
NEW STATISTICAL APPROACHES IN THE SYSTEMIC ANALYSIS OF REGIONAL, INTRA-REGIONAL AND CROSS-REGIONAL FACTORS OF INFORMATION SOCIETY AND ECONOMIC DEVELOPMENT; THE CASE OF MAZOVIA

Jan Grzegorek
Andrzej P. Wierzbicki

ABSTRACT

The paper starts with the issue of estimating duality gaps – the differences in development measured in years between diverse districts (powiat in Polish) – in Mazovia, taking the number of computer laboratories per primary or secondary school as a basic index of the development of information society. This is based on a classical statistical approach of estimating the parameters of logistic curves. It turns out that these gaps amount up to 16-17 years – a very large differences in development even between neighboring districts. Then the paper turns to the issue of degree of causality of relations between diverse socio-economic indicators. It turns out that we can adapt a technique used typically in telecommunication signal analysis – but not in socio-economic analysis – for answering questions such as: do we earn money first and then buy computers and the Internet access, or vice versa, we buy computers and the Internet access first and then earn more money? We address this question on the level of voivodeships in Poland on two examples: the mentioned above example of causality relation of percentage of households having access to the Internet and GDP/capita percentage growth, and the example of causality relation of differences in GDP/capita and migrations between diverse voivodships of Poland. Both these themes illustrate the possibility of treating the Mazovian region as a system of sub-regions (districts) or entire Poland as a system of regions. They also contribute to the general theme “Quo Vadis Polonia” on the example of Mazovian region.

Introduction: research questions and methods

Poland is now one of the last European Union countries when it concerns information society development, which is a significant deterioration of its position during last 10 years [see e.g. Boguszewicz 2009]. There are significant macro-factors impeding this development, to which we return in conclusions; but an important question is to understand the details of information society development in micro-scale, smaller even than regional, on the level of districts (powiatów). There are two essential, two technical and two supplemental research questions that we want to address in this paper.

One of essential questions is the intra-regional diversity of the development of information society factors; we want to address this question on the example of the Mazovian districts, since the Mazovia region in Poland exhibits large duality gaps. A related technical question is the choice of indicators of information society development and of a method of their comparison or prediction; there are several types of indicators, such as the percentage of households having computers at home and those having access to Internet, but we selected a basic indicator of the number of computer laboratories per school. We decided to fit statistically historical data on these indicators in specific districts into logistic curves, with a comparison to similar logistic curves for entire Poland, which results in *advance or delay indicators*, i.e., the number of years of advance (or delay) when compared to the averages of Poland. On the level of districts, such comparisons are most basic and exhibit very large *duality gaps* (sums of maximal advance and delay). On the level of voivodeships, also percentages of households with access to the Internet or having computers at home are available, we can use the data to make also logistic curves for them, but the point is that averaging over large sets of data tends to hide duality gaps that are exhibited only in sufficient detail on the level of districts; on the level of voivodeships, the duality gaps are much smaller.

The second essential research question is the possibility of establishing a causal link between the indicators of information society development and main factors of economic growth, such as increase of GDP per capita. This can be established on the level of voivodeships only, since we need a larger statistical base. The related technical question is the method of analyzing the degree of causality. We selected a novel approach to this question, resulting from statistical analysis of causal relationship of signals in telecommunications, for which purpose auto-correlation function and cross-correlation function of input and output signals are used. For telecommunication signals, the causal relation is clear and the cross-correlation function is non-zero only for positive time differences (output signal can be influenced only by former values of input signal). For general socio-economic indicators, however, various causal relations, including positive or negative feedback loops, can occur; for example, people can buy computers because their income (GDP/capita) increased, or their income increases because they have bought computers. Obviously, both relations occur; the question is which of them is stronger. For answering this question, we propose to compare maximal values of cross-correlation function between such two socio-economic indicators for positive and negative time differences. For example, if the value of such cross-correlation function is maximal for a positive time difference, it means that the former of two indicators has a greater causal impact on the latter indicator than vice-versa. The supplemental research question is the relation of the results of analyzing such causal link and the public policy of supporting information society development in Poland. Because we find that the causal impact of buying computers first on the future income increases is stronger, we conclude that the delay of information society development in Poland results from a too passive public policy in that respect. Somewhat simplifying, we can say that European Union perceived the need of actively supporting the development of information society even in private enterprises, Polish authorities believe that the market will take care of that and it is sufficient to computerize administration.

This research is not done in isolation: we are aware of many publications on related issues, such as (Strużak 2009). However, the novelty of our approach is to analyze intra-regional duality factors defined as differences of developmental delay factors, and to analyze degree of causality of a feedback relation between diverse developmental factors using a cross-correlation function approach. We try also to analyze other causal links, such as the causal link of migration to income differentials. In general, we can say that the novelty of our approach is related to analyzing regional relations in a systemic way, that is, treating Poland as a system of regions and Mazovian region as a system of districts.

Delays in the development of information society in diverse districts of Mazovia

Between diverse indexes of the development of information society, we selected first the most basic one: *the equipment of primary and secondary schools in computer laboratories*, measured by the number of such laboratories per school. This index changes between zero and two, because one computer laboratory is typically not sufficient for a school; some schools might have even more than two computer laboratories, but these are exceptional. This is a very basic index, because information society starts at schools; without education in computer and the Internet use, no development of information society is possible.

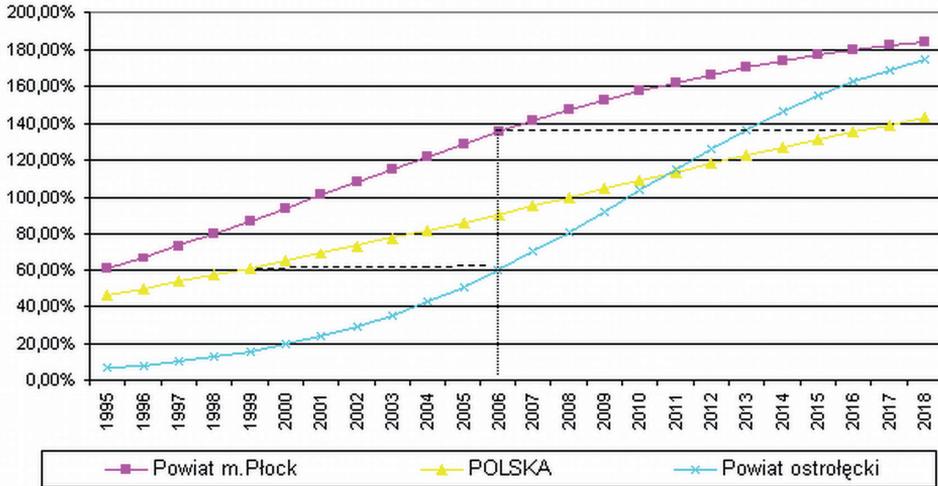
We used a specific method of analyzing such indexes. The data reporting such indexes for a number of years (about seven years of such data are available at http://www.stat.gov.pl/bdr_n/app/wybrane_cechy.wymiary) were fitted to a logistic curve:

$$y = \frac{a}{1 + b \exp(-tc)} ; a, b, c \geq 0 \quad (1)$$

where the parameter $a = 2$ was assumed and parameters b, c were fitted statistically (statistical fit for a gives much less significant results). This resulted in logistic curves as shown in Fig. 1 a, b, c on which we can determine – either from the actual or extrapolated data – the advance or delay times between the development of a given voivodeship or a given district and the average of Poland (actually, 0.9 computer labs per school; the average of Mazovia region is approximately the same) in 2006.

As we can see from Fig. 1, there might be several ways of determining advance and delay times: either on a common level of the average for entire region or country in a base year (2006), or on the separate levels of the development of districts in the base year (as indicated in Fig. 1). We selected the second way, which gives slightly larger estimates of the advance and delay times; the detailed comparison of those diverse ways shall be a subject of a separate study. With the method chosen, the maximal advancement of 10 years ahead of average of Poland in Mazovia region occurs at the city of Płock (which is a separate though small district; but the district around Płock is about 11 years behind the city); the maximal

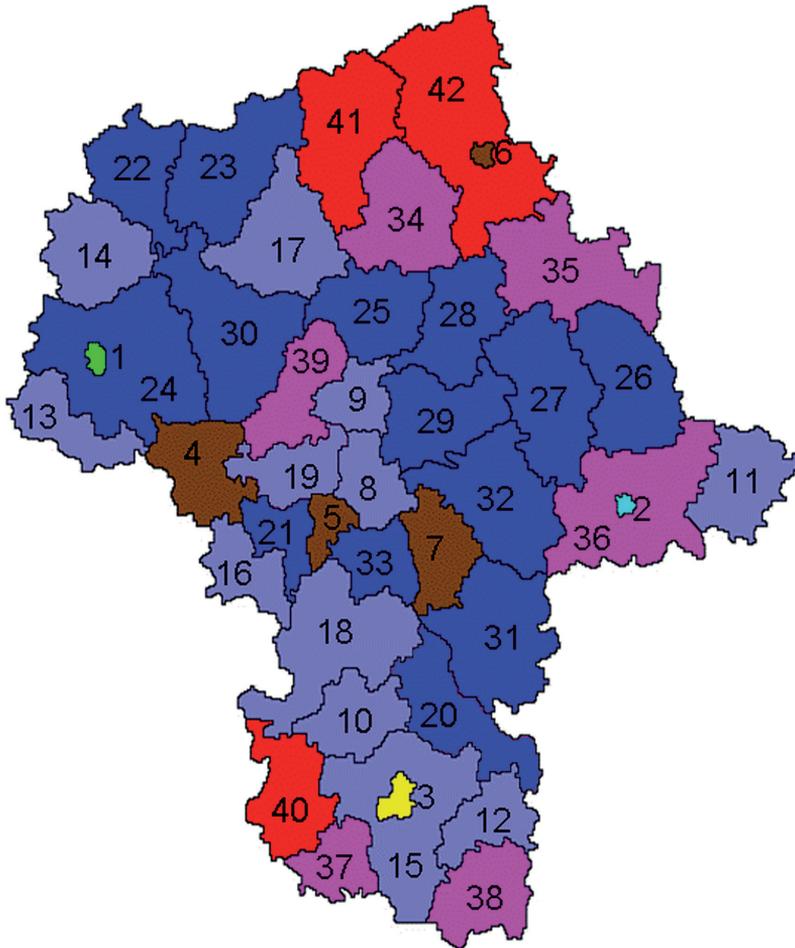
Fig. 1. The logistic curves for the maximal advanced district in Mazovia (district city of Płock, curve a), average of Poland (curve b) and maximal delayed district in Mazovia (district around Ostrołęka, curve c)



delay of 6-7 years behind the average of Poland in Mazovia region occurs in district around Ostrołęka (while the city of Ostrołęka alone is 3 years ahead of Polish average). Thus, *maximal duality gap* in Mazovia region – the sum of (absolute values of) maximal advancement and maximal delay – is tremendous, exceeds 16 years, which confirms the intuitive opinions about large duality gaps in Mazovia. Actually, not all intuitive judgments are confirmed by the analysis; the capital city of Warsaw is only 2 years ahead of Polish average, much more ahead are not only the city of Płock, but also several other city districts: the city of Siedlce (but the district around Siedlce is about 8 years behind the city), the cities of Radom and of Pruszków. This is illustrated graphically in the map presented in Fig. 2, together with a table of codes used in the map.

The same type of analysis performed on the level of voivodeships would give results distorted by averaging – correct, but overseeing much of local duality exhibited on the level of districts and averaged out in voivodeships. The map presented in Fig. 3 illustrates such results for all voivodeships of Poland; we see that the duality gap on the level of voivodeships is rather small, around 3 years (while the average of Poland is about 3-4 years behind the average of European Union in most indexes of information society development). However, this indicates only that we can oversee duality gaps when averaging over large sets.

Fig. 2. The advancements and delays in the equipment of schools of diverse districts in Mazovia in computer laboratories when compared to Poland averages in 2006 (numbers indicate places in ranking)

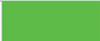


8 – 10 (years)	Districts: city of Płock (1)	
6 – 8	Districts: city of Siedlce (2);	
4 – 6	Districts: city of Radom (3);	
2 – 4	Districts: sochaczewski (4); pruszkowski (5); city Ostrołęka (6); otwocki (7);	
0 – 2	Districts: city of Warsaw (8); legionowski (9); białobrzeski (10); łosicki (11); zwoleniński (12); gostyniński (13); sierpecki (14); radomski (15); żyrardowski (16); ciechanowski (17); grójcecki (18); warszawski zachodni (19);	
-2 – 0	Districts: kozienicki (20); grodzki (21); żuromiński (22); mławski (23); płocki (24); pułtuski (25); sokołowski (26); węgrowski (27); wyszkowski (28); wołomiński (29); płoński (30); garwoliński (31); miński (32); piaseczyński (33);	
-4 - -2	Districts: makowski (34); ostrowski (35); siedlecki (36); szydlowiecki (37); lipski (38); nowodworski (39)	
-7 – -4	Districts: przysuski (40); przasnyski (41); ostrołęcki (42)	

Data based on http://www.stat.gov.pl/bdr_n/app/wybrane_cechy.wymiary

Fig. 3. The advancements and delays of voivodeships in equipment of schools in computer laboratories to the when compared to Poland averages in 2006



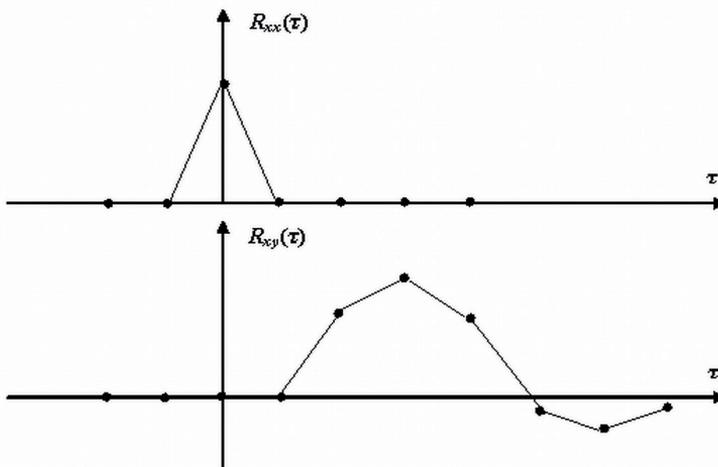
≈ +1.0	Voivodeships: Dolnośląskie, Lubuskie, Łódzkie, Opolskie, Podlaskie, Pomorskie, Śląskie	
≈ +0.5	Voivodeship: Kujawsko-Pomorskie	
≈ 0.0	Voivodeships: Mazowieckie, Warmińsko-Mazurskie, Wielkopolskie, Zachodnio-Pomorskie	
≈ -1.0	Voivodeships: Lubelskie, Podkarpackie, Świętokrzyskie	
≈ -1.5	Voivodeship: Małopolskie	

Causal links between socio-economic indicators: should we first buy computers or first earn more?

The second essential research question is the possibility of establishing a causal link between diverse socio-economic indicators. Usually, such indicators are somewhat related and correlated statistically, but we cannot derive conclusions about causal relations just from statistical correlation. There are many classical examples of the futility of establishing causal relations from statistical data, e.g. the example of correlation between the intensity

of the spots on the Sun and that of electromagnetic storms on Earth; both have a common cause, storms on the Sun, but this knowledge does not result just from the statistical observations. However, there is the issue of defining at least the degree of causality relation between two statistical indicators. For this, we can adapt a method used until now only in telecommunication signal analysis, called there auto-correlation function and cross-correlation function analysis of input and output signals in a transmission line. For such a case, the causal relation is clear and the cross-correlation function is non-zero only for positive time differences (output signal can be influenced only by former values of input signal), see Fig. 4.

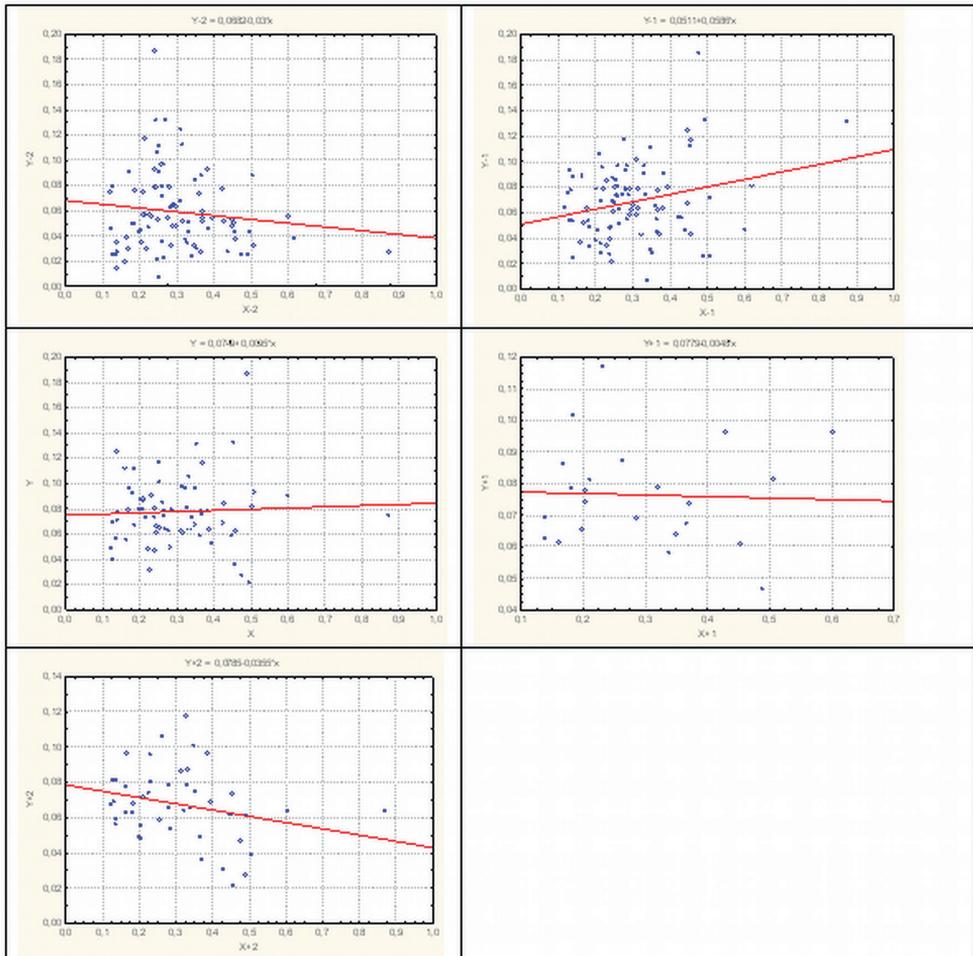
Fig. 4. A typical auto-correlation function $R_{xx}(\tau)$ (for an input signal $x(t)$ correlated with $x(t+\tau)$, of white noise type, without internal correlation) and related cross-correlation function $R_{xy}(\tau)$ (between the input signal $x(t)$ and the output signal $y(t+\tau)$, treated as a function of the time difference τ) for signals in a telecommunication transmission line



However, we can extend this type of reasoning for general socio-economic indicators, where various causal relations, including positive or negative feedback loops, can occur. We can compare maximal values of cross-correlation function between such two socio-economic indicators for positive and negative time differences τ . If the value of such cross-correlation function is maximal for a positive time difference, it means that the former of two indicators has a greater causal impact on the latter indicator than vice-versa. This does not mean that there is a single, clear cut casual relation as between the input and output signals in a transmission line, but we can use the maximal value of the cross-correlation function for positive time differences τ as a measure of the strength of partial causal relation between the former signal or indicator and the latter signal or indicator $y(t+\tau)$.

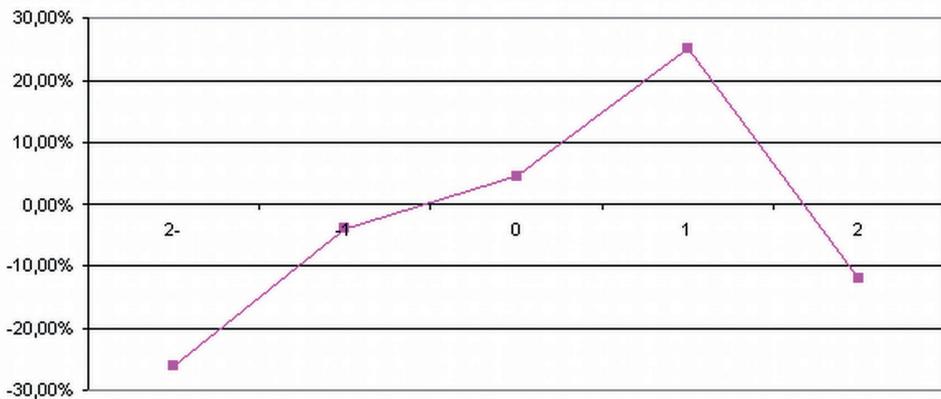
We use this type of analysis first to answer the question: do people buy computers and Internet access because their income (GDP/capita) has increased, or vice-versa, does their income increase because they have bought computers and Internet access? Obviously, both relations occur; the question is which of them is stronger. For answering this question, we use data of the percentage of homesteads with Internet access as the indicator $x(t)$ and of GDP/capita percentage increases as the indicator $y(t+\tau)$, with the statistical base of all voivodeships of Poland and last seven years. The statistical base is illustrated by the points in Fig. 4 a, b, c, d, e (for $\tau = -2, -1, 0, 1, 2$); the resulting cross-correlation function is presented in Fig. 5.

Fig. 5. The statistical base and the correlation trends in relations of percentage Internet access at homesteads $x(t)$ and the percentage increase of GDP/capita $y(t+\tau)$ in all voivodeships and the years 2002-2008 in Poland



The cross-correlation function in Fig. 5 indicates that the correlation between Internet access and per capita GDP growth is not very large (after all, many other factors influence the growth of per capita income), but significantly largest for one year delay: we do buy Internet access first and then our income is growing. This might be not a revolutionary conclusion – after all, the Internet access is an important investment – but it indicates that public policy should take this fact into account, providing even stronger incentives than existing ones for such investments.

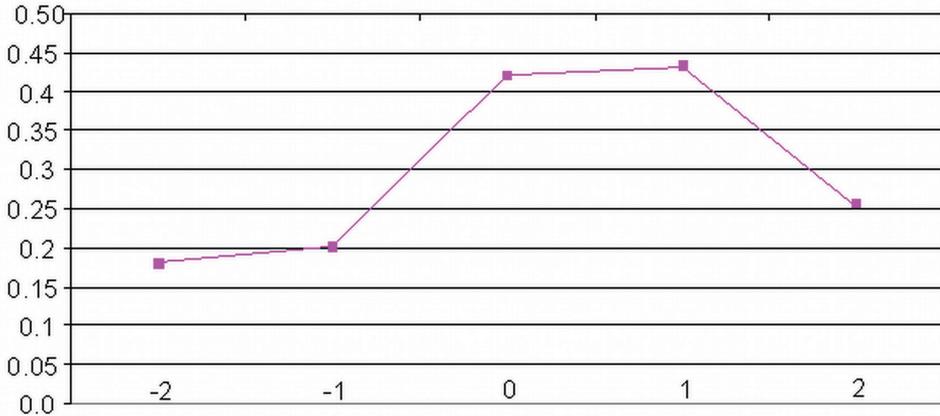
Fig. 6. The cross-correlation function $R_{xy}(\tau)$ of $y(t+\tau)$ (per capita GDP percentage growth) and $x(t)$ (percentage of homesteads with Internet access) depending on the time shift τ of data



Another question are the reasons for cross-regional migrations. We can determine migrations between all voivodeships in Poland for several years, which gives rather large statistical base. We correlated them – also with advancements and delays, to compute cross-correlation function – with cross-regional differences of GDP per capita. The results are presented in Fig. 6.

We see from Fig. 6 that the differences in GDP per capita do indeed have a rather strong impact on cross-regional migrations and vice versa. The impact of the current year and last year differences in GDP per capita motivates the migrations most strongly (correlation coefficients over 0.4), while last year and two years before migrations have also positive impact on per capita growth in the region being migration target (correlation coefficients close to 0.2). This confirms the intuitive judgment that new immigrants contribute positively to the economic growth.

Fig. 7. The cross-correlation function $R_{xy}(\tau)$ of the cross-regional differences in GDP per capita as $x(t)$ and cross-regional migrations as $y(t+\tau)$



Conclusions: is public policy supporting information society in Poland too passive?

The main conclusion of the paper is that looking upon Mazovia as a system of districts, or on Poland as a system of voivodeships, and applying advanced statistical analysis techniques results in interesting results that might be useful for regional development strategies and policies. We are thankful to Prof. Antoni Kukliński for suggesting such systemic approach to the question Quo Vadis Polonia.

However, there are also critical conclusions concerning public policy towards information society issues. Not only is entire Poland lagging behind almost all European Union countries in most indexes of information society development, but inside Poland – as exemplified by the case of Mazovia – we observe very large duality gaps, districts that are far ahead of Polish or even European average and districts that are many years behind these averages. That indicates that public policy towards the issues of development of information society might be just too passive in Poland, leaving too much to the market and local initiatives – which are extremely important, but tend to increase duality gaps beyond reasonable differences.

There are many possible future research topics resulting from our analysis. One is analyzing diverse ways of defining delay or advance times from statistical data. Another is auto-correlation analysis of economic and social data. Still another is combining passively obtained statistical data with actively obtained data from surveys; for example, opinions of children from schools in the analyzed districts might be an important information component for forming regional policies, and performing such surveys should be promoted.

BIBLIOGRAPHY

- Boguszewicz T., 2009, *Polska maruderem Internetu* (in Polish: Poland as an outsider of Internet). "Rzeczpospolita", March 26, 2009, page 1.
- COCOM08-41 FINAL, 2008, *Broadband access in the EU: situation at 1 July 2008 Brussels*, November 2008, DG INFSO/B3.
- Codagnone C., 2009, *Modelling and demonstrating the impact of eInclusion.*, Vienna Study on Inclusive Innovation for Growth and Cohesion: March 2009.
- Eurostat Statistical Books, 2008, *Eurostat yearbook 2008. Europe in figures*.
- Fickinger S., Lumio M., 2008, *Eurostat Population and Social Conditions*, Data in "Focus" n 12/2008
- IŁ, 2005, *Firmowi użytkownicy Internetu w Polsce Raport z badań*, Gdańsk, grudzień 2005.
- ITU, 2008, *Partnership on Measuring ICT for Development ITU Manual on ICT Household Statistics*, www.itu.int/ict September 2008.
- ITU, 2009, *Measuring the Information society - ICT Development Index*.
- Hales C., 2006, *Rola ITC w rozwoju społeczno-gospodarczym*, Zeszyty Naukowe SCENO n 2/2006.
- Löf A., 2008, *Science and technology, population and social conditions: Industry, trade and services*, Data in "Focus" 46/2008.
- Science, Technology and Innovation in Europe, 2008, *Eurostat Pocketbooks 2008 edition*.
- Strużak R., 2009, *Rozwój szerokopasmowego Internetu w Polsce*, (in print), *Telekomunikacja i Techniki Informacyjne*.
- United Nations, 2008, *Partnership o Measuring ICT for Development*, Core ICT Indicators.
- Wierzbicki A. P., 2009, *Delays in Technology Development: Their Impact on the Issues of Determinism, Autonomy and Controllability of Technology*, "Journal of Telecommunications and Information Technology", n 4:1-12, 2008.
- Wilén H., 2007, *Statistics in focus*, "Science and Technology" n 120/2007.
- World Information Society Report, 2007, *Information Society Statistics*.

STRESZCZENIE

Artykuł rozpoczyna się przedstawieniem kwestii oceny tzw. luki dualnej, a więc zróżnicowań rozwojowych między różnymi powiatami Mazowsza przy wykorzystaniu wskaźnika liczby laboratoriów komputerowych w szkołach podstawowych i średnich, który służy do pomiaru stopnia rozwoju społeczeństwa informacyjnego. Jest to klasyczne podejście statystyczne oceny parametrów krzywych logistycznych. Okazuje się, że te zróżnicowania są bardzo duże, nawet jeśli chodzi o sąsiadujące ze sobą powiaty. Kolejna część artykułu poświęcona jest związkowi przyczynowo-skutkowym między różnymi wskaźnikami społeczno-ekonomicznymi. Okazuje się, że można w badaniach zastosować techniki używane w analizach sygnałów telekomunikacyjnych, by odpowiedzieć na pytania, czy najpierw musimy zarobić pieniądze, by kupić komputer i dostęp do internetu, czy też na odwrót: kupujemy najpierw komputery i dostęp do internetu, by zarabiać pieniądze? W artykule odpowiada się na to pytanie, analizując sytuację województw w dwóch ujęciach: związków między liczbą gospodarstw domowych, które mają dostęp do internetu a wzrostem PKB per capita oraz związków między zróżnicowanym PKB per capita i migracjami między województwami w Polsce. Te dwa ujęcia umożliwiają potraktowanie Mazowsza jako systemu subregionów (powiatów) albo całej Polski jako systemu regionów. Ujęcia te są także wkładem w dyskusję nad głównym problemem *Quo Vadis Polonia* na przykładzie Mazowsza.

Jan Grzegorek, ekspert ds. wsparcia operacji Telekomunikacja Polska SA. Specjalizacja: zastosowanie nowych metod statystycznych do analizy wskaźników rozwoju społeczeństwa informacyjnego w Polsce traktowanej jako system regionów.

prof. dr hab. inż. Andrzej P. Wierzbicki, pracuje w Instytucie Łączności (Państwowy Instytut Badawczy) – Zakład Zaawansowanych Technik Informacyjnych, zastępca przewodniczącego Komitetu Prognoz „Polska 2000 Plus” przy Prezydium Polskiej Akademii Nauk, zastępca przewodniczącego Zespołu Polskiej Akademii Nauk d.s. Współpracy z Międzynarodowym Instytutem Stosowanej Analizy Systemowej w Luxemburgu k. Wiednia. Specjalności: telekomunikacja i automatyka, optymalizacja, wspomaganie decyzji, analiza systemowa, inżynieria wiedzy.