation to undertake geospatial activities.

The technical infrastructure is based on SQL server as the standard database environment. Microsoft SQL versions prior to 2008 did not support the use of geometric and spatial SQL. With the introduction of spatial SQL, more and more operations have been moved from the desktop environment to the SQL server environment. Production of most grid statistics or the statistics of small areas has been successfully streamlined in Statistics Sweden by using SQL server. However, many processes, such as highly advanced spatial analyses and map-making, still require desktop GIS. The standard desktop GIS software packages at Statistics Sweden are MapInfo, ESRI ArcGIS, and FME. Recently, Q-GIS has also been evaluated for use.

Currently, Statistics Sweden supplies geospatial statistics through traditional file transfer procedures. This also accounts for most of the intake of data from external organisations. Registers and geospatial data from other producers are downloaded through File Transfer Protocol (FTP), or similar secure services and stored as SQL tables or GIS files on servers.

The statistical geographies produced and maintained by Statistics Sweden are accessible under open data licensing and can be searched for and accessed through the national geoportal, as well as from Statistics Sweden's website. The Statistical Database, based on the PX-web tool, is the main dissemination platform at Statistics Sweden.[14] Data in PX-web can be accessed through an Application Programming Interface (API), and more and more applications, both within and outside Statistics Sweden, are using the API to fetch data from the Statistical Database. However, no solution currently exists for the linking of statistical information from the Statistical Database to interactive maps.

[13] https://inspire.ec.europa.eu/

[14] http://www.statistikdatabasen.scb.se/pxweb/en/ssd/

The Swedish Geodata Portal[15] is used to provide search and view services; links to download services are available. All services are provided as per to the Swedish Metadata Profile.

In 2016, Statistics Sweden launched a new service called REGINA[16], the online service for territorial units, to enable users to easily track all changes in administrative geographies since 1952. A new feature was added in 2017 to enable users to view older geographies through a map service.

Processes

This element covers the preferred set of steps or tasks undertaken by the business to perform a particular activity in an efficient and effective manner. These will reference methods, standards and systems where necessary.

Statistics Sweden obtains data on cadastral parcels, buildings and address locations weekly, and data are stored in historicised tables, as well as annual, situational extracts used for regular statistical production. The data transfer process has been automated through a special service.

Geo-referencing and grid data services follow a standardised procedure based on a relatively limited set of core data comprising mainly the GDB for spatial locations and statistical data retrieved from the Population Register, Real Estate Tax Assessment Register, and the Business Register, among others.

Production of geo-referenced official statistics, and/or official statistics without explicit geo-referencing, but where geospatial applications are used, includes processes based on both the core data mentioned above and a wide range of other geospatial data/map data from different agencies.

Statistical production at Statistics Sweden is carried out according to guidance material available in a specific 'process support environment'; for each process in the production chain, many checklists, methodology support documents, and templates have been developed over the last few years. However, guidance material on geocoding practice or geocoding and management of geospatial data is currently missing. Guidance material concerning the use of geographies for dissemination of data could also be improved.

Further, there is a need for a systematic approach to the geospatial integration of all registers, resulting in improved geospatial accuracy. This will save time in the processes later on, because no additional quality improvements and checking should be required.

The results from the GEOSTAT 2 project show that the GSBPM could be extended to include geospatial data management. There is still a lot of work to be done to include geospatial processes into the statistical process chart for Statistics Sweden.

[15] https://geodata.se/
[16] http://regina.scb.se/

Institutional arrangements

This element covers the legal and institutional infrastructure required to conduct and support the activities.

The Swedish Spatial Data Infrastructure, coordinated by the Geospatial Agency, is based on a number of important cornerstones. These are the Swedish Act and Ordinance on Spatial Information, National Geodata Strategy, the Geodata Cooperation, the standardisation work and the technical solution with a national geodata portal, and the links to the European INSPIRE Geoportal. The Geodata Cooperation began in 2011 and the revised agreement in 2015 acknowledged the fact that more and more spatial data is being provided as open data for free. For an annual fee, a joint licence is available that grants unlimited access to the geospatial data.

The legal framework for production of geospatial statistics in Sweden is closely related to the use of administrative sources and registers. Virtually all geocoded statistics are derived from administrative sources – the spatial references and the statistical micro data. Access to administrative sources for statistical production is regulated through the ordinance for official statistics (2001:100), which states that public agencies and authorities are obligated to contribute any data they have, if needed for production of official statistics. The Geospatial Agency provides the Swedish Real Property Register (Cadastre), regulated under the Real Property Register Act (2000:224), to Statistics Sweden.

In order to improve the coordination of the management and use of administrative and statistical geographies within the National Statistical System (NSS), Statistics Sweden as the coordinating body of the NSS needs to address this issue and aim for an agreement between agencies responsible for official statistics. All the geographies used for dissemination of official statistics within the NSS should be disseminated as harmonised services that are well known and used by all producers.

The level of collaboration between Statistics Sweden and the Geospatial Agency is already outstanding, but further improvements are always possible. There is a need for further collaboration with the standards community.

Statistics at the small area level might be sensitive from a privacy and confidentiality perspective; thus, the application of national privacy laws and agreed privacy standards is a top priority to ensure public trust.

The Swedish Geodata Strategy has open data as one of its priorities. Statistics Sweden provides geospatial statistics both as open data (part of official statistics) and as commissioned services. Land use statistics, for example, are published in the statistical databases and illustrated in various articles and through infographics on social media.

There is a strong trend towards more open data and there is a need to make a good business case that highlights the possibility of making detailed and interesting facts at the regional level accessible to society. Starting in 2019, more statistics at the local level will be released by Statistics Sweden as open data.

The way forward

The 2030 Agenda for sustainable development will not be achieved without finding new ways to measure a nation's society, economy, and environment. Many of the targets need to be addressed at the local level, so that facts on various local levels showing the current situation and changes over time will help in leaving no one behind. For NSIs, this poses the great challenge of collaborating with different data providers, such as agencies with administrative registers, geospatial agencies and space agencies, as well as with academia and others to apply new methods to produce relevant indicators. It also provides opportunities for NSIs to become even more relevant to their users, learning how to describe our society, economy, and environment in diverse ways.

Linking society and economy to a place helps in explaining the local and regional situation and bringing the numbers closer to the citizens. Statistics at the local level helps NSIs to stay relevant to citizens looking for facts showing the differences within a region. However, making statistics at the local level into official statistics also comes with challenges regarding safeguarding privacy and providing easy-to-use visualisation tools.

By implementing the Global Statistical Geospatial Framework in the national context NSIs get support in how to setup a successful production of statistics at local level. Adding a capability aspect to the framework can help in assessing where the NSIs need to improve. . Engaging with users, building the necessary capability, and setting cooperation mechanisms in place are parts of the process. Australia can serve as a very good example with their history of good collaboration between the NSI and the Geospatial Agency. The GEOSTAT 2 project gives useful guidance on how to start with statistical geospatial integration, pointing to the benefits of using the GSBPM.

The way forward is to modernise production of official statistics, so that statistics at the local level becomes part of the statistical infrastructure of more and more countries. Finding some countries to benchmark against is a good way to start, using common frameworks and models is helpful with the goal to build capacity and improve the services to citizens.

Maybe the greatest challenge ahead is not how to increase the use of geospatial information, but how to get access to suitable statistical information from administrative registers to match with the geospatial information. Without regularly updated register-based information on the population qualified/suited to be linked to place, the users will not get timely and relevant statistics at the local area level.

Acknowledgements

I would like to acknowledge my colleague Jerker Moström at Statistics Sweden, who is behind many of the innovative results shared in this paper. I would also like to thank Martin Brady at the Australian Bureau of Statistics for valuable advice on how to improve the content and make the text more relevant.

REFERENCES

HALDORSON, M.–ZACCHEDDU, P.–FOHGRUB, B.–PETRI, E. (2016): Geospatial information management in Europe – responding to the user needs *Statistical Journal of the IAOS* 32 32(4): 481–487. 10.3233/SJI-161010

HALDORSON, M.–MOSTRÖM, J. (2018): *Implementing the Statistical Geospatial Framework at Statistics Sweden, National Report as part of the GEOSTAT 3 project* https://www.efgs.info/wp-content/uploads/geostat/3/Impementing-the-GSGF-at-Statistics-Sweden-v0.96_Draft_for_GEOSTAT.pdf (downloaded: August 2018)

LANTMÄTERIET (2017*): Swedish Spatial Data Infrastructure and the National Geodata Strategy. Country Report of Sweden* Submitted to United Nations Committee of Experts on Global Geospatial Information Management Seventh Session, New York. http://ggim.un.org/country-reports/documents/Sweden_Land_report_2018_UNGGIM_Final.pdf (downloaded: August 2017)

MOSTRÖM, J. (ed.) et al. (2017): *A Point-based Foundation for Statistics, Final report from the GEOSTAT 2 project. EFGS and* Eurostat http://www.efgs.info/wp-content/uploads/2017/03/GEOSTAT2ReportMain.pdf (downloaded: August 2018)

UN EG-ISGI (2016): *Background Document on Proposal for a Global Statistical Geospatial Framework Advanced Draft as of 28/07/2016*. http://ggim.un.org/meetings/GGIM-committee/documents/GGIM6/Background-Paper-Proposal-for-a-global-statistical-geospatial-framework.pdf (downloaded: August 2017)

UNITED NATIONS GENERAL ASSEMBLY (2015). *Resolution 70/1. Transforming our world: the 2030 Agenda for Sustainable Development* https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf (downloaded: August 2018)

VAN HALDEREN, G.–MINCHIN, S.–BRADY, M.–SCOTT, G. (2016): Integrating statistical and geospatial information, cultures and professions: International developments and Australian experience *Statistical Journal of the IAOS* 32(4) 457–470. https://doi.org/10.3233/SJI-161018

WALLGREN, A.–WALLGREN, B. (2007): *Register-based Statistics: Administrative Data for Statistical Purposes* John Wiley & Sons, Ltd Hoboken. https://doi.org/110.1002/9780470061350