Comparative Economic Research, Volume 21, Number 1, 2018
10.2478/cer–2018–0005

DE GRUYTER
OPEN

UNIWERSYTET
ŁÓDZKI

ENGIN SORHUN*

Is Economic Integration a Historical Shock to City-size Distribution?

Abstract

Based on the assumption that the economic integration process contributes, via market reforms, to the dynamics of the space distribution in candidate countries, this study examines (i) whether agglomeration forces or dispersion forces are dominant; (ii) whether EU-integration causes a structural break to the space distribution over time; (iii) whether EU-integration makes the city-size distribution more even or uneven in eight eastern European Union members (EU–8). To carry out the analysis, the Ziwot-Andrew and Cusum Square tests are used to detect structural breaks; the ARDL Bound test is used to reveal the interaction between long-run and short-run equilibrium; and the Granger test is used to determine the direction of the causality among the variables. The main results are: the integration with the EU (i) caused a structural break to the city-size distribution, (ii) made the city-size distribution more uneven and (iii) stimulated the agglomerating forces over the spreading forces in the EU–8.

Keywords: *City-size distribution, rank-size rule, Zipf's Law, EU integration, new member and candidate countries, ARDL, Bound test, Granger causality test*

JEL*: R12, R15*

---

[*] Assoc. Prof., Istanbul 29 Mayis University, Faculty of Economics and Administrative Sciences, Department of Economics, e-mail: esorhun@29mayis.edu.tr

1. Introduction

The distribution of people and economic activities represent regularities that have been subject to a wide range of studies in the empirical literature. Krugman (1995) described three particular empirical regularities: the gravity model of trade, the market potential analysis, and the equation underlying the rank-size distribution. Focusing on the latter, this study investigates whether joining an economically integrated entity causes a shock to the rank-size distribution of cities through the accession process.

In empirically examining the impact of an economic integration on the city-size distribution (CSD), the accession of the Central and Eastern European Countries (CEECs) to the European Union (EU) represents a concrete case. In order to become a member, the CEECs had to execute the demanded reforms (e.g. implement the *acquis communautaire*). Each step toward the EU membership brought about concerns relative to the impacts of regional integration on the economic, social, and legal structures of the CEECs. As the EU Commission reported, the implementation of the more-than-thirty-chapter long *acquis communautaire* accelerated the convergence of CEECs to the EU–15's level.

Under the assumption that the EU integration process has an impact, through market reforms, on the distribution of economic activities, this paper examines, in a dynamic perspective, first what kind of forces (agglomeration or dispersion forces) are dominant; second, to what extent the economic integration explains the developments in the city-size distribution; and third whether the EU integration makes the CSD more even or uneven for the eight CEECs (the EU–8[1]).

The paper is organized as follows: the next section 2 contains a brief review of the literature on the meaning of CSD and the debates on measurement issues. Section 3 presents the estimation model to test the impact of economic integration as a historical shock on space distribution in the candidate countries. Section 4 defines data sets. Section 5 presents the econometric methodology and empirical results. The final section offers conclusions from the study.

2. The Meaning of Zip's Law as a Measure of CSD

The most commonly used model to compute rank-size distribution is Zipfs's Law (Zipf, 1949). The number of cities of large size seems to decrease according to a rather regular geometric progression, which depends on their rank in the

---

[1] Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovak Republic, Slovenia.

urban hierarchy. The law expresses the assumption that when cities are ordered by population size, regressing the logarithm of their rank on the logarithm of their size yields a slope coefficient close to minus one. Using Pareto distribution, Zipf tried to approximate the distribution of sizes of cities where a linear regression of log-rank on log-size of cities shows the largest city in the sample is more than *b* times as large as the *b*th largest city. Taking exponents, the relationship can be seen to be a special case of a power rule relating the size rank of a city (*R*) to some power of its population size (*M*).

$$lnR_j = \alpha - \beta ln M_j \tag{1}$$

where α is constant, M_j is the size of city *j* (measured by its population), and R_j is the rank of city *j* (rank 1 for the biggest city, rank 2 for the second biggest city, etc.). In empirical research β is the estimated coefficient, giving the slope of the log-linear relationship between city size and city rank. This means that if and only if $\beta=1$, Zipf's law holds. If β is smaller than unity, a more even distribution of city sizes results than predicted by Zipf's law. That is to say, if $\beta=0$ all cities are of the same size. If β is larger than unity, the large cities are larger than Zipf's law predicts, implying more urban agglomeration (the larger city is more than *b* times as large as *b*th largest city. Empirically, if the rank-size distribution holds, the question to pose is whether $\beta=1$ or not.

Although the empirical validity of Zipf's law is debated by many authors, they reach a consensus on two points: (i) Zipf's law holds *proximately* but not *absolutely* true (the coefficient's value varies around the unity), and (ii) the CSD (measured through the estimated β value) changes over time. Geographical Economy and Urban Economy explain the intertemporal variation in the β value through two major effects: the population effect (proportional growth in urban populations) pointing out the natural development of the urban hierarchy over time, and that historical shocks create break points and radical transformations through disasters, policies, social events, wars, etc. that create an additional effect to the variation in the β value. Both fields explain not only the dynamic variation in a single urban hierarchy over time, but also the cross-country difference in the β value as a result of the balance between the agglomerating forces (economies of scale, low transport costs, market potential, spillovers, externalities, high wage levels etc.) and the spreading forces (diseconomies, congestion, high land costs, costs of living, pollution etc.) differs between individual cities. Since the balance between agglomerating and spreading forces differs between individual countries (Rosen and Resnick, 1980; Soo, 2005), it helps to explain the differences in the estimated coefficients for CSDs of different countries.

The effect of great shocks (war, disease, natural catastrophe, revolution, etc.) having the potential to radically change the CSD has been empirically tested in previous studies, such as: the impacts of historical shock (Davis and Weinstein, 2002);

government policies (Anderson and Ge, 2005); socio-economic transformations (Benguigui and Lieberthal, 2007) on the CSD dynamics; and the impacts of World War II (Bosker et al, 2008); and the German division and reunification (Redding and Sturm, 2005) on the German CSD dynamics. However, they also emphasized that historical shocks can, but not necessarily does, have *permanent* impacts on the relative position of cities within the hierarchical distribution.

As for the impact of historical shocks on the CSD, there is no consensus on the fact that can be related to the Zipfs's Law in the previous literature: Fujita et al (1999) and Brakman et al (2001) noted that CSD is sensitive to shocks and that this does not necessarily adhere to Zipf's law. On the one hand "insofar as the size distribution is tied to technological characteristics […], then as we move from hunter-gatherer societies, to agriculture-based societies, through feudalism, into and out of autarky, and finally to a modern industrial economy, one might guess that there would be radical shifts over time in […] the densely populated regions" on the other hand "even strong shocks should shortly be reversed as the particular locations reassert themselves in relatively rapid growth rates on the path to recovery" (Davis and Weinstein, 2002). Regardless of whether the shocks are permanent or temporary, considering the impact of large shocks on the CSD, it is important to have a theoretical means to explain the spatial change. To stress the importance of theoretical attempts to model the eventual impacts of historical shocks on the CSD, Krugman (1998) pointed out that "the new work is highly suggestive, particularly in indicating how historical accident can shape economic geography".

Many authors have described "forces" or "effects" to determine developmental stages of cities.[2] Brakman et al (2001) adapted Kooij's (1988) urbanization pattern,[3] which is coherent with agglomeration and spreading forces to their rank-size distribution analysis for the Netherlands and makes a useful interpretation of the dynamic β coefficient: change in the β value over time depends on changes in the balance between forces. Following a progressive pattern, the developmental stage

---

[2] Weber's (1909) "location triangle", Myrdal's (1957) "backwash and spread effects", Hirscman's (1958) "backward-forward linkage", and Krugman's (1993) "first nature-second nature" are the major examples.

[3] Kooij (1988) distinguishes three stages of urbanization: (i) Pre-industrialization, characterized by high transport costs, substitute products, and production being dominated by immobile farmers; (ii) Industrialization, characterized by declining transport cost and the growing industrial production with increasing return to scale; and (iii) Post-industrialization, characterized by the declining importance of industrial production and increasing importance of negative externalities, like congestion. In the first stage, there is a low level of integration due to high transport costs. In the second stage, the decrease in transport cost pushes some cities to expand and become larger. Agglomeration forces dominate during this period. In the third stage, transport costs remain low, but the manufacturing sector is characterized by differentiated products and increasing returns to scale. Nevertheless spreading forces, or so-called congestion effects, like diseconomies, traffic jams, pollution, criminality, rising land rents in larger cities etc. emerge in this period.

of a city begins with β well below unity (pre-agglomeration). As the city expands, the β value increases (agglomerating forces). And when it reaches a certain maturity level, then the β value starts to decrease (spreading forces).

3. Testing Economic Integration as a Historical Shock to CSD

The last two enlargement of the EU constitute a recent example that will support the assumption that integration causes a historical turning point. To become a member of the EU, the CEECs had to pass a certain number of stages (adoption of the *acquis communautaire*). Each step of the EU membership brought about concerns relative to the impacts of regional integration on the economic, social, legal etc. structures within the new member states. And so, in order to test our expectation that integration into a supranational zone leads to a structural change in the dynamic CSD, an estimation model involving bilateral trade integration, FDI inflows, as well as convergence parameters is built:

$$\begin{aligned}\left|\beta_t-\beta_{t-1}\right| &= \alpha_0+\alpha_1 lnTRADECON_t+\alpha_2 lnEUFDI_t+\alpha_3 lnGDPCDIF_t+ \\ &+\alpha_4 lnPRICEDIF_t+\beta\alpha_5 lnUNEMPDIF_t+\alpha_6 lnRWAGEDIF_t+e_t\end{aligned} \tag{2}$$

TRADECON: Trade concentration is measured by the trade intensity index[4]. It takes a value between 0 and +∞. Values greater than 1 indicate an intense trade relationship.

EUFDI: Foreign direct investment inflows. The weight of the FDI inflows from the EU–15 to the EU–8 in the total FDI inflows. A sustainable rise signals a convergence.

GDPCDIF: GDP per capita differential in terms of purchasing power. It is basically the proportion of the GDP per capita average of the EU–8 to that of the EU–15. It takes values between 0 and +∞. Its unity represents a full convergence.

PRICEDIF: Price level differential. This is the proportion of the average price level of the EU–8 to that of the EU–15. It takes values between 0 and +∞. Its unity represents a full convergence.

---

[4] $TRADECON=\dfrac{\sum_{sd}X_{sd}/\sum_{sw}X_{sw}}{\sum_{wd}X_{wd}/\sum_{wy}X_{wy}}$, where *s* is the set of countries in the source, *d* is the destination, *w* and *y* represent the countries in the world, and *X* is the bilateral flows of total exports. In words, the numerator is the export share of the source region to the destination; the denominator is export share of the world to the destination.

UNEMPDIF: Unemployment differential. This is the proportion of the difference between the EU–8 and the EU–15 unemployment rates to the unemployment rate of the EU–15. It takes values between 0 and +∞. Zero indicates a full convergence.

RWAGEDIF: Real wage differential. This is the proportion of the real wage of the EU–8 to that of the EU–15. It takes values between 0 and +∞. Its unity represents a full convergence.

All the explanatory parameters are in log linear form. e_t is the error term.

In the next section the main outcomes of the estimation model will be presented under the assumption that candidate countries should meet a set of required structural market reforms (before and even after joining the EU) that affect the space distribution of activities (proxied by the β-coefficient), depending on the balance between agglomeration and spreading forces. The more reforms a country realizes on the way toward EU membership, the more its β value is subject to variation over time. Which dynamics (agglomeration or spreading forces) the membership process accelerates is another problematic issue. As a few empirical studies have pointed out, the EU membership process contributes to agglomerating forces rather than spreading forces (Combes and Overman 2003; Dupunch et al, 2004; Brakman et al, 2001).

4. Description of the data

In this paper both city proper (administrative boundaries) and agglomeration (functional urban areas-FUA methodology, proposed by OECD, 2012) definitions are adopted. And as suggested in Brakman et al. (2001), the smallest units are excluded from the sample and adopted the same city proper and agglomeration definition.[5] 14 agglomerations (based on FUA definition) and 81 cities (based CP definition) are determined. The EU–8 urban system is treated as a whole.

The full data set that involves all variables' time series goes back to 1995. City census data (covering 1995–2017) were collected from the statistics institutes of the EU–8[6], while other data were taken from Eurostat. The time series analysis, which covers a 21-year-period between 1995 and 2017, gives a considerable degree

---

[5] One tenth of the number of inhabitants of the total population of the countries is considered as the threshold of agglomeration. For the countries with less than 1,000,000 inhabitants, the threshold is 100,000 inhabitants; for the countries with less than 2.000,000 inhabitants, the threshold is 200,000; so on. For example, if there are two agglomerations with more than 200,000 inhabitants in a country with a population of more than 2,000,000, then they are considered as an agglomeration.

[6] Czech Statistical Office (CSO), Statistics Estonia (SE), Hungarian Central Statistical Office (HCSO), Central Statistical Bureau of Latvia (CSBL), Statistics Lithuania (SL), Statistics Poland (SP), Statistical Office of the Slovak Republic (SOSR), Statistical Office of the Republic of Slovenia (SORS).

of freedom. The β coefficient is estimated for each year with cross-section treatment, and a time series is obtained from the estimated values.

5. Empirical Results

The descriptive statistics are presented in Table 1. The values of Jarque-Bera show that all the variables of the model have zero mean and finite covariance; this confirms that the data sets are normally distributed.

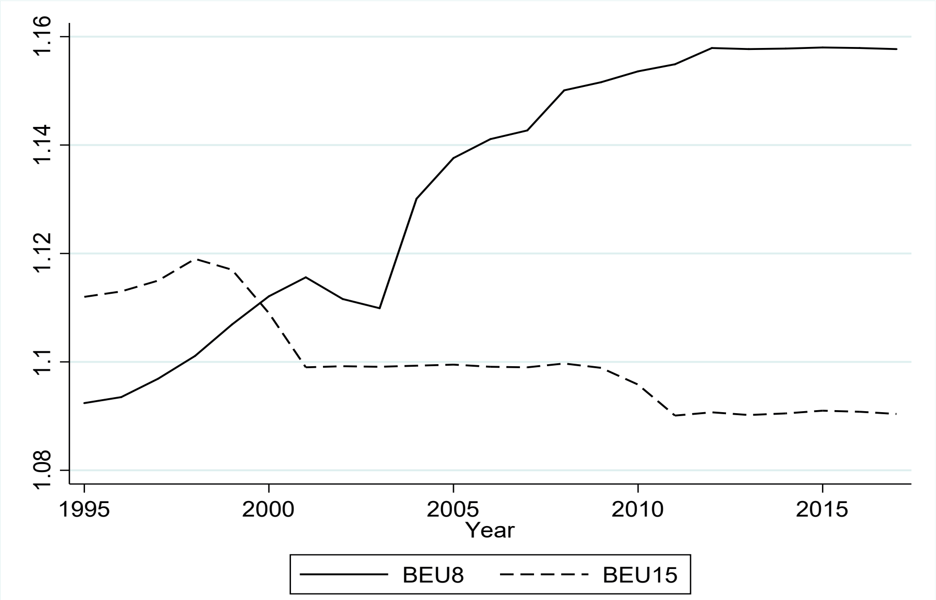
Table 1. Descriptive Statistics

	$\lvert\beta_t - \beta_{t-1}\rvert$	TRADECON	EUFDI	GDPCDIF	PRICEDIF	RWAGEDIF
Mean	0.0083	0.290	0.276	0.429	0.461	0.350
Median	0.0079	0.312	0.298	0.542	0.415	0.389
Maximum	0.0129	0.556	0.499	0.695	0.71	0.596
Minimum	0.0036	0.213	0.146	0.169	0.060	0.215
Std. Dev.	0.0423	0.142	0.229	0.121	0.132	0.395
Probability	0.455	0.322	0.765	0.365	0.654	0.564
Jarque-Bera	1.523	2.029	2.023	0.964	0.626	1.269
Observation	21	21	21	21	21	21

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

Before dealing with the estimation model outcomes, the study first examines what kind of forces (agglomeration or spreading) is dominant. Because in a sense this will alter the interpretation of the estimated coefficient of the model. Graphic 1 shows 21-year long course that the β value of the EU–15 and the EU–8 followed. However, the two subgroups' β values have different patterns: The EU–15's β value is moving in a narrow range, whereas the EU–8's β value is moving in a relatively wide range. The former seems to follow a relatively stable pattern compared to that of the latter. What's more important are the trends of the β values over time: the β value of the integrated bloc has a steadily declining trend, whereas the β value of the integrating block has an increasing trend (it seems to have stabilized recently). This signals that spreading forces dominate over agglomerating forces in the EU–15, whereas agglomeration forces prevail over spreading forces in the EU–8. This trend is distinct for EU–8 countries after the year 2004. Following the accession of the EU–8 to the EU, the β's slope changed explicitly. In the next step, the temporal variation of the city-size distribution and the eventual impact of the accession to the EU on it will be tested through the estimation model.

Stationary and non-stationary time series data lead to different outcomes. In the event of a unit-root problem (involvement of non-stationary data), the regression results may be spurious. The time series regression analysis is based on the assumption that the series used are stationary. In the literature, several unit root tests are available for testing this critical assumption and making the data stationary: Augmented Dickey Fuller (ADF), Philips-Perron (PP), Dickey Fuller Generalized Least Squares (DF-GLS), Ziwot-Andrew[7] (ZA). With regard to their features, all of them are used.



Graphic 1. The Estimated β coefficient between 1995–2017

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

Table 2 presents the results of the conventional unit-root tests: the ADF test indicates, except for lnPRICEDIF and lnUNEMPDIF, which are stationary at the first difference, the other variables are stationary at the level while the PP and DG-GLS confirm that whole variables are stationary at the level.

---

[7] In the presence of any temporal structural variation or break, the conventional unit-root tests cannot decide about the stationarity of series. The ZA may be used to test it, taking into consideration the structural variation or break.

Table 2. The Results of the Unit-Root Tests without a structural break

At the Level			
Variable	**ADF**	**PP**	**DG-GLS**
$\left\|\beta_t - \beta_{t-1}\right\|$	–2.61***	–10.7**	–3.49**
lnTRADECON	–2.72**	–9.4***	–2.89***
lnEUFDI	–0.92***	–11.6**	–3.69**
lnGDPCDIF	–2.94**	–4.2***	–1.46***
lnPRICEDIF	0.302	–12.1**	–3.11**
lnUNEMPDIF	0.105	–8.4***	–3.02***
lnRWAGEDIF	–1.99***	–10.9**	–2.78***
At the First Difference			
Variable	**ADF**	**PP**	**DG-GLS**
$\left\|\beta_t - \beta_{t-1}\right\|$	–2.69**	–12.1**	–3.57*
lnTRADECON	–2.88**	–16.1*	–3.19**
lnEUFDI	–1.24***	–9.0***	–1.45***
lnGDPCDIF	–2.97**	–15.3*	–3.81*
lnPRICEDIF	–0.52***	–11.8**	–3.25**
lnUNEMPDIF	–0.05***	–12.5**	–3.38**
lnRWAGEDIF	–2.27***	–10.3**	–3.41**

*Note: The asterisks *, ** and *** denote the significant at the 0.01, 0.05 and 0.10 levels, respectively. ADF is estimated through the model as a constant but with no trend. If the calculated tau value is less than the critical value in the table above, then we have a significant result; otherwise we accept the null hypothesis that there is a unit root and the time series is not stationary. PP is estimated through the process with a random walk without drift. DG-GLS performs a modified Dickey–Fuller t test for a unit root in which the series has been transformed by a generalized least-squares regression. The optimal lag selection is determined by the Schwarz Bayesian Criterion. One lag is considered optimal for the three tests.*

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

However, these conventional unit root tests (ADF, PP and DF-GLS) do not allow for the possibility of a structural break. Assuming the time of the break as an exogenous phenomenon, Perron (1997) indicates that the possibility of rejecting the presence of a unit root decreases when a structural break is ignored. Zivot and Andrews (1992) propose an alternative unit-root test under the assumption of unknown time of the break-point. The ZA is performed using three models to test for a unit root: (1) model A, which permits a one-time change in the level of the series (intercept); (2) model B, which allows for a one-time change in the slope of the trend function (trend), and (3) model C, which combines one-time changes in the level and the slope of the trend function of the series (intercept + trend).

Table 3. The Results of the Ziwot-Andrew Unit-root Test with a structural break

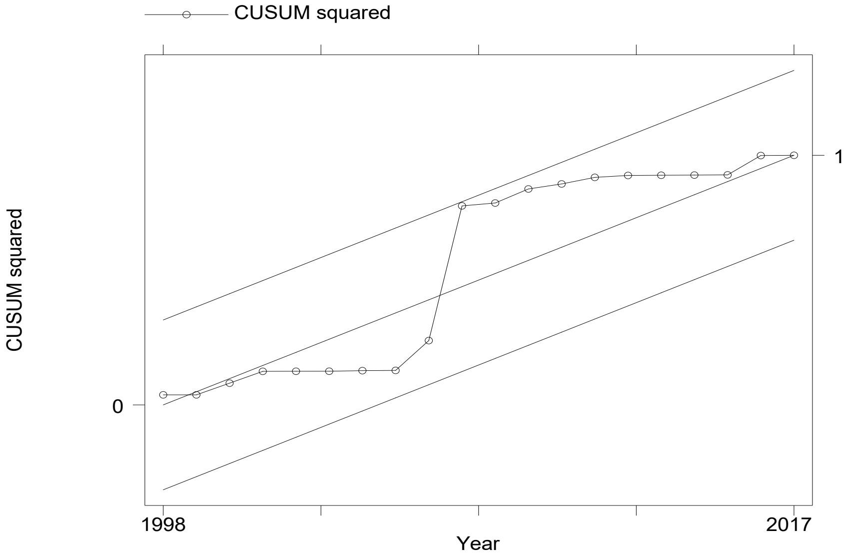
Variable		Model A (Intercept)	Model B (Trend)	Model C (Intercept + Trend)
$\left\|\beta_t - \beta_{t-1}\right\|$	t-statistics	–5.24*	–2.46***	–5,09*
	Break point	2003	2004	1999
lnTRADECON	t-statistics	–4.74**	–3.12***	–5,05**
	Break point	1997	2003	2000
lnEUFDI	t-statistics	–3.31***	–3.56***	–4,79***
	Break point	1996	2004	1999
lnGDPCDIF	t-statistics	–4.59**	–3.27***	–4,92**
	Break point	1996	2003	2003
lnPRICEDIF	t-statistics	–4.55**	–4.09***	–4,84**
	Break point	1996	2004	2004
lnUNEMPDIF	t-statistics	–3.71***	–4.19**	–4,32***
	Break point	1998	2005	2003
lnRWAGEDIF	t-statistics	–4.69**	–4.10***	–3,69***
	Break point	1997	2005	2004
Critical Values				
%1		–5,34	–4.93	–5.57
%5		–4,80	–4.42	–5.08
%10		–4,58	–4.11	–4.82

*Note: The asterisks *, ** and *** denote the significance at the 0.01, 0.05 and 0.10 levels, respectively. If the calculated value is higher than the critical value, then the series is stationary. The optimal lag selection is determined by the Akaike Information Criteria. One lag is considered optimal for the ZA.*

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

Table 3 presents the results of the ZA unit-root test with a structural break: the three models indicate different structural break years for the variables that are stationary at the level and have a different significance level. However, model A shows that the variables have a structural break in the intercept in the years 1996, 1997, and 1998 (the candidacy period before accession), while model B shows that they have a structural break in the trend in 2003, 2004, and 2005 (just before and after their integration). The break points in the intercept and the break points in the slope may be interpreted as the impact of the transition period and the impact of accession to the EU, respectively. Except for the ADF test result, the other used unit-root tests verify that the series are stationary at the level.

Alternatively, the cumulative sum of the squared test (Cusum SQ) is performed to determine whether the coefficients in a time-series regression are stable over time. The test statistic is constructed from the cumulative sum of either the recursive residuals or the ordinary least-squares (OLS) residuals. As with model B of the ZA, the Cusum SQ test indicates a high structural variation in 2005 and 2006, although it is not considered as a "break" because the trend line lies between the two critical lines which indicate that the estimated model is stable.



Graphic 2. Cusum Square Structural Break Test

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

Subsequently, to investigate the long-term relationship among the city-size distribution variation and the explanatory variables, the Autoregressive Distributive Lag (ARDL) bound testing approach is used. The first practical superiority of the ARDL cointegration approach to the conventional tests such as the Eagle-Grenger test, the Johansen test and the Johansen-Juselius test is that the series does not need to have the same order of integration. But the conventional cointegration tests assume that the series are non-stationary at the level and they become stationary at the same order of difference. Second, the conventional cointegration tests do not provide information about the structural breaks in time series data and also have a low power of prediction. The ARDL bound test approach gives efficient and valid detailed information about the structural breaks in the data. Third, this approach can be used for a small sample size.

Table 4. The Results of the Bound Test

k	R^2	F-Statistics	Level	Lower Bound	Upper Bound
6	0.61	4.76	99%1	3.43	4.99
			95%	2.86	4.38
			90%	2.57	4.04

Note: If the calculated F is less than the critical lower bound, then there is no cointegration among the variables. If it is greater than the critical upper bound, then a cointegration among the variables exists. If it is between the lower and upper bounds, then we cannot say anything about the presence of a cointegration relationship and need to perform other cointegration tests.

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

Table 5. The Estimated Long-term Coefficients of the ARDL (2,2,1,2,1,2)

Variables	Coefficients	Sign Interpretation
Constant	0.05959** (1.99890)	(+) indicates a contribution to agglomeration forces while (-) indicates a contribution to spreading forces.
lnTRADECON	0.007750 * (4.23591)	
lnEUFDI	0.000235 *** (2.00920)	
lnGDPCDIF	–0.009475** (–3.69240)	(+) indicates a contribution to spreading forces while (-) indicates a contribution to agglomeration forces.
lnPRICEDIF	–0.000459 ** (–2.76251)	
lnUNEMPDIF	–0.001012** (–4.59415)	
lnRWAGEDIF	–0.002370 *** (–3.96264)	
R^2	0.64	
χ^2_{BG}	0.207	
χ^2_{RAM}	0.056	
χ^2_{JAB}	0.000	
χ^2_{BP}	0.301	

*Note: The optimal lag selection of the variables is determined by the Akaike Information Criteria. Standard errors are presented in parentheses. The asterisks *, ** and *** denote the significance at the 0.01, 0.05 and 0.10 levels, respectively.* χ^2_{BG}, χ^2_{RAM}, χ^2_{JAB}, χ^2_{WT} *are the p-values for the Breush-Godfrey autocorrelation test, the Ramsey model specification test, the Jacque-Bera normal distribution test and the Breuch-Pagan heteroscedasticity test statistics.*

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

The steps in the ARDL bound test approach are: (i) Optimal lag order selection; (ii) testing for autocorrelation in the error term of the model; and (iii) bound test estimation. If a long-run cointegration relationship among the parameters is detected, then the Vector Error Correction Model (VECM) is applied to reveal the short-term and the long-term properties. The detection of a cointegration among the parameters is particularly conclusive to clear the doubts about a spurious regression. The calculated F-statistics (4.76) is greater than the upper bound (4.38) critical value at 0.05. So the existence of cointegration among the variables confirms the long-run relationship between the city-size distribution parameter and the convergence parameters.

The long-term relationship is presented in Table 5. The coefficient of trade convergence shows there is a positive and significant relationship with the city-size distribution. The results show that a 1 percent increase in trade convergence with the EU–15 creates 0.007 point increase in the temporal variation of the β value in the EU–8, and this relationship is significant at 1 percent. As the trade concentration with the EU–15 increases, this contributes to agglomeration forces. There is a negative and significant relationship between the city-size distribution and the differentiation in terms of GDP per capita, price level, unemployment, and real wages. A 1 percent increase in the weight of FDI inflows originated from the EU–15 explains a 0.000235 point positive variation in the β value. The estimated results show that a 1 percent decrease in these divergence parameters accounts for a 0.013 point increase in the variation of the β value, and this cumulative relationship has a 5 percent level of significance. As the EU–8 converges to the EU–15 level, this feeds agglomeration forces over spreading forces.[8] The whole of the estimated parameters accounts for a variation of 0.021 point in the city-size distribution in the EU–8 and explains 64% of the cases. Considering the trend in the β value, this result may be interpreted as the contribution of the economic integration process within the EU to the agglomeration over dispersion in the long-run. Taking into consideration the trend of the β value (Graphic 1) and the results of the long-term model estimation, the EU-integration process makes the city-size distribution more uneven (the β value moving off from unity) for the EU–8.

The short-term relationship is presented in Table 6. The short-term dynamics between the city-size distribution ($d\left|\beta_t - \beta_{t-1}\right|$) and the explanatory variables (dlnTRADECON, dlnEUFDI, dlnGDPCDIF, dlnPRICEDIF, dlnUNEMPDIF, dlnRWAGEDIF) are estimated using the VECM. An error correction term (EC(–1)) is the lagged value of the error term series of the long-term model. Except for the constant term and price level convergence parameter, the estimated coefficients are significant and have the same signs as the long-term model. This indicates that the acces-

---

[8] A dummy variable to represent the years before and after the accession of the EU–8 to the EU is also added into the model. But it is estimated to be not significant. Besides, it affects the degree of significance of the other variables negatively. For this reason, the dummy is excluded from the model.

sion processes and integration has a weak positive impact even in the short-term on the agglomeration forces in the EU–8. The negative and significant EC(–1) coefficient (–0.25215) is theoretically correct and shows the speed of adjustment from the short-term to the long-term equilibrium. This indicates that there is both a short- and long-term equilibrium in the system. The coefficient of one period lag residual is negative and significant, which represents the long-term equilibrium. The coefficient shows that the system (long-term equilibrium) corrects its previous period disequilibrium (short-term deviation) at the rate of 25.2% annually to reach a steady state.

Table 6. The Estimated Short-term Coefficients of the VECM, ARDL(2,2,1,1,1,1)

Variables	Coefficients	Sign Interpretation
Constant	1.13490 (0.56981)	(+) indicates a contribution to agglomeration forces while (-) indicates a contribution to spreading forces.
dlnTRADECON	0.032952 * (6.23581)	
dlnEUFDI	0.01489 *** (1.79350)	
dlnGDPCDIF	–0.09222*** (–2.00140)	(+) indicates a contribution to spreading forces while (-) indicates a contribution to agglomeration forces.
dlnPRICEDIF	0.02459 (0.76251)	
dlnUNEMPDIF	–0.012142** (–2.90415)	
dlnRWAGEDIF	–0.063110 *** (–3.26460)	
ECM(–1)	–0.25215** (–3.11280)	
R^2	0.45	
χ^2_{BG}	0.111	
χ^2_{RAM}	0.061	
χ^2_{JAB}	0.002	
χ^2_{BP}	0.055	

*Note: The optimal lag selection of the variables is determined by the Akaike Information Criteria. Standard errors are presented in parentheses. The asterisks *, ** and *** denote the significance at the 0.01, 0.05 and 0.10 levels, respectively.* χ^2_{BG}, χ^2_{RAM}, χ^2_{JAB}, χ^2_{WT} *are the p-values for the Breush-Godfrey autocorrelation test, the Ramsey model specification test, the Jacque-Bera normal distribution test and the Breuch-Pagan heteroscedasticity test statistics.*

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

The estimation results of Granger-Causality Tests (GCT) are presented in Table 7. The last column shows the direction of the causality. Trade concentration, FDI inflows from the EU–15, and convergence in the unemployment rate and in real wages cause a variation in the β value of the EU–8. This is unidirectional causality. However, convergence in GDP per capita causes a variation in the β value and vice versa. This shows a bidirectional causality. No significant relationship is detected between price level convergence and β variation. GCT mostly confirms the causality direction revealed in the long-term and the short-term models.

Table 7. Pairwise Granger Causality Tests

Null Hypothesis	p-value	Result	Causality Direction
TRADECON does not Granger-cause β variation	0.023	Rejected	TRADECON→ β
β variation does not Granger-cause TRADECON	0.279	Accepted	–
EUFDI does not Granger-cause β variation	0.044	Rejected	EUFDI→ β
β variation does not Granger-cause EUFDI	0.611	Accepted	–
GDPCDIF does not Granger-cause β variation	0.038	Rejected	GDPCDIF→ β
β variation does not Granger-cause GDPCDIF	0.041	Rejected	β →GDPCDIF
PRICEDIF does not Granger-cause β variation	0.776	Accepted	–
β variation does not Granger-cause PRICEDIF	0.674	Accepted	–
UNEMPDIF does not Granger-cause β variation	0.427	Rejected	UNEMPDIF→ β
β variation does not Granger-cause UNEMPDIF	0.571	Accepted	
RWAGEDIF does not Granger-cause β variation	0.029	Rejected	RWAGEDIF→ β
β variation does not Granger-cause RWAGEDIF	0.684	Accepted	–

Note: The optimal lag selection of the variables is determined by the democracy of Akaike Information Criteria (one lagged), Bayesian information criterion (one lagged). and Hannan and Quinn information criterion (two lagged). One lag is determined optimal.

Source: Own calculation based on the data from Eurostat, CSO, SE, HCSO, CSBL, SL, SP, SOSR, SORS (26.06.2017).

6. Conclusions

Under the assumption that the EU-integration process contributes via market reforms to the dynamics of the space distribution in the EU–8, this study analyzes three problematic issues: Do agglomeration or spreading forces prevail? Does the economic integration cause a structural break to the space distribution over time? And does the integration process makes the space distribution more even or uneven?

First, the dynamic space distribution of the EU–15 and the EU–8, captured by the estimated β value, in conformity with the previous literature confirms that Zipf's law holds proximately but not absolutely, and that the law represents a dynamic but not static regularity. The β value of the EU–15 has a steadily declining trend (going down from 1.1191 to 1.0792), whereas the β value of the EU–8 has an increasing trend (going up from 1.0973 to 1.1574). Adopting the interpretation of Brakman et al (2001) and Kooij (1998), the study shows that the two blocs have different patterns: spreading forces dominate over agglomerating forces in the EU–15, whereas agglomeration forces dominate over spreading forces in the EU–8. This trend is distinct for the EU–8 countries after the year 2004. Following the accession of the EU–8 to the EU, the β's slope changed explicitly.

Second, the results of Ziwot-Andrews detect the existence of structural breaks. However, for the most of the variable the detected structural breaks in the intercept occurred in 1996, 1997, 1998 (before the integration) while the detected structural breaks in the trend occurred in 2003, 2004, 2005 (right before or after the integration). This can be interpreted as follows: the breaks in the intercept can be related to the transition period, while the breaks in the trend are related to the accession process. To confirm the former interpretation, we need to test it with data also covering the transition period. But the latter interpretation is also confirmed by the Cusum SQ test that showed a high structural variation in 2005 and 2006. Both tests indicate that the EU-integration caused a historical shock to the space distribution in the EU–8.

Third, the temporal variation of the city-size distribution and the eventual impact of accession to the EU on it were tested through the ARDL bound test approach. The result of the Bound test confirms the existence of a significant long-term relationship between the space distribution and the convergence parameters. As the EU–8 converges to the EU–15 level, this feeds agglomeration forces over spreading forces. The whole of the estimated parameters account for a variation of 0.021 point in the city-size distribution in the EU–8 and explains 64% of the cases. Considering the trend in the β value, this result may be interpreted as the contribution of the economic integration process within the EU to agglomeration over dispersion in the long-term. Additionally, the long-term model estimation indicates that the EU-integration process makes the city-size distribution more uneven for the EU–8.

Fourth, the short-term dynamics between the space distribution and the convergence variables are estimated using VECM. The short-term model estimation indicates that EU-integration has a weak but positive effect on the agglomeration forces in the EU–8. The negative and significant error correction coefficient (–0.25) is theoretically correct and shows the speed of annual adjustment from short-term to long-term equilibrium.

Fifth, the direction of causality is revealed by the Granger-causality tests. GCT mostly confirms the causality direction from the convergence parameters (except for price level) toward the space distribution revealed in the long-term and the short-term models.

Finally, the study concludes that the EU-integration process, causing a structural break to the city-size distribution which gradually became more uneven during the period, feeds agglomeration forces over spreading forces in the EU–8.

References

Anderson G. and Ge Y. (2005), *The size distribution of Chinese cities*, 'Regional and Urban Economics', No. 35.

Benguigui L. and Lieberhal E.F. (2007), *Beyond the power law- a new approach to analyses city size distributions*, 'Environment and Urban Systems', No. 31.

Bosker M., Brakman S. Garretsen H., Schramm M. (2008), *A century of shocks: the evolution of the German city size distribution 1925–1999*, Regional Science and Urban Economics, No. 38(4).

Brakman S., Garretsen H., and van Marrewijk C. (2001), *An Introduction to Geographical Economics,* Cambridge, UK: Cambridge University Press (Third edition, 2005).

Combes P.P. and Overman H. (2003), *The Spatial Distribution of Economic Activities in the European Union*, [in:] edited by V. Henderson and J.F. *Thisse Handbook of Regional and Urban Economics* Volume IV .

Davis D.R. and Weinstein D.E. (2002), *Bones, bombs and breakpoints: the geography of economic activity*, 'American Economic Review', No. 92.

Dupuch S., Jennequin H. and Mouhoud E.M. (2001), *Intégration européenne, élargissement aux PECO et économie géographique*, Région et Développement, No. 13.

Fujita M., Krugman P. and Vanables A. (1999), *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge MA: MIT Press.

Hirschman A. (1958), *Strategy of Economic Development*, Yale University Press, New Haven.

Kooij P. (1988), *Peripheral cities and their regions in the Dutch urban system until 1900*, 'Journal of Economic History', No. 48.

Krugman P. (1993), *First nature, second nature and metropolitan location*, 'Journal of Regional Science', No. 33.

Krugman P. (1995), *Development, Geography, and Economic Theory*, MIT Press.

Krugman P. (1998), *What's New about the New Economic Geography?*, 'Oxford Review of Economic Policy', No. 14(2).

Myrdal G. (1957), *Economic Theory and Underdeveloped Regions*, Duckworth, London.

OECD (2012), Redefining Urban: A New Way to Measure Metropolitan Areas, OECD Publishing, Paris.

Perron P. (1997), *Further evidence on breaking trend functions in macroeconomic variables*, 'Journal of Econometrics', No. 80.

Redding S. and Sturm D.M. (2005), *Cost of remoteness: evidence from German division and reunification*, CERP discussion paper, Vol. 5015.

Rosen K.T. and Resnik M. (1980), *The size distribution of cities: An examination of Pareto Law and Primacy*, 'Journal of Urban Economics', No. 8.

So K.T. (2005), *Zipf's law for cities: a cross-country investigation*, 'Regional Sciences and Urban Economics', No. 35.

Zipf G.K. (1949), *Human Behavior and the Principle of Least Effort*, Addison-Wesley Press.

Zivot E. and Andrews D., (1992), *Further evidence of great crash, the oil price shock and unit root hypothesis*, 'Journal of Business and Economic Statistics', No. 10.

Weber A. (1909), *Theory of the Location of Industries, translated in 1929 by Carl J. Friedrich*, Chicago: The University of Chicago Press.

Streszczenie

CZY INTEGRACJA GOSPODARCZA STANOWI HISTORYCZNY SZOK DLA ROZKŁADU WIELKOŚCI MIAST?

Opierając się na założeniu, że proces integracji gospodarczej wpływa, poprzez reformy rynkowe, na dynamikę procesu kształtowania przestrzeni w krajach kandydujących, w niniejszym opracowaniu podjęto próbę udzielenia odpowiedzi na pytanie (i) czy dominujące są siły aglomeracyjne czy rozpraszające (ii) czy integracja z UE powoduje przerwy strukturalne w rozkładzie przestrzeni w czasie; (iii) czy integracja z UE powoduje bardziej równomierny czy nierównomierny rozkład wielkości miast w ośmiu wschodnich państwach członkowskich Unii Europejskiej (UE–8). W toku analizy wykorzystano test Zivota-Andrewsa i test CUSUMSQ w celu zidentyfikowania przerw strukturalnych; test ARDL Bound posłużył do pokazania związku między równowagą długookresową i krótkookresową; test Grangera posłużył do określenia kierunku przyczynowości między zmiennymi. Główne wnioski z analizy: integracja z UE (i) spowodowała przerwę strukturalną w istniejącym rozkładzie wielkości miast, (ii) powiększyła nierównomierność rozkładu wielkości miast i (iii) spowodowała przewagę sił aglomeracyjnych nad siłami rozpraszającymi w państwach UE –8.

Słowa kluczowe*: rozkład wielkości miasta, reguła wielkości-kolejności, Prawo Zipfa, integracja z UE, nowe państwa członkowskie i państwa kandydujące, ARDL, bound test, test przyczynowości Grangera*