Application of Classification Trees in the Analysis of the Population Ageing Process

Justyna Wilk

Abstract A process of socio-economic development is continuously accompanied by a process of population ageing. In terms of a policy of regional development, it is valuable to identify factors of ageing to mitigate or impede its undesirable impact on a national economy. The paper discusses how to model the process of ageing using classification trees and presents an empirical study. The main research question is if the populations similar in their degrees of ageing, feature common demographic conditions of this process as well.

1 Introduction

The socio-economic development process is inseparably accompanied by the population ageing process (see Uhlenberg, 2009; Martinez-Fernandez et al, 2012). Population ageing relates to the changes in the age distribution and an increase in the percentage share of senior citizens in the general population. This results from reductions in mortality which are followed by reductions in the number of births. It is mostly determined by social, economic, cultural, environmental and other factors, such as: the intentional delay of procreation

---

Justyna Wilk
Wrocław University of Economics, Department of Econometrics and Computer Science, Nowowiejska 3 Street, 58-500 Jelenia Góra, Poland,
✉ justyna.wilk@ue.wroc.pl

ARCHIVES OF DATA SCIENCE, SERIES A — DOI 10.5445/KSP/1000058747/04
KIT SCIENTIFIC PUBLISHING — ISSN 2363-9881
Vol. 1, No. 1, S. 57–76, 2016

time and changing life priorities, leading more and more towards a healthy lifestyle, progress in medicine etc.

Population ageing strongly affects the situation of a country in terms of its financial, social and economic conditions. In the long-term prospect, without taking appropriate actions, it may lead to disturbing the retirement system, decreasing the efficiency of social systems and an increasing gap in the labour market etc. (see Magnus, 2008; Weil, 1997; United Nations, 2013; Légaré, 2006).

Population ageing is a natural process in economically well developed countries (see Prskawetz and Lindh, 2011; Lindh and Malmberg, 2009; Martinez-Fernandez et al, 2012). The most intensive process of population ageing is seen in Asian countries, especially in Japan, where one in four people is 65 years old or older (see World Population Prospects, 2012). This process also occurs in the majority of European Union (EU) countries of which Italy and Germany are demographically the oldest. Although the population of countries which joined the EU in 2004 and 2007 is relatively young (Slovakia, Poland and Cyprus are the youngest), they also suffer from ageing (see Muenz, 2007; Giannakouris, 2008; Grundy, 1996).

The diversification of demographic situations relates not only to the national economies but also occurs within countries. This results from socio-economic disparities and also cultural, social and environmental differences etc. Martinez-Fernandez et al (2012) see ageing as one of the most substantial factors of global demographic change and the shrinkage of cities and regions as well.

Regions differ in their nature of ageing; its intensity and demographic conditions, e.g. extremely low fertility, high out-migrations of young people, older and older workforce etc. Different situations of regions require different activities to mitigate the effects of ageing and also to impede this process. The identification of factors of ageing may be helpful in creating a policy of regional development.

This paper proposes using classification trees to model the population ageing process and presents an empirical study. The key research question is if regions similar in their intensity of ageing, feature common factors of this process as well.

2 Modelling the population ageing process

An examination of ageing covers its intensity and the demographic determinants. A lot of research studies are limited to examining the degree of ageing. This requires defining the threshold of ageing and determining the age structure of a population.

The threshold of ageing is the age which classifies a person as being older. Some statistical studies use the retirement age as the threshold of ageing, which is usually separate for men and women. Other research studies assume the age of 60 or 65 as threshold and the same threshold for both, men and women (see e.g. Sauvy, 1948). Sanderson and Scherbov (2008) see the age for which the remaining life expectancy is 15 years as the threshold of ageing.

Several researchers propose a lot of rates indicating the age structure of a population (see Clarke, 1965; Beaujeu-Garnier, 1966; United Nations, 2013). The first group includes age-related measures e.g. measures of location (e.g. median age), the old-age rate, ageing index, old-age dependency ratio. The second group covers the graphical tools such as, the population pyramid and Ossan's triangle etc.

Population ageing results from reductions in mortality and the number of births. At the national level, fertility and life expectancy are seen as the main determinants of the age structure (see Preston et al, 1989; United Nations, 2013). At the regional level, it may also be affected by massive out-migration of young people. The process of ageing is also intensified by the ageing of the adult population, if the share of older adult people is increasing. Moreover, the advancement of ageing is seen, in a population with a very high proportion of people older than 80 years, in the population aged 60 and older.

The study of ageing requires examination if regional sub-populations with similar intensity of ageing also feature common demographic conditions of this process. We can use a contingency table which displays the multivariate frequency distribution. This approach is simple for categorical data or if a division of numerical data is given.

Otherwise we can use multivariate data analysis (see Colley and Lohnes, 1971; Johnson and Wichern, 1992; Everitt and Dunn, 2001; Hair et al, 2006). A regression analysis can identify an overall impact of the demographic factors on the intensity of population ageing and helps to select variables for further studies. But in terms of creating a policy of regional development, it is valuable to identify common factors of ageing in sub-populations exhibiting a similar intensity of this process.

Discriminant analysis is useful when simple, linear interactions between a dependent and the explanatory variables exist, and when the random variables in the model follow a multivariate normal distribution (see Fisher, 1936; Klecka, 1980; Lachenbruch, 1975). In contrast to this, classification trees are nonparametric and the assumption of a multivariate normal distribution is not relevant. They are useful, when complicated and nonlinear relations exist (see Breiman et al, 1984; Ripley, 1996; Rokach and Maimon, 2008).

Classification trees are widely used in medicine, computer science, botany, psychology, finance, marketing, engineering etc. rather than in demography. But Ninčević et al (2010) used them to examine the impact of various factors on life expectancy. Toulemon (2006) applied them to model the transition to adulthood and find differences between Austrians' and Italians' paths. High predictive performance, simplicity and transparency of the classification scheme make them a promising tool in examining the population ageing as well.

3 Classification trees in the analysis of population ageing

This paper examines the situation of Poland which is one of the fastest developing EU countries. Although the Polish population is relatively young, it exhibits an intensive demographic ageing process. The identification of the degree of ageing and its regional diversification, and finding its demographic factors is crucial in order to impede or mitigate the effects of ageing and program the right regional development policy. The main interest of the study is to examine if the sub-populations similar in their degrees of ageing, feature common demographic conditions of this process as well.

The investigation was carried out for 66 Polish subregions. These NUTS 3 units are statistical rather than administrative territorial units. They come from a division of 16 Polish provinces. As Polish subregions represent relatively homogeneous territorial areas in respect to economic, social, cultural and environmental features, we can assume that they also exhibit individual demographic situations. The study covers the period 1995-2012 because the demographic changes can only be seen over a long time period. Empirical data comes from the Central Statistical Office of Poland.

This paper uses classification trees in the comparative analysis of demographic conditions of population ageing. Classification trees are used to predict a membership of units in the categories of a categorical dependent (response) variable from their measurements on a set of predictors (explanatory variables). In this study, they help to find differences between predefined classes of subregions and identify profiles of these classes. The next sections follow a classification procedure:

1. Specifying the demographic features of population ageing. Sect. 4 identifies the degrees of ageing in Polish subregions, while Sect. 5 discusses the demographic factors affecting this process.
2. Classifying a set of subregions. Sect. 6 describes the construction and estimation of a classification tree and presents the classification results.
3. Profiling classes. Sect. 7 identifies demographic conditions of population ageing and makes policy suggestions. Sect. 8 gives comments and open questions.

4 The intensity of the population ageing process

The comparative studies of demographic ageing usually use a typology of the population age. Each of them takes some indicator of a population structure and distinguishes a set of categories of demographic age according to a value of this indicator. The first typology was developed by Sündbarg (1900), the others were proposed by Sauvy (1948), Beaujeu-Garnier (1966), and Veyret-Verner (1971) etc.

In their recent works, the United Nations (UN) distinguishes five demographic ages according to the share of senior citizens in the total population. If the percentage of people aged 65 or older is less than 4%, the population is demographically young. A share above 4% and below 7% indicates a mature population. A share below 14% means an ageing population while a share between 14% and 21% defines an old population. More than 21% is typical of the aged population (see Coulmas, 2007, p. 5).

The population age defines the present demographic situation of a population but does not indicate the expected changes in the age structure. Therefore, in analyzing a relatively young population such as Poland, we should also consider the increase rate of the share of senior citizens. A value of the rate shows a scale of changes and can be used to foresee the future age structure.

Table 1 Basic statistics of the share of people aged 65 or older in 2012, and the increase rate of this share in the period from 1995 to 2012 in 66 Polish subregions (NUTS 3 units)

Variable name	Share of senior citizens in 2012 (%)	Increase rate of the share in the years 1995-2012 (%)
Poland	14.2	1.4
Minimum	9.9	0.2
Maximum	18.6	3.4
Median	13.9	1.4
Mean	14.1	1.5
Standard deviation	1.7	0.7
Coefficient of variation (%)	12.2	46.6
Pearson's correlation $[-1,1]$	0.08	

Over the last thirty years, the share of young people in Poland has significantly decreased, while the number of older people increased. The yearly increase of the share of senior citizens was 1.4% from 1975 to 2012 to reach 14.2% of people aged 65 or older in 2012 (see Table 1).

This situation is not homogeneous in the whole country. High socio-economic disparities and also cultural, social and environmental differences translate into a territorial diversification of the age structure and its changes in time. The values of the share of senior citizens were in the range of [9.9, 18.6] in 66 Polish subregions in 2012 (see Table 5 for details). According to the UN classification, we can distinguish subregions with ageing (53%) and subregions with old (47%) population, respectively.

All subregions showed yearly increases of this share; the majority of them showed a yearly increase between 0.5 and 2.0%. According to the national average value (1.5%), two groups of subregions were distinguished: relatively low (56% of subregions) and high progress of ageing (44% of subregions).

It is interesting that the share of senior citizens is not statistically correlated with the increase rate of this share. We cannot conclude that the higher the increase rate of people aged 65 and older in the total population, the older the population (and inversely). It is also not true that the older the population, the lower the increase rate of the share (and inversely). We can use both features, the demographical age of a population and the progress of ageing to divide subregions into four classes:

1. The "old-high" class includes 24.2% of subregions with a high share of senior citizens and also a high growth rate of this share.

2. The "old-low” class consists of 22.7% of subregions with a high share of senior citizens but with a low growth rate of this share.
3. The "ageing-high” class is represented by 19.7% of subregions with a low share of senior citizens but with a high growth rate of this share.
4. The "ageing-low" class is the biggest and covers 33.3% of subregions with a low share of senior citizens and also with a low growth rate of this share.

In general, the most intensive ageing relates to the majority of subregions located in south western Poland (see Fig. 1). This process is also seen in the biggest Polish cities. Demographically old, but not ageing quickly, is the population of east Poland and also some subregions of central Poland which belong to the Łódzkie and Świętokrzyskie provinces). Still relatively young but quickly ageing is the population of north western and west Poland and also the Warmińsko-mazurskie province.

5 Factors of the population ageing process

This section examines five demographic factors of ageing (discussed in Sect. 2) relevant for the Polish population. An indicator of the generation replacement is the total fertility rate. It is the average number of children a woman would bear over the course of her lifetime if current age-specific fertility rates remained constant throughout her childbearing years. Values between 2.10 and 2.15 provide the replacement of a generation without excessive growth or shrinkage of a population.

The rate reached approximately 1.30 in Poland and from 0.9 to 1.6 in subregions but 85% of them did not reach 1.4 in 2012 (see Table 2). Although the replacement condition was not satisfied, the fertility rate has been increasing. But this growth is extremely slow and we cannot expect any extra changes in the next 20 years.

An indication of life duration is the life expectancy of people at the age of 65. It is the mean number of years still to be lived by a man or woman who has reached the age of 65, if subject throughout the rest of his/her life to the current mortality conditions (age-specific probabilities of dying). The Polish population is living longer and longer. Yearly increases in the life expectancy of men and women at the age 65 are observed in all subregions. An average 65 year old man will live to be 80.4 years old which is approximately 4.0 years less than an average woman. The statistical correlation of the indicator values

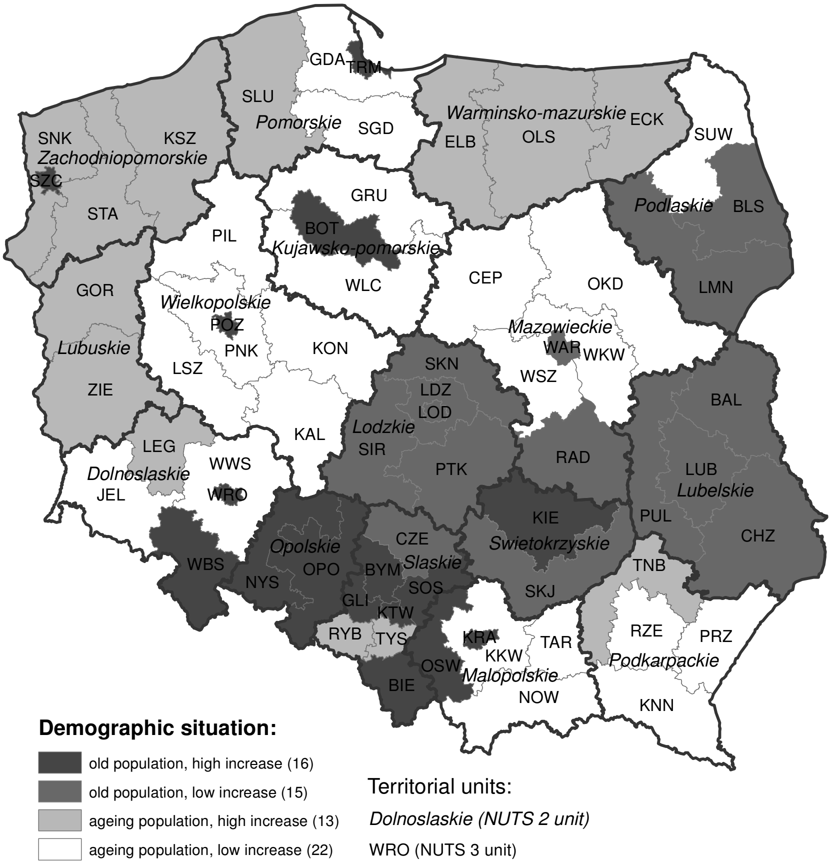

Fig. 1 Four classes of 66 Polish subregions (NUTS 3 units) according to the demographic age in 2012 and the progress of population ageing in the years 1995-2012 (see Table 5 for details)

for men and women in subregions is very high. Thus the study includes only the situation of men.

An indicator of changes in a demographic structure is the net migration of people in the age between 20 and 59. It is the difference between registered (domestic and international) migration inflow and outflow of people with a permanent residence in a region to the average population of the region. The rate took values between –55 and +100 in 2012 in Polish subregions. In 2012 nearly 75% of subregions had a negative balance of migration (higher outflow than inflow of people). Yearly increases of the rate were seen in the majority of

Table 2 The demographic factors of population ageing in Polish subregions (NUTS 3 units)

Abbreviation	Variable name	Year*	Poland	Median	Minimum	Maximum
Fertility	Total fertility rate (person)	2002	1.249	1.312	0.893	1.614
		2012	1.299	1.298	1.091	1.632
Life	Life expectancy of men	2007	14.6	14.3	13.6	16.4
	at the age 65 (year)	2012	15.4	15.2	14.2	16.9
Migration	Net migration rate of population	1995	*NA*	*NA*	*NA*	*NA*
	aged 20-59	2012	*NA*	–21.2	–55.2	100.2
Working	Working-age population ageing	2009	37.8	37.7	33.7	41.7
	rate (%)	2012	37.4	37.7	34.4	41.1
Oldest	Oldest-old-age population	2005	20.3	20.3	16.2	24.2
	rate (%)	2012	26.3	26.3	19.5	31.3

*depending on the data availability, *NA* – not applicable.

subregions. A much higher inflow than outflow of young people is seen in the biggest Polish cities and their surroundings.

An indicator of ageing of an adult population is working-age population ageing rate. The working age is the age between 18 and the retirement age. The age 45 distinguishes the immobile and mobile working age. The working-age population ageing rate means the percentage of immobile working-age people in the working-age population. The rate was 37.4% in Poland in 2012 but four subregions exceeded the threshold of 40.0%.

The progress of ageing of the senior population is shown by the oldest-old-age population rate. It is the percentage of people aged 80 and older in the total population aged 65 and older. The rate reached 26.3% in Poland in 2012. That was 6 percentage points more than in 2005. But two subregions exceeded 30% while one subregion did not reach 20.0%. Yearly increases of the rate were seen in all subregions.

Two separate linear regression models served to examine the statistical significance of these demographic factors in explaining the share of senior citizens (see Table 3) and its increase rate (see Table 4). The estimate of the total fertility rate is statistically significant exclusively for the increase rate which means that the impact of fertility is perceptible in the long term rather than in the short term. The higher the fertility, the smaller the progress of ageing. A Similar situation relates to the net migration rate whose estimate is significant only for the increase rate of the share of seniors. The lower the net migration rate, the older the population.

Both estimates of life expectancy are significant and positive. The increase of life expectancy in a population causes an increase of the share of senior citizens more intensively than the increase rate of this share. A similar situation is presented by the working-age population ageing rate. The oldest-old-age population rate is also related to both dependent variables but proved bilateral impact. The older the senior population, the older the whole population but the lower the progress of ageing.

6 Construction and estimation of a classification tree

Poland has high regional disparities according to the share of senior citizens and its increase rate. Four classes of subregions presenting different degrees of ageing were distinguished in Sect. 4. Although all demographic factors affecting the situation of subregions discussed in Sect. 5 are statistically significant, there are no obvious differences between classes according to the average values of demographic factors (see Table 5). For example, the averages of the working-age population ageing rate are very similar in old-high, old-low and ageing-high classes.

We cannot profile classes on the basis of these results. In this situation we construct a classification tree to depict relations between the degrees of population ageing and a set of demographic factors of this process. The set of four classes of subregions served as the realizations of categorical dependent variable, while all demographic factors formed a set of explanatory variables.

In the study, the CART algorithm proposed by Breiman et al (1984) is applied to profile pre-assigned classes for subregions. CART produces a tree-structured model using recursive binary partitioning. The algorithm asks a sequence of questions which split a set of objects into two subsets. Splitting is determined by a condition of the value of a single explanatory variable which is satisfied by the observation or it is not.

The starting point of a tree, called a root, consists of the entire learning set. A set of nodes originates from the root. A nonterminal (or parent or internal) node is a node that splits into two daughter nodes. A node which has stopped splitting is called a terminal node (or a leaf). A path from the root to the terminal node shows the classification rules on which the units are assigned to a class.

The construction and estimation of a tree includes three steps: the partitioning of the dataset, determining the complexity of the tree and validating the results.

Table 3 Estimation results for the share of senior citizens in 2012 (least-squares method)

Name*	Coefficient	Standard error	Student's t-test	p-value
Constant	–52.664	7.196	–7.318	6.05e-010[a]
Fertility	x	x	x	x
Life	2.005	0.285	7.034	1.87e-09[a]
Migration	x	x	x	x
Working	0.821	0.108	7.601	1.95e-010[a]
Oldest	0.205	0.066	3.086	0.003[a]
R-Squared		0.604778		

* Fertility: the total fertility rate (person), Life: the life expectancy of men at the age 65 (year), Migration: the net migration rate of population aged 20-59, Working: the working-age population ageing rate (%), Oldest: the oldest-old-age population rate (%)
[a] denotes statistical significance at the 99% level
x denotes statistical insignificance

Table 4 Estimation results for the increase rate of the share of senior citizens in the years 1995-2012 (least-squares method)

Name*	Coefficient	Standard error	Student's t-test	p-value
Constant	0.855	3.714	0.230	0.8187[b]
Fertility	–1.981	0.546	–3.626	0.0006[a]
Life	0.452	0.111	4.080	0.0001[a]
Migration	–0.004	0.001	–3.914	0.0002[a]
Working	0.084	0.045	1.869	0.0665[b]
Oldest	–0.265	0.019	–14.260	6.27e-021[a]
R-Squared		0.817918		

* Fertility: the total fertility rate (person), Life: the life expectancy of men at the age 65 (year), Migration: the net migration rate of population aged 20-59, Working: the working-age population ageing rate (%), Oldest: the oldest-old-age population rate (%)
[a] and [b] denote statistical significance at the 99% and 90% levels respectively

Table 5 The average values of demographic factors determined for four classes of regions in 2012

Variable name*	Old-high class	Old-low class	Ageing-high class	Ageing-low class
Fertility	1.199	1.289	1.288	1.379
Life	15.5	15.2	14.9	15.3
Migration	–7.6	–16.2	–21.4	3.5
Working	38.1	37.9	38.1	36.5
Oldest	25.3	28.1	24.6	26.1

* Fertility: the total fertility rate (person), Life: the life expectancy of men at the age 65 (year), Migration: the net migration rate of population aged 20-59, Working: the working-age population ageing rate (%), Oldest: the oldest-old-age population rate (%)

The tree growing procedure is based on the partitioning rules. They divide subsets of the learning set with respect to a dependent variable and create the daughter nodes from a parent node. The data in each of the daughter nodes is obtained by reducing the number of cases that has been misclassified. All of the possible ways of splitting are tested and the one which leads to the greatest increase in node purity is chosen. The goodness of a potential split is indicated by an impurity function. This is a function of the proportion of the learning sample belonging to the possible classes of the dependent variable. This study uses Gini's index of impurity.

The second problem of classification is to select a tree of the right size in such a way as not to overfit the learning sample, as well as to achieve an exhaustive representation of the data. In this study, the dataset is relatively small, so we use a strategy of growing a fully expanded tree and then pruning it back, or removing some of its nodes to produce a tree with a smaller number of terminal nodes. This produced a finite sequence of nested subtrees from which the best solution was chosen.

The membership accuracies in a sequence of subtrees can be compared using some estimates of their misclassification rates. One of the possible solutions, if one has enough data, is to distinguish an independent test set. In this study, due to having a not very large number of cases, a V-fold cross-validation estimate of the misclassification rate is used. The entire dataset used as a learning set is randomly divided into V parts of approximately equal sized, disjoint subsets. In this study V was equal to 5.

The subtree with the smallest estimated misclassification rate equal to 12.1% is selected to be the final tree-based classification model. Table 6 presents a tree growing scheme. The classification tree has 10 internal nodes and 12 leaves which determine the profiles of classes. The node numbers indicate the rules of a division. For example, the leaf node with the number of 2.2.2.2.2.1 is a subdivision of an internal node with the number of 2.2.2.2.2 according to the values of the life expectancy of men at the age 65 higher than 14.85 years. All calculations were made in Statistica 10 software.

An example of the classification rules for a subclass of the ageing-low class is the leaf node number 2.2.2.1.2 in Table 6. This subclass includes subregions with the rate of adult population ageing between 37.09% and 37.78% which is very close to the Polish national average equal to 37.4% in 2012. The percentage of people aged 80 and older in the total population is lower than 27.60%. The life expectancy of men at the age 65 is less than 15.15 years which is rather shorter than the Polish national average equal to 15.4% in 2012. The total

Table 6 Classification tree

Node number	Type of node	Input variable*	Splitting criterion	Number of sub-regions (66)	Number of old-high class members (16)**	Number of old-low class members (15)**	Number of ageing-high class members (13)**	Number of ageing-low class members (22)**
1	internal	Working	≤37.09	26	3	4	1	**18**
1.1	leaf	Fertility	>1.25	18	0	1	0	**17**
1.2	internal	Fertility	≤1.25	8	**3**	3	1	1
1.2.1	leaf	Working	≤35.99	3	**3**	0	0	0
1.2.2	internal	Working	>35.99	5	0	**3**	1	1
1.2.2.1	leaf	Migration	≤−18.18	2	0	0	1	**1**
1.2.2.2	leaf	Migration	>−18.18	3	0	**3**	0	0
2	internal	Working	>37.09	40	**13**	11	12	4
2.1	leaf	Oldest	>27.60	6	0	**6**	0	0
2.2	internal	Oldest	≤27.60	34	**13**	5	12	4
2.2.1	internal	Life	>15.15	9	**8**	0	1	0
2.2.1.1	leaf	Life	>15.30	6	**6**	0	0	0
2.2.1.2	leaf	Life	≤15.30	3	**2**	0	1	0
2.2.2	internal	Life	≤15.15	25	5	5	**11**	4
2.2.2.1	internal	Fertility	>1.37	5	0	**3**	0	2
2.2.2.1.1	leaf	Working	>37.78	3	0	**3**	0	0
2.2.2.1.2	leaf	Working	≤37.78	2	0	0	0	**2**
2.2.2.2	internal	Fertility	≤1.37	20	5	2	**11**	2
2.2.2.2.1	leaf	Fertility	>1.28	11	0	1	**9**	1
2.2.2.2.2	internal	Fertility	≤1.28	9	**5**	1	2	1
2.2.2.2.2.1	leaf	Life	>14.85	3	0	**1**	1	1
2.2.2.2.2.2	leaf	Life	≤14.85	6	**5**	0	1	0

* Fertility: the total fertility rate (person), Life: the life expectancy of men at the age 65 (year), Migration: the net migration rate of population aged 20-59, Working: the working-age population ageing rate (%), Oldest: the oldest-old-age population rate (%)
** Bold values indicate a predicted class for which a cost of misclassification is the lowest

fertility rate is higher than 1.37 which is much more than the Polish national average equal to 1.299 in 2012. A population of subregions belonging to this subclass is relatively young and does not age very quickly due to presenting a relatively short life duration and high fertility as well.

7 Demographic conditions of population ageing within Poland

The results of a classification tree help to profile 4 classes of subregions. The first (old-high) class with an old and still ageing population consists of 3 subclasses. The most intensive ageing is seen in the big socio-economic centers in Poland:

the cities of Cracow (KRA), Wrocław (WRO) and Poznań (POZ). See Table 7. The GDP *per capita* in these cities is 150% of the national average value (Poland: 37,096 PLN). In spite of the relatively young working-age population, extremely low fertility accompanied by high life expectancy is seen in these subregions.

This probably results from the intentional delay of procreation time and changing life priorities from family-related to profession-related. The situation requires a redefinition of social policy. Exemplary activities include promoting starting a family, giving help to women to reconcile their career with caring for children, e.g. by providing social infrastructure (e.g. increasing the availability of nurseries), giving an opportunity for men to take paternal leave etc.

The second subclass covers economically well developed subregions with the cities inside and also the city of Szczecin (SZC). They exhibit an extremely old working-age population which is accompanied by low fertility and high life expectancy. These subregions experience serious demographic problems which are difficult to solve. The population of young people is shrinking. This can disturb the regional labour market and diminish demand for jobs. Regional policy, e.g. creating new jobs and services, should attract people to live in these subregions.

The intensive process of population ageing is also typical of weak, neighbouring subregions of south-western Poland apart from the Katowicki (KTW) subregion. Extremely old working-age population and low fertility, and also a big migration outflow of young people occur in this area. These subregions suffer from very serious demographic and economic problems which affect a depopulation process and may lead these subregions to become extinct.

The second (old-low) class with an old but very slowly ageing population also distinguishes three subclasses. The first subclass covers a few subregions together with Warsaw (WAR), the capital city of Poland. The situation of this subclass is very similar to the situation of the cities of Cracow, Wrocław and Poznań. But it will be much more difficult to increase fertility in Warsaw due to social and cultural (e.g. life style) and economic (e.g. high cost of living) adversities.

The old-low class also includes the weakest Polish subregions (less than 75% of the national average value of GDP *per capita* value in 2012) with common borders such as the Chełmsko-zamojski (CHZ), Puławski (PUL), Radomski (RAD) and Sandomiersko-jędrzejowski (SKJ) subregions. Extremely high migration outflow of young people from these subregions and also a lot of oldest-old age people directly result from the long-term economic problems.

Table 7 Share of people aged 65 or older and its increase rate in 66 Polish subregions (NUTS 3)

Name	Full name	Share of of senior citizens in 2012 (%)	Increase rate of the share in 1995-2012 (%)	Name	Full name	Share of of senior citizens in 2012 (%)	Increase rate of the share in 1995-2012 (%)
BAL	Bialski	14.1	0.4	OPO	Opolski	15.0	2.8
BIE	Bielski	14.4	1.7	OSW	Oświęcimski	14.5	1.8
BLS	Białostocki	14.6	1.3	PIL	Pilski	11.9	1.3
BOT	Bydgosko-toruński	14.3	1.9	PNK	Poznański	10.7	0.5
BYM	Bytomski	15.4	2.8	POZ	City of Poznań	16.4	1.6
CEP	Ciechanowskopłocki	13.9	1.1	PRZ	Przemyski	13.9	0.9
CHZ	Chełmsko-zamojski	15.3	0.7	PTK	Piotrkowski	14.3	0.9
CZE	Częstochowski	15.8	1.3	PUL	Puławski	15.6	1.0
ECK	Ełcki	12.3	1.9	RAD	Radomski	14.1	0.9
ELB	Elbląski	12.2	1.6	RYB	Rybnicki	13.6	3.4
GDA	Gdański	9.9	1.4	RZE	Rzeszowski	13.6	1.2
GLI	Gliwicki	15.3	3.2	SGD	Starogardzki	11.2	1.4
GOR	Gorzowski	12.5	1.6	SIR	Sieradzki	15.0	0.8
GRU	Grudziądzki	12.6	1.1	SKJ	Sandomiersko-jędrzejowski	16.1	0.3
JEL	Jeleniogórski	13.8	1.3	SKN	Skierniewicki	15.5	1.0
KAL	Kaliski	13.5	1.0	SLU	Słupski	11.9	1.9
KIE	Kielecki	15.3	1.8	SNK	Szczeciński	11.4	1.9
KKW	Krakowski	13.0	0.2	SOS	Sosnowiecki	15.7	1.9
KNN	Krośnieński	13.7	1.4	STA	Stargardzki	12.3	1.6
KON	Koniński	13.0	1.3	SUW	Suwalski	13.9	1.0
KRA	City of Kraków	16.6	1.9	SZC	City of Szczecin	16.0	2.2
KSZ	Koszaliński	13.0	2.1	TAR	Tarnowski	13.9	1.5
KTW	Katowicki	16.2	2.7	TNB	Tarnobrzeski	13.4	1.8
LDZ	Łódzki	15.5	1.1	TRM	Trójmiejski	17.0	2.3
LEG	Legnicko-głogowski	12.7	2.5	TYS	Tyski	12.4	2.7
LMN	Łomżyński	16.2	0.9	WAR	City of Warsaw	18.0	1.2
LOD	City of Łódź	18.6	1.2	WBS	Wałbrzyski	15.2	1.5
LSZ	Leszczyński	12.1	0.8	WKW	Warszawski wschodni	12.8	0.6
LUB	Lubelski	14.6	1.3	WLC	Włocławski	13.4	1.2
NOW	Nowosądecki	12.5	1.4	WRO	City of Wrocław	16.6	1.7
NYS	Nyski	14.4	1.7	WSZ	Warszawski zachodni	13.6	0.5
OKD	Ostrołęcko-siedlecki	13.9	0.5	WWS	Wrocławski	12.1	0.8
OLS	Olsztyński	12.5	2.0	ZIE	Zielonogórski	12.8	1.5

This is a very similar situation to the last subclass of the old-low class. An effective social policy and economic measures to stimulate regional economies may help to keep young people and impede shrinkage.

The third (ageing-high) class is formed by subregions with a relatively young population which experiences the ageing process. It is the most homogeneous class according to demographical conditions but differs in economic and environmental factors. This class consists of selected subregions of the Lubuskie region (the Gorzowski (GOR) and Zielonogórski (ZIE) subregions) located in western Poland, subregions of the Śląskie region (the Rybnicki (RYB) and Tyski (TYS) subregions) located in southern Poland, the Warmińsko-mazurskie region (the Elbląski (ELB) and Ełcki (ECK) subregions) located in northern Poland, and also the Słupski (SLU) and Legnicko-głogowski (LEG) subregions. Low fertility, accompanied by an old working-age population, which results from the migration outflow of young people, are typical features of this class. Some of these subregions will probably join the old-high class in the future.

The fourth (ageing-low) class presents the smallest population ageing but is the most diversified according to its demographic conditions. The majority of its subregions has a relatively young working-age population but also high life expectancy. Extremely weak economic situation (GDP *per capita* less than 60% of the national average value) translates into a high migration outflow of young people.

Subregions surrounding the biggest economic centers of Poland (the cities of Warsaw, Wrocław, Poznań, Cracow and Gdańsk) show an extremely high positive net migration of young people. This results from the sub-urbanization process and expansion of these cities. The rest of the subregions of the ageing-low class home a relatively old working-age population and low fertility, accompanied by a high migration outflow of young people. This subclass will probably join the ageing-high class if no preventive measures are taken.

8 Discussion

Population ageing is not simple to examine because its foundations do not lie only in demographic factors. For example, in general, we can assume that subregions similar in their economic situation are also similar demographically. However, when we compare the situation of the biggest Polish cities with the other economically well developed subregions, we cannot expect that the demographic situation of such different types of agglomerations will be similar.

The occurring sub-urbanization process may also disturb the analysis results. Thus we cannot expect that the youngest population lives in the biggest cities.

Many young people work in a city but live outside the city. It is a problem how to interpret this situation. Is the population of economic centres really demographically old?

Some demographic factors of the population ageing process which were not disclosed in this study also exist. For example, eastern Poland is economically the weakest macro-region, so we can suppose that the population of the subregions of eastern Poland is demographically the oldest. But there are subregions in the rest of Poland which are economically well developed and also show a very intensive process of population ageing. For example, some pensioners, after working many years outside, come back to their native region, e.g. to the Śląskie region, to spend the rest of their lives there.

Although demographic changes are not rapid and take a long-time, the analysis shows that results are not stable in time. Changes of economic trends, social and economic policies, actions of local authorities, and even some misfortunes affect the demographic situation. Carrying out additional studies is recommended.

9 Conclusions

The paper conducts a comparative study of demographic features of populations exhibiting the ageing process using classification trees. It presents an original approach in the analysis of population ageing which is not well recognized in the literature.

1. It examines the demographic determinants of population ageing besides the intensity of ageing.
2. It concerns the internal situation of a country rather than international comparisons.
3. It uses classification trees which are not frequently applied in demography and population ageing studies.
4. It indicates regional problems and conditions of ageing in Poland.

The main advantage of using classification trees is to discover two, previously unknown situations difficult to find with other tools. The most important result of the study is to prove that regions similar in their intensity of population ageing may feature different demographic conditions of the process. Only subregions with young but quickly ageing population are common in their factors of ageing.

The second, unexpected finding is that the intensity of population ageing is not correlated with the economic situation of subregions. But subregions similar in their economic situation exhibit common demographic factors of ageing as well. In general, well developed regions mostly suffer from very low fertility, while poor regions struggle with a high out-migration of young people and an old working-age population. This requires an individual approach in mitigating the effects of ageing and impeding this process within a country and, besides an effective national social policy and also coordinated activities of regional and local authorities.

References

Beaujeu-Garnier J (1966) Geography of population. Longmans, London

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and Regression Trees. Statistics/Probability Series, Wadsworth, Belmont, California

Clarke JI (1965) Population geography. Pergamon Press, Oxford

Colley WW, Lohnes PR (1971) Multivariate data analysis. Wiley, New Jersey

Coulmas F (2007) Population decline and Ageing in Japan – the social consequences. Routledge, New York

Everitt BS, Dunn G (2001) Applied Multivariate Data Analysis. Wiley, London

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics 7(2):179–188, URL http://hdl.handle.net/2440/15227

Giannakouris K (2008) Aging characterises the demographic perspectives of the european societies. Eurostat Statistics in Focus 72:1–11

Grundy E (1996) Population ageing in europe. In: Coleman D (ed) Europe's Population in the 1990s, Oxford University Press, Oxford, pp 267–299

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) Multivariate Data Analysis, 6th edn. Prentice Hall, New Jersey

Johnson RA, Wichern DW (1992) Applied multivariate statistical analysis, 3rd edn. Prentice-Hall, New Jersey

Klecka WR (1980) Discriminant analysis. Sage Publications, Beverly Hills

Lachenbruch PA (1975) Discriminant analysis. Hafner Press, New York

Légaré J (2006) Economic, social and cultural consequences of the ageing of the population. In: Caselli D, Vallin J, Wunsch G (eds) Demography - Analysis and Synthesis: A Treatise in Population, Elsevier, London, pp 327–336

Lindh T, Malmberg B (2009) European union economic growth and the age structure of the population. Economic Change and Restructuring 42(3):159–187, DOI 10.1007/s10644-008-9057-1

Magnus G (2008) The age of aging: how demographics are changing the global economy and our world. Wiley, New York

Martinez-Fernandez C, Kubo N, Noya A, Weyman T (2012) Demographic Change and Local Development: Shrinkage, Regeneration and Social Dynamics. OECD Publishing, Paris, DOI 10.1787/9789264180468-en

Muenz R (2007) Aging and demographic change in european societies: Main trends and alternative policy options. SP Discussion Paper 703, World Bank Group

Ninčević I, Ćukušić M, Garača Z (2010) Mining demographic data with decision trees. In: MIPRO, Proceedings of the 33rd International Convention, IEEE, Opatija, Croatia, pp 1288–1293

Preston SH, Himes C, Eggers M (1989) Demographic conditions responsible for population ageing. Demography 26(4):691–704, DOI 10.2307/2061266

Prskawetz A, Lindh T (2011) The relationship between demographic change and economic growth in the EU. Austrian Academy of Sciences Publishing House, Wien

Ripley BD (1996) Pattern recognition and neural networks. Cambridge University Press, Cambridge

Rokach L, Maimon O (2008) Data mining with decision trees: theory and applications. World Scientific, New Jersey Publishing

Sanderson W, Scherbov B (2008) Rethinking age and ageing. Population Bulletin 63(4):1–16

Sauvy A (1948) Social and economic consequences of the ageing of western european populations. Population Studies 2(1):115–124, DOI 10.1080/00324728.1948.10416342

Sündbarg G (1900) Sur la répartition par âge et sur les taux de la mortalité. Bulletin de l'Institut international de statistique 12(1):89–98

Toulemon L (2006) Multidimensional exploratory analysis. In: Caselli D, Vallin J, Wunsch G (eds) Demography - Analysis and Synthesis: A Treatise in Population, Elsevier, London, pp 683–684

Uhlenberg P (2009) International Handbook of Population Aging. Springer, Dordrecht, NL, DOI 10.1007/978-1-4020-8356-3

United Nations (2013) World population ageing 2013. ST/ESA/SER.A/34, department of Economic and Social Affairs

Veyret-Verner G (1971) Populations vieillies. Types, variétés des processus et des incidences sur la population adulte. Revue de géographie alpine 59(4):433–456, DOI 10.3406/rga.1971.1446

Weil DN (1997) The economics of population aging. In: Rosenzweig MR, Stark O (eds) Handbook of Population and Family Economics, Elsevier, New York, pp 967–1014

World Population Prospects (2012) World population prospects: The 2012 revision. http://esa.un.org/wpp/, department of Economic and Social Affairs, United Nations