TOMASZ NIEDZIELSKI

WROCŁAW UNIVERSITY, POLISH ACADEMY OF SCIENCE

POPULATION DYNAMICS IN LARGE POLISH CITIES SINCE 1980: A STATICTICAL APPROACH

ABSTRACT. The analysis of population dynamics in large Polish cities in the period 1980-2001 is performed by means of statistics. Decomposition of time series into deterministic and stochastic components is applied to datasets (number of inhabitants). In order to deal with trends unlike each other the differential index of population dynamics (DIPD) is introduced. A spatial organization of cities revealing the great dynamics of population changes is detected. Seeking dependency of residuals is performed by means of the Ljung-Box test and the sample autocorrelation function. It is discovered that five of six cities exhibiting the high DIPD reveal dependent autoregressive residuals. The great dynamics of population changes may be driven by dependent autoregressive demographic processes.

KEY WORDS: population dynamics, large city, time series, statistics, Poland.

INTRODUCTION

The objective of this paper is to study dynamics of population changes in large Polish cities in the period 1980–2001 by means of time series theory. The study is based upon a differential analysis of the deterministic trend functions and a stochastic analysis of residuals. The problem of spatio-temporal analysis of demographic changes in Poland is of particular interest of researchers. Parysek and Kotus (1997), for instance, focused on the dynamics of development of the largest 60 Polish towns in the period 1950-1990. They proposed to group towns according to the size criteria and carried out the analysis in terms of this classification. Moreover, the authors stressed the diversified and turbulent properties of the growth rate curves (Parysek and Kotus, 1995) which indicates a need for seeking indices of population dynamics covering the mentioned diversity. Herein, we aim to find the differential criteria coping with the description of dynamics of population changes described by deterministic trend functions which are elements of the classical decomposition of the input time series. The need for taking into consideration a time series approach in the analysis of population changes theory is also emphasized by Parysek and Kotus (1995). In this article we use: (1) a trend analysis applied here in order to detect a spatial organization of cities exhibiting a great dynamics of population changes and (2) the time series methods to recognize the correspondence between the dynamics of population changes and the stochastic structure of residuals.

As an indicator of population changes the number of inhabitants has been chosen. The data analysis corresponding to this variable is based upon the decomposition of the time series into deterministic components (trend and seasonality) and residuals (sample path of a certain discrete stochastic process) (e.g. Brockwell and Davis, 1996). Thus, in process of studying we face deterministic functions which describe general trajectories of the analysed population changes. However, the data consist of many fluctuations which may indicate the random nature of the described phenomena. In this paper, trends as well as residuals are therefore taken into consideration. On one hand, the analysis of trend functions based on the introduced differential index of population dynamics (DIPD) allows us to discover dynamics of population changes. On the other hand, the analysis of stochastic residuals carried out by means of autoregressive moving average processes (ARMA) provides the description of fluctuations. Therefore, empirical studies are combined with modelling what is, according to Burch (2003), typical of demography. He stresses a role of balance between a model-based approach and an empirical approach. This justifies our time series analysis which contains the two.

For the analysis the ten largest Polish cities (i.e. Bydgoszcz, Cracow, Gdańsk, Katowice, Lublin, Łódź, Poznań, Szczecin, Warsaw, and Wrocław) have been chosen. The number of inhabitants of each city exceeds 300 thousand. In five of ten cases (Warsaw, Łódż, Cracow, Wrocław, and Poznań) we face cities with population of over half of a million. The choice of the time period (namely 1980-2001) follows from the fact that it reflects the crucial political changes in Poland and thus allows us to incorporate into the analysis the effects of socio-political transformation. The stability of the administrative boarders of the considered city is confirmed. Thus, the analysed population changes are mainly of demographic origins.

METHODS

Herein, we consider M cities in respect of the number of inhabitants altering in time. The time space is discrete and may be denoted as $T=\{t_1,\ldots,t_n\}$, where t_i is a discrete-time moment of observation. The sampling intervals ought to be the same length. In case of necessity one has to derive lacking data by applying, for instance, a moving average filter or a linear interpolation. Therefore, the input data consist of M time series which are assumed *a priori* to be mutually (thus spatially) independent. As stated before, this study is based upon the analysis of trend functions and residuals. They are elements of the classical decomposition model which may be written down in the following way:

$$X_t = m_t + s_t + Y_t, \quad (1)$$

where X_t represents the input data, m_t is a trend component, s_t is a seasonal component, and Y_t is a random noise component (the trajectory of an unknown discrete stationary stochastic process). In this paper, we do not consider seasonal components because the data underlying the analysed variables do not exhibit periodicity. However, in case of the presence of seasonality, one has to remove the periodic component as well (see for details: Brockwell and Davis (1996)).

The first approach consists of seeking similarities or differences between trend components (m_t) of the time series corresponding to M cities. For each set of data there is a need to compute the trend function which fits the considered time series well. This may be performed, for instance, by applying the least squares procedure (e.g. Brockwell and Davis, 1996). It is possible to fit, for example, a linear or a quadratic trend. These are, of course, simple functions but, in turn, we aim to yield a simple model. The above-mentioned functions are continuous and theirs time-domain is a real line. Once the trends are derived one may compare results attained for different cities. This may be done by applying the differential index of population dynamics (DIPD) which is defined herein.

The DIPD is based upon the trend of the number of inhabitants of the analysed city altering in time. There is a need to compute a trend derivative $\mathrm{d}m_t/\mathrm{d}t$ in order to obtain the rate of population changes. Subsequently, one should derive its values for all discrete time points considered. As positive and negative numbers are likely to occur we propose to apply absolute values in order to cover the general dynamics of population changes caused by increasing as well as decreasing tendencies. Finally, the sum of the derived numbers gives a differential index of population dynamics (DIPD). Thus, the DIPD ought to be computed for each city considered. The lowest value of the DIPD indicates the lowest dynamics of the considered changes while the larges value of the DIPD corresponds to the largest dynamics. This allows us to compare population changes across the analysed M cities. This may lead to the analysis of spatial distribution of cities exhibiting the considerable dynamics of population changes in the period in question. It is worth noting that this procedure holds only for differentiable trend functions. One can label this part of an analysis as a deterministic one.

The second approach is based on stochastic processes. At the beginning of the procedure one needs to yield the residuals. This should be performed by removing the previously computed trend component and, if necessary, the seasonality from the non-transformed data. As we aim to compare residuals corresponding to M analysed cities, the uniformity in deriving a trend function becomes a necessary condition which has to be fulfilled. This means that after the analysis of graphs of M time series we have to choose the most appropriate set of trend functions which is to be used in the subsequent study. Thus, once the trend components are computed by means of the uniform method and are subtracted from the original data and, additionally, the average values are subtracted, the time series of residuals (Y_t) are obtained. However, we expect the residuals to exhibit stationarity because we aim to fit an ARMA model. Indeed, if one applies this procedure and does not produce a stationary trajectory of a stochastic process, estimation of the model may be disturbed. Assessing stationarity may be done by statistical testing (e.g. Phillips – Perron test) or by the analysis of graphs of the obtained residuals. The plots cannot consist of extreme values, and the variance should not be increasing or decreasing. Yielding stationary residuals (Y_t) allows us to carry out the subsequent stochastic approach.

Each case study consists of the analysis of residuals (Y_t) carried out in order to detect stochastic processes governing fluctuations. These results allow us to detect differences between residuals of the number of inhabitants of the considered cities. Indeed, seeking differences between demographic changes may be performed, for instance, in terms of studying autoregressive moving average processes (ARMA). This approach takes into consideration stochastic behaviours of stationary residuals. Hence, in this part of the study we do not focus on deterministic components of a process but, in turn, we study its random part. This implies that the direction of changes is uncertain in this part of the analysis. Fluctuations of demographic changes, on the other hand, are intrinsic to our stochastic approach because they may reflect random demographic processes governing population changes.

The objective of the subsequent study is to show that the considered cities may be classified according to ARMA models which are fitted to the respective residuals. For every realization of the stochastic process (Y_t) an ARMA model has to be approximated. The further analysis and hence classification requires grouping cities together if the respective residuals are described by an ARMA model of the same orders. The model selection is based on the classical methods of time series theory presented in the textbook of Brockwell and Davis (1996). We focus on estimation of the orders p and q for an ARMA model. If this is performed there is a need to group cities in which the residuals of the number of inhabitants are assumed to be governed by an ARMA model of identical orders (p and q, explicitely). As the theory is not an objective of this study we omit stating useful mathematical results and relate to Brockwell and Davis

(1996) for the probabilistic background of ARMA order selection. In summary, this may be performed, in the simplest way, by an analysis of the sample autocorrelation function, and the partial sample autocorrelation function. Additionally, we apply the Ljung-Box test in order to test for the iid (independent, identically distributed random variables) structure of the time series. Rejection of the iid hypothesis means that the considered time series exhibits dependence. Moreover, there exist additional criteria, e.g. FPE and AICC statistics (Brockwell and Davis, 1996) which support selection of the ARMA orders. Once the models are fitted the subsequent inference is case specific and, therefore, is included in the next section.

RESULTS AND DISCUSSION

This study focuses on ten cities thus, according to the previous notation, $M = 10$. We gain therefore ten time series. Each sample path of a stochastic process consists of twenty two numbers corresponding to the number of inhabitants in the period 1980-2001. The analysed data do not exhibit periodicity. The input data are taken from *Demographic Yearbooks of Poland.*

For the stochastic approach, residuals have been prepared by applying the uniform method which, according to the underlying data, is based upon removing a quadratic trend. This provides the satisfying goodness-of-fit in most cases (Fig. 1). Additionally, the mean of each time series has been subtracted. Thus, in process of preparing residuals we obtain the quadratic trend functions (Fig. 1) which are to be analysed firstly.

On one hand, we aim to compare the properties of the previously derived trend functions by means of the analysis of the DIPD. The results of computations are presented in Table 1. The time domain has been transformed in order to consider integers $\{1,2,\ldots,22\}$ instead of $\{1980,1981,\ldots,2001\}$.

Table 1. The DIPD in ten largest Polish cities.

City	DIPD	DIPD rank
Bydgoszcz	46,190	5
Cracow	44,396	6
Gdańsk	20,128	10
Katowice	40,935	7
Lublin	59,062	3
Łódź	83,468	2
Poznań	55,749	4
Szczecin	39,452	8
Warsaw	115,623	1
Wrocław	38,806	9

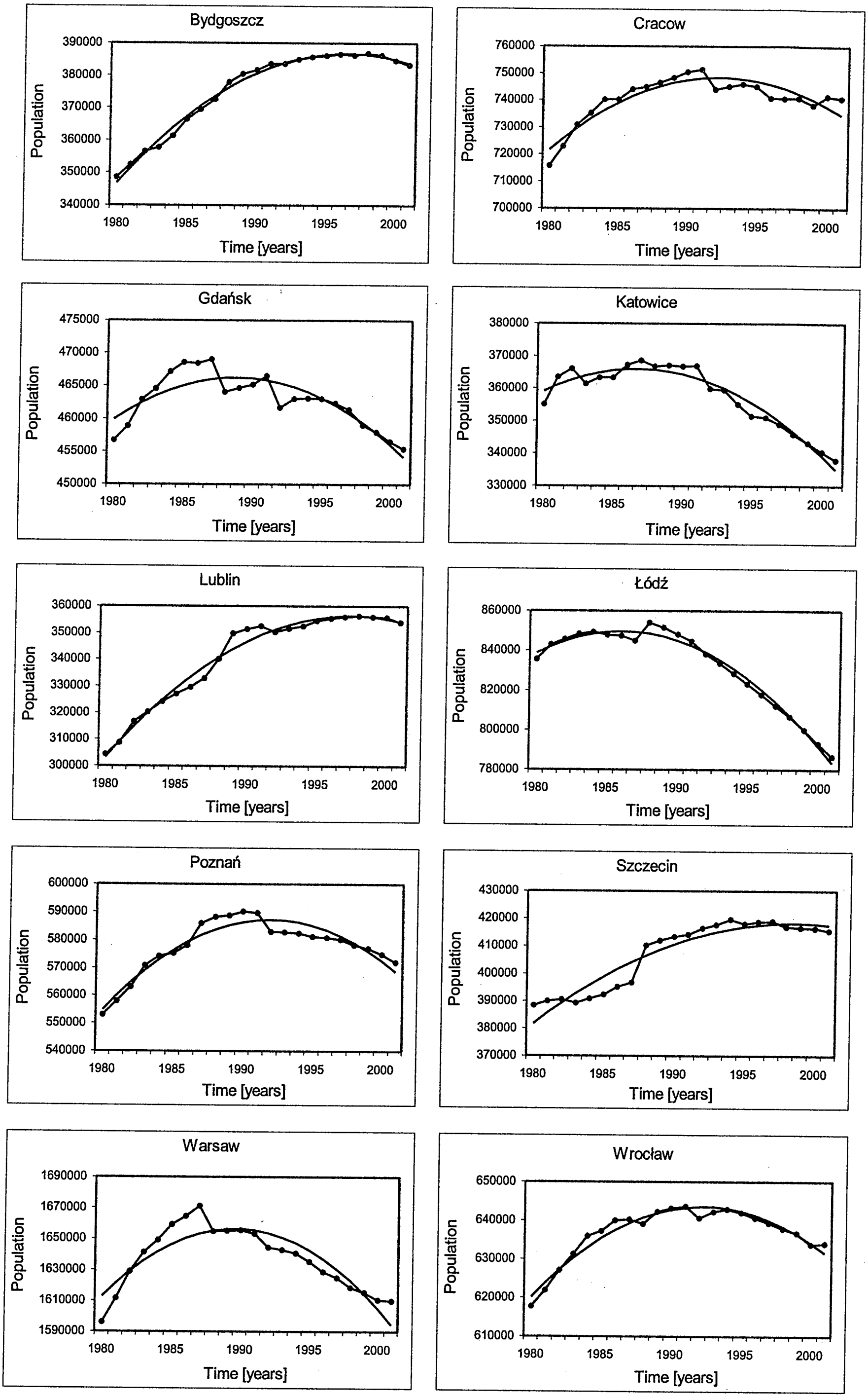

Fig. 1. The quadratic trends fitted by the least squares procedure.

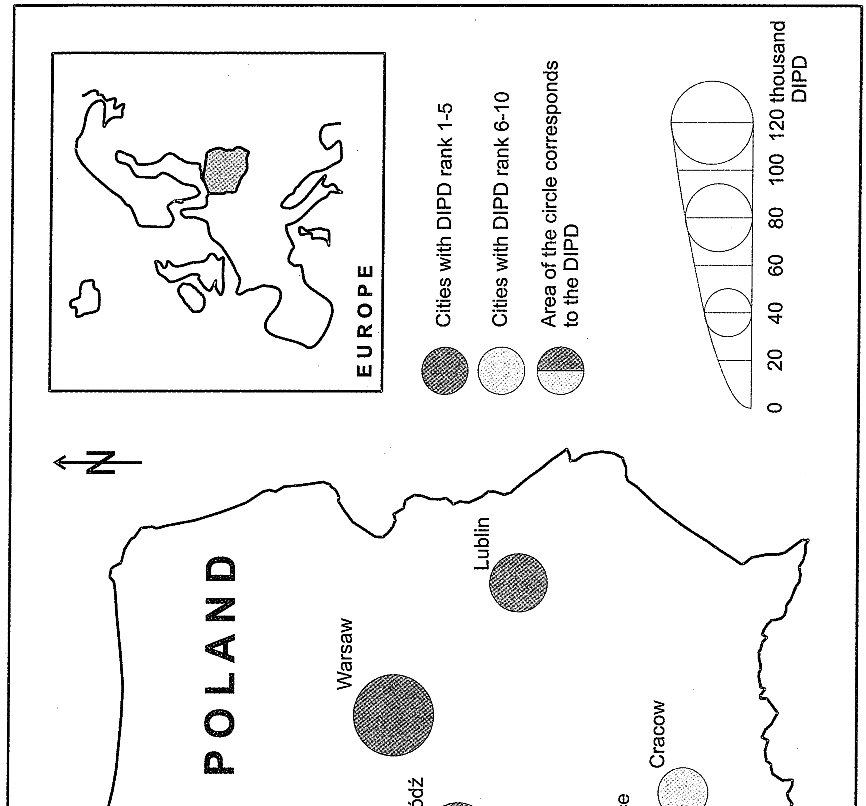

Fig. 2. The DIPD index derived for ten largest Polish cities. Reference map based on 'Geograficzny atlas świata', Warszawa, PPWK, 1997.

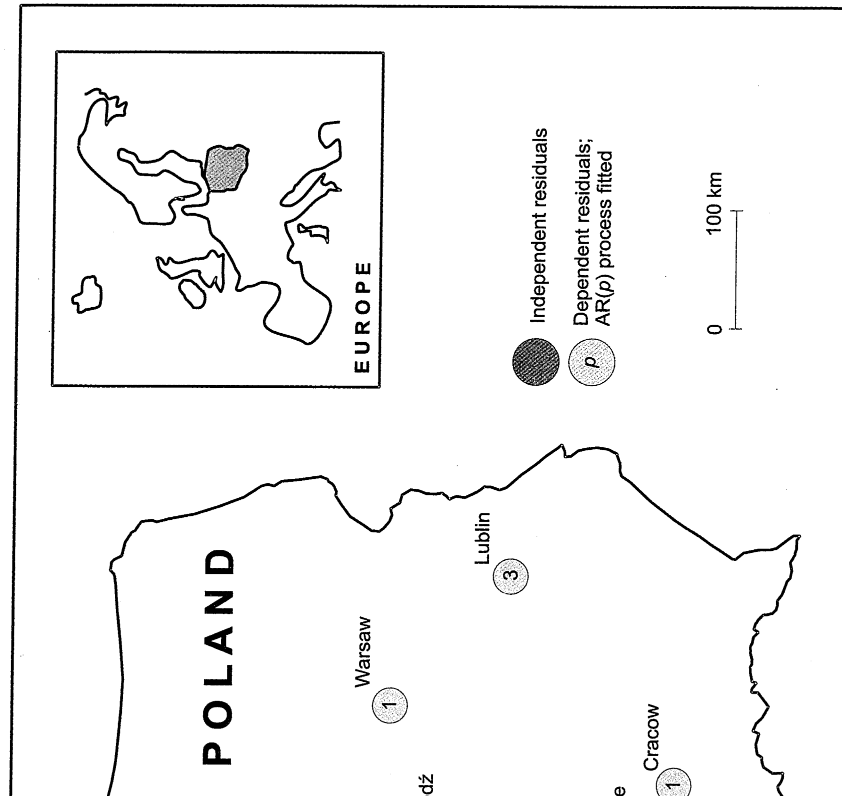

Fig. 3. AR models fitted ro residuals of the number of inhabitants for largest Polish cities. Reference map based on 'Geograficzny atlas świata', Warszawa, PPWK, 1997.

The DIPD analysis shows that the population dynamics of ten larges Polish cities considered independently in the period 1980-2001 revealed the spatial organization. Indeed, investigating cities with ranks 1-5 of the DIPD (the half of the considered sample) shows that the greatest dynamics of the population changes was associated with the geographical pattern: Poznań – Bydgoszcz – Łódź – Warsaw – Lublin (Fig. 2). The cities create the setup of the considerable and high population dynamics. This spatial distribution may be driven by different processes. In what follows, Warsaw and Poznań exhibit the considerable population dynamics due to the cumulated effect of (1) a fast increase in the number of inhabitants until early 1990s and (2) a subsequent population decrease (Fig. 1). According to Górecka and Kozieł (2004), this is associated with a low migration to large cities at the beginning of transformation period. This is due to the domestic economic constraints i.e. problems in the job market and in the housing market. The considerable dynamics of population changes in the city of Bydgoszcz and the city of Lublin is of other origins. They reveal the fast almost monotonic increase in the number of inhabitants over the entire period in question (Fig. 1).

This may be caused by the size of the two cities. This implies that there are no many relevant public problems (e.g. with transportation) which does not force people to move away. Moreover, Bydgoszcz and Lublin remain the regional centers and thus are still attractive places to live. On the other hand, the high dynamics of population changes in the city of Łódź is driven by a fast decrease in the number of inhabitants since middle of 1980s (Fig. 1). Thus, in this case the considerable dynamics is connected with declining trend in the domain of years 1986-2001. Górecka and Kozieł (2004), for instance, point out a collection of various negative demographic indices which explain declining trends of the number of inhabitants in the city of Łódź and the great dynamics of The stochastic approach is based upon ARMA order selections applied to the realizations of unknown discrete stochastic processes corresponding to ten cities. Firstly, in order to verify the hypothesis of independence we applied the Ljung-Box test to the considered time series (Table 2). The results have been combined with the analysis of the sample autocorrelation functions (ACF) by counting the numbers of autocorrelation values outside the confidence band (Brockwell and Davis, 1996). These two approaches allow us to infer dependent structure of residuals of population changes in the period in question in six cities considered. Indeed, residuals of the number of inhabitants of Cracow, Lublin, Warsaw, Bydgoszcz, Poznań and Szczecin have been assumed dependent. Secondly, we take a use of (1) the sample autocorrelation function (ACF) and (2) the sample partial autocorrelation function (PACF) in order to fit the ARMA models to each time series considered. The exponential decay of all ACFs says about the autoregressive structure of the dependent processes. Thus, the considered residuals in the above-mentioned cities may be fitted by an AR(p)

processes (p is the order of the autoregressive process). The analysis of PACFs lead to the following order selection: AR(1) – Cracow, Warsaw, Bydgoszcz, Poznań and Szczecin; AR(3) – Lublin (Fig. 3). In summary, this allows one to describe residuals by three classes of models: (1) independent and identically distributed random variables - iid, (2) autoregressive processes of order 1 - AR(1), and (3) autoregressive process of order 3 - AR(3).

Table 2. Noise sequence properties and stochastic models for residuals; the limiting value i.e. the 0.95 quantile of the chi-square distribution with 18 degrees of freedom (the distribution of Q statistics) is equal to 28.87; IID - independent and identically distributed random variables; AR(p) – p-th order autoregression.

City	Ljung-Box Q statistics	Amount of autocorrelation values outside the confidence interval	Dependence (D)/ /Independence (I)	Stochastic model
Bydgoszcz	34.76	3	D	AR(1)
Cracow	38.86	1	D	AR(1)
Gdańsk	27.68	1	I	IID
Katowice	16.99	0	I	IID
Lublin	36.88	3	D	AR(3)
Łódź	18.86	0	I	IID
Poznań	52.15	4	D	AR(1)
Szczecin	41.88	3	D	AR(1)
Warsaw	61.98	3	D	AR(1)
Wrocław	25.25	0	I	IID

An unexpected result is that the five of six cities which have revealed the greatest values of the DIPD do exhibit the dependent structure of the analysed residuals. Indeed, Cracow, Lublin, Warsaw, Bydgoszcz and Poznań yield the considerable or high DIPD values exhibiting simultaneously the autoregressive dependence of the respective residuals. The interpretation of this specific convergence may be the following. As the residuals are deviations from a given trajectory of population change (deterministic trend) they may reflect various random demographic processes (e.g. population growth and migrations). This assumption leads to the conclusion that the significant dynamics of population changes is associated with dependent demographic stochastic processes (hence, revealing a certain random mechanism). Furthermore, we have shown that the considered structures are purely autoregressive. This implies the certain rules of stochastic dependence and thus stochastic memory. Thus, the above-mentioned correspondence may lead to an interpretation that the high dynamics of population changes is caused by certain autoregressive demographic processes.

CONCLUSIONS

In this paper we analyse the population changes in large Polish cities in the period 1980-2001 by means of (1) the introduced herein differential index of population dynamics (DIPD) and (2) the stochastic analysis of the residuals of the considered time series. The first approach yields a ranking of cities created in terms of dynamics of population changes and a spatial organization of cities exhibiting the high and the considerable population dynamics. The second approach gives an unexpected correspondence between the great dynamics of population changes detected by the DIPD and the dependent autoregressive structure of residuals. According to our interpretation, the high dynamics may be caused by autoregressive demographic processes. The detection of these processes appears to be a crucial goal in further studies.

REFERENCES

Brockwell, P.J. and Davis R.A. 1996: *Introduction to time series and forecasting*, New York: Springer.
Burch, T.K. 2003: Demography in a new key: A theory of population theory, *Demographic Research* 9, p. 263-284.
Demographic Yearbook of Poland 1980-1989 and 1997-2001, Central Statistical Office, Poland, Warsaw.
Górecka, S. and Kozieł, R. 2004: Procesy demograficzne w największych miastach Polski w okresie transformacji. In Słodczyk, J., editor, *Przemiany demograficzne i jakość życia ludności miast*, Opole, p. 27-40.
Parysek, J.J. and Kotus J. 1995: Development patterns of Polish towns in the years 1950-1990, *Geographia Polonica* 66, p. 87-110.
Parysek, J.J. and Kotus, J. 1997: Powojenny rozwój miast polskich i ich rola w procesie urbanizacji kraju, *Przegląd Gegraficzny* 69, p. 33-54.
Statistical Yearbook – Demography 1990-1992, Central Statistical Office, Poland, Warsaw.
Statistical Yearbook of Demography 1993-1996, Central Statistical Office, Poland, Warsaw.

CORRESPONDENCE TO:

Tomasz Niedzielski, Institute of Geography and Regional Development, Wrocław University, Pl. Uniwersytecki 1, 50-137 Wrocław, Poland.
[e-mail: niedzielski@geom.uni.wroc.pl]

Space Research Centre, Polish Academy of Sciences,
Bartycka 8A, 00-719 Warsaw, Poland
[e-mail: niedzielski@cbk.waw.pl]