10.1515/cer-2016-0044

GRAŻYNA TRZPIOT*, AGNIESZKA ORWAT-ACEDAŃSKA**

Spatial Quantile Regression In Analysis Of Healthy Life Years In The European Union Countries

Abstract

The paper investigates the impact of the selected factors on the healthy life years of men and women in the EU countries. The multiple quantile spatial autoregression models are used in order to account for substantial differences in the healthy life years and life quality across the EU members. Quantile regression allows studying dependencies between variables in different quantiles of the response distribution. Moreover, this statistical tool is robust against violations of the classical regression assumption about the distribution of the error term. Parameters of the models were estimated using instrumental variable method (Kim, Muller 2004), whereas the confidence intervals and p-values were bootstrapped.

Keywords: *quantile regression, multiple spatial quantile autoregression, spatial analysis, healthy life years*

1. Introduction

Public health is treated as one of the most important factors determining strength of an economy. It affects the productivity of labour, labour supply, human capital as well as public spendings. Investment in health is one of the

* Full Professor, Department of Demography and Economic Statistics, Faculty of Informatics and Communication, University of Economics in Katowice, e-mail: trzpiot@ue.katowice.pl

** Ph.D., Department of Demography and Economic Statistics, Faculty of Informatics and Communication, University of Economics in Katowice, e-mail: agnieszka.orwat@ue.katowice.pl

priorities of the Europe 2020 strategy, which aims at promoting a sustainable development in Europe. In order to assess improvements in the public health or effects of healthcare programs health measures are utilized which take into account not only negative indicators of mortality or life expectancy. In the face of the population ageing, one of the most popular positive health measures is the healthy life years (HLY). Generally, they are defined as the expected remaining healthy life years without a disability.

The expected healthy life years (HLY) at birth of an average European is equal to 61,7 years for a male and 62 years for a female.[1] For men, the difference between the healthy life years in Latvia (the country with the smallest value of the HLY for men – 51,7 years) and Malta (the country with the highest value of the HLY for men – 71,6 years) reaches almost 20 years. For women, the difference between Lithuania (54,2 years) and Malta (72,7 years) is equal to 18,5 years. Such a big gap clearly shows that Europe is severely differentiated as far as quality of life is concerned. To identify the factors determining the observed differences in the HLY statistical procedures accounting for spatial effects as well as the substantial heterogeneity of the error term should be applied because the impact of the investigated socio-economic factors and health determinants could vary across the countries. Using the classical statistical methods only that largely focus on mean values (like the spatial models based on classical regression) can lead to falsely results on the HLY determinants.

Therefore we extend the *spatial autoregression* models (SAR) using *quantile spatial autoregression regression* models (QSAR). The quantile estimation of the spatial model put a more light on the spatial dependencies in different parts of the healthy life years distribution.

The aims of the paper are: to identify the health and socio-economic factors associated with the expected healthy life years in the EU countries and to verify the multiple, spatial, quantile autoregression method.

The multiple, quantile, spatial autoregression models are used in order to account for substantial differences in the healthy life years and life quality between the EU members. Quantile regression allows studying dependencies between variables in different quantiles of the response distribution. Moreover, this statistical tool is robust against violations of the classical regression assumption about the distribution of the error term. Parameters of the models were estimated using instrumental variable method (Kim, Muller 2004), whereas the confidence intervals and p-values were bootstrapped.

The paper is organized as follows. In the first section, we describe the SAR model. The next section introduces the concept of the quantile regression. Then,

[1] According to EUROSTAT data for 2015 year.

we blend the two methodologies and introduce the QSAR model. The fourth section contains the empirical analysis. It consists of two subsections. First, we describe the explanatory variables used in the study and the main assumptions used in the empirical study. Then, we present and discuss the results.

2. Spatial Autoregressive Model – SAR model[2]

SAR model has the following form (Anselin 1988; Suchecki 2010):

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where: $\mathbf{Y}$ – vector of dependent variable realizations, $\mathbf{W}$ – spatial weight matrix, ρ – autoregression parameter, $\mathbf{X}$ – matrix of covariates realizations, $\boldsymbol{\beta}$ – vector of parameters, $\boldsymbol{\varepsilon}$ – error terms vector.

Model (1) is a linear regression model with the additional spatial autoregression term. Spatial autocorrelation measures the correlation between value of a variable in one localization and its value in the other localization (region, for example). Spatial autoregression is represented by the spatial lag term $\rho\mathbf{WY}$ of the dependent variable. Vector of the error terms has a multivariate normal distribution:[3]

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma\mathbf{I}) \tag{2}$$

The least squares estimator of the parameters of model (1) is inconsistent (Lee 2002). Therefore, many consistent alternatives have been proposed, particularly maximum likelihood, instrumental variables (Anselin 1988), generalized method of moments or two-stage least squares (Lee 2007). For large-scale spatial models, Bayesian estimation is also employed by LeSage (1997) and Lum, Gelfand (2012).

In case of the error term with asymmetric distribution, fat tails or heteroscedasticity as well as outliers, the standard estimation and inference techniques have low power. This comes as a consequence of large estimation errors of the parameters or the error term variance. More importantly, the standard SAR methodology, as a conditional expected value model, focuses solely on the relationships observed in the central part of the outcome distribution. Therefore, it cannot provide any insights into dependencies in the other parts of the distribution.

---

[2] In the literature, a few names of the model are utilized interchangeably. Using SAR term we follow LeSage, Pace (2009).

[3] In the paper, we do not consider the model with correlated error terms. The quantile versions of such model can be estimated using Bayesian methods only.

This is not the case for quantile regression which allows studying an impact of covariates on the outcome in aby point of the outcome distribution.

3. Quantile Regression Model – QR model

We analyse a problem of estimation of a vector of parameters $\boldsymbol{\beta}$ for a sequence of random variables $Y_1, Y_2, \ldots, Y_n$ drawn from a distribution $P(Y_i < y) = F(y - \mathbf{X}_i'\boldsymbol{\beta})$, where $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ik})'$ is a column of $n \times k$ covariate matrix $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)'$, $i = 1, 2, \ldots, n$ and the distribution F is unknown.

The point of departure for quantile regression is the conditional quantile function of a random variable Y:

$$Q_{\mathbf{Y}}(\tau|\mathbf{X}) = F^{-1}(\tau|\mathbf{X}), \tag{3}$$

where $\tau \in [0, 1]$ denotes the order of a quantile. The quantile regression model of order τ takes the following form:

$$Y_i = \mathbf{X}_i' \boldsymbol{\beta}^{(\tau)} + \varepsilon_i^{(\tau)}, \tag{4}$$

where $Y_i \equiv Q_{(\tau)}(Y_i \mid \mathbf{X}_i)$, $\boldsymbol{\beta}^{(\tau)} = (\beta_1^{(\tau)}, \beta_2^{(\tau)}, \ldots, \beta_k^{(\tau)})'$ is vector of the sensitivity coefficients of the conditional quantile on the changes in values of covariates, and $Q_{(\tau)}(\varepsilon_i^{(\tau)} \mid \mathbf{X}_i) = 0$. A distribution of independent random variables $\varepsilon_i^{(\tau)}$ is left unspecified, which is one of the virtues of the method as far as robustness to outliers is concerned. If $\boldsymbol{\beta}(\tau)$ is independent from τ, then the quantile model collapses to a model $E(Y_i \mid \mathbf{X}_i) = \mathbf{X}_i'\boldsymbol{\beta}$ with a constant variance of an error term. Otherwise, the model implies the variance that a quantile of distribution of Y_i depends on $\mathbf{X}_i$.

The model estimation stage[4] is performed for a given quantile τ. Assuming that observations y_i, $i = 1, 2, \ldots, n$ are treated as a random sample of the regression process $u_i = y_i - \mathbf{x}_i'\boldsymbol{\beta}$ with unknown distribution F, Koenker and Basset (1978) defined a τ-th quantile regression estimator $\mathbf{b}^{(\tau)} = (b_1^{(\tau)}, b_2^{(\tau)}, \ldots, b_k^{(\tau)})'$, which solves the following problem:

$$\min_{\mathbf{b}\in\Re^k}\left[\sum_{i\in\{i:\, y_i \geq \mathbf{x}_i'\mathbf{b}\}} \tau\left|y_i - \mathbf{x}_i'\mathbf{b}^{(\tau)}\right| + \sum_{i\in\{i:\, y_i < \mathbf{x}_i'\mathbf{b}\}} (1-\tau)\left|y_i - \mathbf{x}_i'\mathbf{b}^{(\tau)}\right|\right]. \tag{5}$$

The problem (5) has always a solution and for continuous distributions, it is unique. Since the problem (5) can be transformed to a linear optimization problem its solution can be found using the internal point method (Portnoy, Koenker 1997). The approach is regarded as a nonclassic method due to its robustness. Like robust estimation, the quantile approach detects relationships missed by traditional data analysis. Robust estimates detect the influence of the bulk of the data, whereas quantile estimates detect the influence of covariates on alternate parts of the conditional distribution. Applications of the quantile regression method for the Polish capital market can be found in Trzpiot (2008), Trzpiot (2009a), Trzpiot (2009b), Trzpiot (2010) or Orwat-Acedańska, Trzpiot (2011), among others.

4. Quantile Spatial Autoregressive Model – QSAR model

QSAR model of order τ blends the two approaches mentioned above. It can be written as follows (Kostov 2009; Trzpiot 2012):

$$\mathbf{Y} = \rho^{(\tau)}\mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta}^{(\tau)} + \boldsymbol{\varepsilon}^{(\tau)} \tag{6}$$

where $\mathbf{Y} \equiv Q_{(\tau)}(\mathbf{Y} \mid \mathbf{X})$, $\rho^{(\tau)}$ – quantile spatial autoregression parameter of order τ, $\boldsymbol{\beta}^{(\tau)}$ – vector of the parameters. Vector $\boldsymbol{\varepsilon}^{(\tau)}$ contains independent and identically distributed random variables whose distribution is not specified.

[4] Semi-parametric character of estimation of the model (4) follows from the fact that the error term distribution is left unspecified. Parametric approach is also available provided the error term follows asymmetric Laplace distribution.

Because of the endogeneity problems in the models (6) and (1) (on the right hand side we have spatial lags of the dependent variable $\rho\mathbf{WY}$) their parameters are estimated using instrumental variables procedures (see Chernozhukov, Hansen 2006); Kim, Muller 2004). In the paper, we use the procedure proposed by Kim, Muller (2004). It consists of the following steps:

Estimate the ordinary quantile regression model of order τ for **WY**:

$$\mathbf{WY} = \mathbf{X}\boldsymbol{\beta}^{*(\tau)} + \mathbf{WX}\boldsymbol{\gamma}^{*(\tau)} + \boldsymbol{\varepsilon}^{*(\tau)} \tag{7}$$

1. Calculate the predicted values from (7):

$$\widehat{\mathbf{WY}} = \mathbf{X}\hat{\boldsymbol{\beta}}^{*(\tau)} + \mathbf{WX}\hat{\boldsymbol{\gamma}}^{*(\tau)} \tag{8}$$

2. Use the predicted values as explanatory variable in the original model:

$$\mathbf{Y} = \rho^{(\tau)}\widehat{\mathbf{WY}} + \mathbf{X}\boldsymbol{\beta}^{(\tau)} + \boldsymbol{\varepsilon}^{(\tau)} \tag{9}$$

and estimate its parameters using another ordinary quantile regression by solving the optimization problem (5).

Applications of the spatial quantile regression method and above procedure can be found in Orwat-Acedanska, Trzpiot (2016).

4. Empirical analysis

4.1. Data and the empirical procedure

In the empirical part, we try to identify factors affecting the HLY in $n = 30$ European countries. We include 27 members of the EU (excluded Luxembourg excluded). We work with yearly data where most of the series are from 2014. If unavailable, the previous year is used. The data are taken from Eurostat and WHO database.

The dependent variable in the studied model, namely HLY, is one of the European Structural Indicators monitored by Eurostat. It is also called *Disability-Free Life Expectancy* (*DFLE*) (Gromulska, Wysocki, Goryński 2008; Robine, Jagger, Egidi 2000). The HLY indicator is calculated for men (HLY_M) and women (HLY_W) separately. Its values have been compiled and published since 2004 year for the EU member states (Wróblewska 2008, pp. 153–154). The measure blends data on mortality (age at death) with susceptibility to disease (the age specific proportion of population with and without disabilities). Good health is defined by the lack of limitations resulting from disability. It is calculated as follows:

$$HLY = \frac{\sum_{i=0}^{\omega}\{L_i(1-prev_i)\}}{l_i} \tag{10}$$

where L_i – number of person years lived in the age group i, $prev_i$ – the fraction of disabled persons in age i, l_i – number of survivors of age i.

Selecting the potential exogenous variables, we focus on the health determinants and also socio-economic factors. Eleven exogenous variables are studied: (X_1) – Air pollution – carbon dioxide emission in tons per capita **(AP)**; (X_2) – Education – fraction of population with tertiary education **(E)**; (X_3) – GDP per capita **(GDP)**; (X_4) – Material deprivation – fraction of population with 4 or more important housing items missing **(MD)**; (X_5) – Social protection expenditures to GDP **(SP)**; (X_6) – Population density **(PD)**; (X_7) – Beds in hospitals per 100000 inhabitants **(BH)**; (X_8) – Doctors per 100000 inhabitants **(D)**; (X_9) – Alcohol consumption in liters per capita **(AC)**; (X_{10}) – Cigarettes – fraction of regular smokers in population **(C)**; (X_{11}) – Obesity – fraction of obese inhabitants in population **(OP)**. Data for the most of variables are taken from Eurostat. Some of them, mostly the health determinants, come from the WHO database. Completeness and reliability of the publicly available series served as the primary criteria for selection of the explanatory variables. For example, the data on consumption of fruits and vegetables were not taken into account because of many missing entries, that could not be completed easily. Together with the constant term, matrix X has k = 12 columns and 30 rows.

The parameters are estimated using QSAR model (6) for the following quantiles: τ = 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9. The instrumental variables procedure proposed by Kim, Muller (2004) is employed.[5] Confidence intervals and p-values for the estimates are calculated using the residual bootstrap with 1000 subsamples.

To build the weight matrix **W**, we calculate distances between the centers of the subregions. In the baseline version of the study, we use the four nearest neighbours weight matrix. That is, for a given subregion i, we set w_{ij} = 0,25 if a region j belongs to the four nearest neighbours of i and w_{ij} = 0 otherwise.

[5] We show the results obtained from the Kim, Muller (2004) procedure which are much more stable as far as the spatial autoregression coefficient is concerned compared to those from the Chernozhukov, Hansen (2006).

Alternatively, we also consider the inverse weight matrix: $w_{ij} = 1/d_{ij}$ if $i \neq j$ and $w_{ij} = 0$, otherwise, where d_{ij} denotes the distance between regions i and j.[6]

All computations are carried out in Matlab using the authors' own routines (for example bootstrap estimation) as well as procedures written by Koenker and also LeSage (Spatial Econometric Toolbox).

4.2. Results

We start with the initial spatial analysis of the healthy life years of women (HLY_W) and healthy life years of men (HLY_M) at birth. HLY_W and HLY_M in the investigated countries are depicted on figures 1 and 2 respectively.

Figure 1. Healthy life years of women (HLY_W)

Source: Own calculation based on Eurostat data.

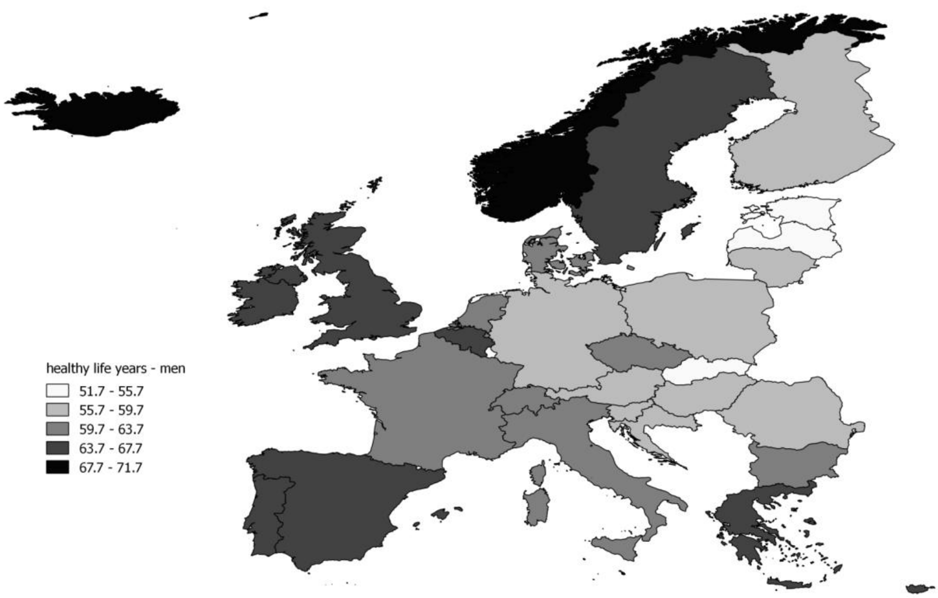
The smallest values of the HLY of women are reported for Latvia (54,2 years), Slovakia (54,3 years) and Finland (56,2 years). The Netherlands, Germany, Romania and Estonia also belongs to this group (the range 54,2–57,9

[6] Because the results for the inverse distance weight matrix are very similar, we do not report them.

years). These countries are coloured in white on figure 1. On the other hand, women in Malta (72,7 years), Norway (68,6 years) and Ireland (68 years) enjoy the longest period of healthy life. Similar values of the HLY are observed in Sweden, Iceland and Bulgaria (69–72,7 years). This group is marked with black colour on figure 1.

The difference in the HLY between women in Latvia and Malta reaches 18,5 years. Interestingly, the HLY is not exactly associated with the geographical location, as the countries belonging to the same groups shown on figure 1 are located in different parts of Europe.

Figure 2. Healthy life years of men (HLY$_M$)



Source: Own calculation based on Eurostat data.

The smallest values of the HLY are observed in Latvia (51,7 years), Estonia (53,9 years) and Slovakia (54,5 years). This group is marked with white colour on figure 2. On the contrary, Iceland (71,7 years), Malta (71,6 years) and Norway (71 years) are the leaders as far as the HLY is concerned. These are coloured in black. For men, the distance in the HLY between Latvia and Malta is equal to 20 years. Like for women, the distribution of the measure does not overlap with the geographical location. This initial analysis confirms our intuition on the substantial heterogeneity of the HLY measure across Europe.

Second stage of the empirical analysis focuses on point and interval estimation of the QSAR model (12) as well as significance testing of the estimates for different variants of the HLY measure.

$$\begin{aligned}\mathbf{HLY}_W &= \rho^{(\tau)}\mathbf{W}\,\mathbf{HLY}_W + \beta_0^{(\tau)} + \beta_1^{(\tau)}\mathbf{AP} + \beta_2^{(\tau)}\mathbf{E} + \beta_3^{(\tau)}\mathbf{GDP} + \beta_4^{(\tau)}\mathbf{MD} + \\ &+ \beta_5^{(\tau)}\mathbf{SP} + \beta_6^{(\tau)}\mathbf{PD} + \beta_7^{(\tau)}\mathbf{BH} + \beta_8^{(\tau)}\mathbf{D} + \beta_9^{(\tau)}\mathbf{AC} + \beta_{10}^{(\tau)}\mathbf{C} + \beta_{11}^{(\tau)}\mathbf{OP} + \boldsymbol{\varepsilon}^{(\tau)}\end{aligned} \qquad (11)$$

The estimates of the parameters together with 90% confidence intervals and p-values are reported below in table 1.

Table 1. Estimates for HLY_w for different quantiles, confidence intervals and p-values for model (11)

	Air pollution	Education	GDP	Material deprivation	Social protection	Population density	Beds in hospitals	Doctors	Alcohol	Cigarettes	Obese population	Autocorr. ceofficient
Abbrev.	AP	**E**	**GDP**	**MD**	**SP**	PD	BH	D	AC	C	OP	
Quantile	**β_1**	β_2	β_3	β_4	β_5	**β_6**	β_7	**β_8**	β_9	**β_{10}**	**β_{11}**	**ρ**
0,1	**-0,186**	0,300	0,000	0,222	-0,149	**0,009**	0,005	**0,051**	-0,640	**-0,461**	**0,155**	**0,982**
0,2	**-0,197**	0,247	0,000	0,168	-0,157	**0,009**	0,008	**0,053**	-0,492	**-0,506**	**0,200**	**0,990**
0,3	**-0,007**	0,151	0,000	0,155	-0,240	**0,007**	0,003	**0,043**	-0,399	**-0,340**	**0,374**	**0,918**
0,4	**-0,123**	0,198	0,000	0,094	-0,193	**0,007**	0,009	**0,034**	-0,595	**-0,224**	**0,383**	**0,817**
0,5	**-0,173**	-0,123	0,000	0,038	-0,161	**0,004**	-0,003	**0,022**	-0,323	**-0,048**	**1,113**	**0,596**
0,6	**-0,083**	-0,082	0,000	0,175	0,039	**0,002**	-0,002	**0,022**	-0,707	**-0,017**	**1,398**	**0,444**
0,7	**-0,159**	0,040	0,000	0,168	0,096	**0,003**	0,002	**0,019**	-0,693	**-0,053**	**1,453**	**0,376**
0,8	**-0,414**	0,109	0,000	0,038	-0,157	**0,006**	0,006	**0,035**	-0,353	**-0,059**	**0,680**	**0,634**
0,9	**-0,414**	0,109	0,000	0,038	-0,157	**0,006**	0,006	**0,035**	-0,353	**-0,059**	**0,680**	**0,634**
Quantile	**p-values**											
0,1	0,275	0,061	0,087	0,017	0,145	0,019	0,211	0	0,101	0,002	0,249	0
0,2	0,264	0,102	0,052	0,056	0,156	0,023	0,097	0	0,15	0	0,185	0
0,3	0,486	0,211	0,122	0,1	0,054	0,038	0,256	0	0,185	0,008	0,049	0
0,4	0,32	0,135	0,385	0,236	0,122	0,015	0,088	0,008	0,113	0,036	0,102	0
0,5	0,207	0,223	0,059	0,36	0,14	0,103	0,292	0,028	0,18	0,291	0,001	0,002
0,6	0,35	0,3	0,121	0,06	0,393	0,264	0,377	0,026	0,04	0,418	0	0,007
0,7	0,231	0,377	0,178	0,082	0,237	0,206	0,331	0,077	0,059	0,343	0	0,018
0,8	0,025	0,227	0,206	0,434	0,19	0,057	0,162	0,006	0,18	0,306	0,007	0,001
0,9	0,029	0,232	0,206	0,447	0,205	0,059	0,173	0,007	0,193	0,328	0,009	0,001

Quantile	**Lover bounds of 90% confidence intervals**											
0,1	-0,604	-0,022	-0,001	0,061	-0,495	0,003	-0,006	0,029	-1,451	-0,709	-0,294	0,767
0,2	-0,625	-0,092	-0,001	-0,009	-0,466	0,002	-0,003	0,031	-1,402	-0,758	-0,240	0,809
0,3	-0,481	-0,176	-0,001	-0,044	-0,533	0,001	-0,008	0,023	-1,164	-0,574	0,006	0,681
0,4	-0,623	-0,095	0,000	-0,112	-0,480	0,002	-0,002	0,013	-1,429	-0,454	-0,104	0,533
0,5	-0,617	-0,361	0,000	-0,159	-0,424	-0,001	-0,013	0,003	-0,985	-0,239	0,589	0,347
0,6	-0,490	-0,322	0,000	-0,016	-0,220	-0,004	-0,012	0,003	-1,355	-0,204	0,867	0,190
0,7	-0,596	-0,200	0,000	-0,029	-0,180	-0,003	-0,008	-0,002	-1,376	-0,244	0,912	0,070
0,8	-0,896	-0,134	0,000	-0,195	-0,401	0,000	-0,005	0,013	-1,045	-0,261	0,248	0,370
0,9	-0,906	-0,150	0,000	-0,202	-0,394	0,000	-0,005	0,011	-1,044	-0,262	0,220	0,354
Quantile	**Upper bounds of 90% confidence intervals**											
0,1	0,374	0,587	0,000	0,524	0,122	0,016	0,020	0,076	0,170	-0,254	0,696	0,990
0,2	0,380	0,541	0,000	0,460	0,179	0,016	0,025	0,077	0,355	-0,297	0,889	0,990
0,3	0,554	0,387	0,000	0,392	0,002	0,013	0,017	0,065	0,369	-0,128	0,938	0,990
0,4	0,364	0,470	0,000	0,315	0,098	0,013	0,022	0,056	0,201	-0,029	0,883	0,990
0,5	0,219	0,125	0,001	0,215	0,103	0,008	0,007	0,040	0,354	0,125	1,580	0,882
0,6	0,285	0,178	0,000	0,368	0,306	0,006	0,008	0,038	-0,051	0,165	1,853	0,736
0,7	0,247	0,307	0,000	0,365	0,396	0,007	0,013	0,035	0,069	0,139	1,943	0,667
0,8	-0,090	0,359	0,000	0,203	0,131	0,010	0,016	0,050	0,354	0,128	1,142	0,912
0,9	-0,074	0,342	0,000	0,204	0,147	0,010	0,017	0,050	0,396	0,147	1,165	0,932

The bolded estimates are statistically significant (α = 0,05) for at least two quantiles.

Source: Own calculation.

The following variables turned out to be insignificant for any studied quantiles: the socio-economic factors: Education **(E)**, **GDP**, Material deprivation (**MD)**, Social protection (**SP)**, and the healthcare factors: Beds in hospitals (**BH)**, alcohol consumption **(AC)**. Therefore, we respecified the model by excluding those variables. As a result, we were left with the following model:

$$\mathbf{HLY}_W = \rho^{(\tau)}\mathbf{W}\,\mathbf{HLY}_W + \beta_0^{(\tau)} + \beta_1^{(\tau)}\mathbf{AP} + \beta_2^{(\tau)}\mathbf{PD} + \beta_3^{(\tau)}\mathbf{D} + \beta_4^{(\tau)}\mathbf{C} + \beta_5^{(\tau)}\mathbf{OP} + \boldsymbol{\varepsilon}^{(\tau)} \quad (12)$$

The point estimates of the model (12) together with the p-values of the significance tests for the analyzed quantiles are shown in table 2. The associated confidence intervals are depicted on figure 3, where the solid lines represent the point estimates and the 90% confidence interval bounds are marked by the dotted lines.

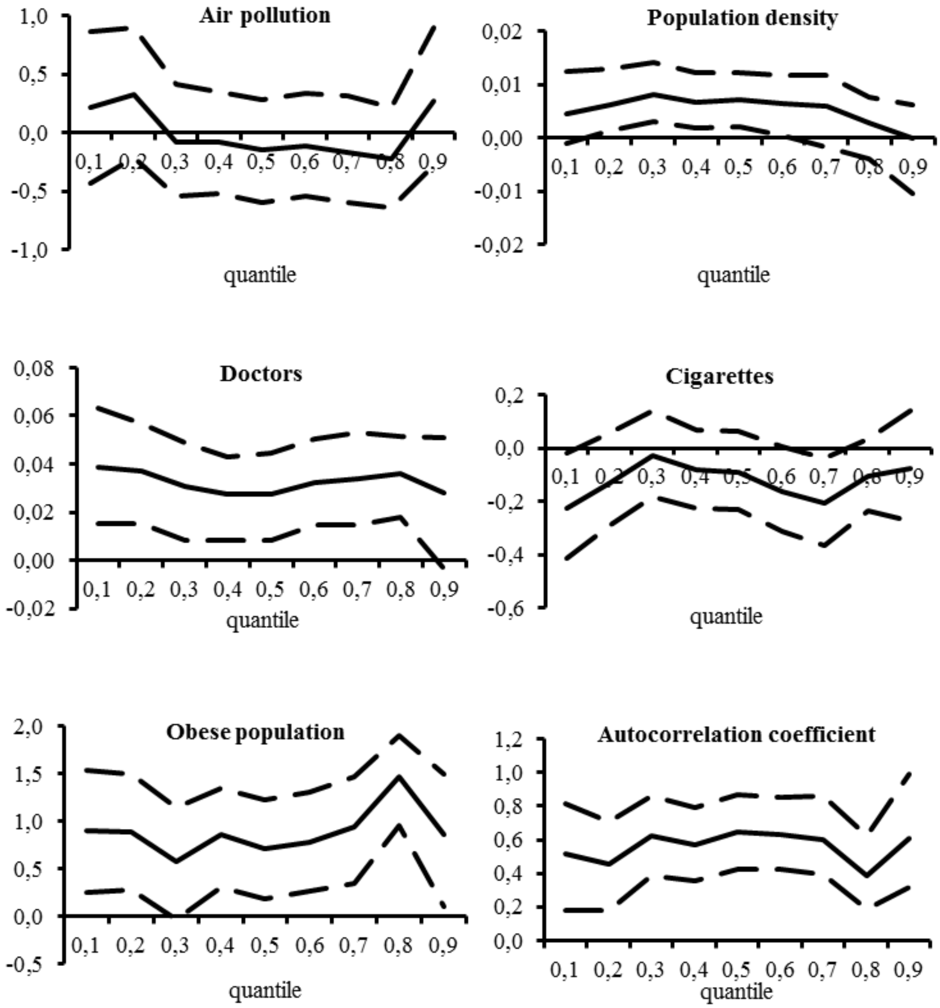
Table 2. Estimates for HLY_w for different quantiles, confidence intervals and p-values for model (12)

	Air pollution	Population density	Doctors	Cigarettes	Obese population	Autocorrelation coefficient
Abbreviation	AP	PD	D	C	OP	
Quantile	β_1	β_2	β_3	β_4	β_5	ρ
0,1	0,222	0,004	0,039	-0,224	0,893	0,517
0,2	0,330	0,006	0,037	-0,128	0,880	0,458
0,3	-0,084	0,008	0,031	-0,030	0,568	0,622
0,4	-0,079	0,007	0,028	-0,080	0,861	0,566
0,5	-0,143	0,007	0,027	-0,091	0,705	0,645
0,6	-0,108	0,006	0,032	-0,161	0,771	0,630
0,7	-0,170	0,006	0,034	-0,206	0,934	0,600
0,8	-0,219	0,003	0,036	-0,103	1,467	0,382
0,9	0,275	0,000	0,028	-0,076	0,862	0,609
Quantile	**p-values**					
0,1	**0,259**	**0,071**	0,003	0,04	0,025	0,006
0,2	**0,129**	0,033	0,001	**0,134**	0,011	0,007
0,3	**0,361**	0,006	0,011	**0,405**	0,059	0
0,4	**0,356**	0,021	0,015	**0,177**	0,003	0
0,5	**0,285**	0,017	0,008	**0,144**	0,011	0
0,6	**0,332**	0,045	0,003	**0,058**	0,007	0
0,7	**0,284**	**0,112**	0,001	0,022	0,004	0
0,8	**0,163**	**0,3**	0,001	**0,091**	0	0,006
0,9	**0,214**	**0,336**	0,05	**0,271**	0,025	0,001

Bolded are the p-values greater than 0,05.

Source: Own calculation.

Figure 3. Parameter estimates with 90% confidence intervals for model (12)



Source: Own calculation.

The variables that were found insignificant in the first step of the empirical analysis (E, GDP, MD, SP, BH and AC) can be mostly labelled as the socio-economic indicators. These variables seem not to affect the HLY for women for all considered quantiles, which means that for all countries regardless of the length of disability-free life. The remaining variables are the health determinants and the healthcare factors (population density is the only exception). The impact of these variables on the response is different for the different quantiles. More specifically:

- **Population density** matters primarily for the countries with the moderate values of the HLY for women (the variable is significant for the quantiles 0,2–0,6). It seems not to affect the dependent variables in countries with the smallest values (the Netherlands, Germany, Slovakia, Romania, Latvia, Estonia and Finland) and the highest values of the HLY (Malta, Iceland, Norway, Sweden and Bulgaria).
- The **number of doctors** is significantly associated with the HLY for all the studied countries – the magnitude of its impact is approximately constant regardless of the quantiles considered.
- The variable **cigarettes** is significant for the quantiles 0,1 and 0,7, which means that smoking has the strongest negative effect on the HLY in the countries with the smallest (the Netherlands, Germany, Slovakia, Romania, Latvia, Estonia and Finland) and moderately high values of the response.
- There is also a significant correlation between **obesity** and the HLY for all studied countries.
- Additionally, the spatial **autocorrelation parameter** is statistically significant for all the studied quantiles which justify utilization of the spatial autocorrelation models for investigating the factors associated with the HLY for women.

We conduct the similar analysis for the HLY of men starting with the following specification:

$$\mathbf{HLY}_M = \rho^{(\tau)}\mathbf{W}\,\mathbf{HLY}_M + \beta_0^{(\tau)} + \beta_1^{(\tau)}\mathbf{AP} + \beta_2^{(\tau)}\mathbf{E} + \beta_3^{(\tau)}\mathbf{GDP} + \beta_4^{(\tau)}\mathbf{MD} + \\ + \beta_5^{(\tau)}\mathbf{SP} + \beta_6^{(\tau)}\mathbf{PD} + \beta_7^{(\tau)}\mathbf{BH} + \beta_8^{(\tau)}\mathbf{D} + \beta_9^{(\tau)}\mathbf{AC} + \beta_{10}^{(\tau)}\mathbf{C} + \beta_{11}^{(\tau)}\mathbf{OP} + \boldsymbol{\varepsilon}^{(\tau)} \tag{13}$$

The estimates of the parameters together with 90% confidence intervals and p-values are depicted below.

Table 3. Estimates for *HLY*

	Air pollution	Education	GDP	Material deprivation	Social protection	Population density	Beds in hospitals	Doctors	Alcohol consumption	Cigarettes	Obese population	Autocorrelati on ceofficient
Abbrev.	AP	E	GDP	MD	SP	PD	BH	D	AC	C	OP	
Order	β_1	β_2	β_3	β_4	β_5	β_6	β_7	$\boldsymbol{\beta_8}$	$\boldsymbol{\beta_9}$	$\boldsymbol{\beta_{10}}$	$\boldsymbol{\beta_{11}}$	$\boldsymbol{\rho}$
0,1	-0,050	0,129	0,000	0,095	-0,382	0,005	0,002	**0,064**	**-0,859**	**-0,440**	**0,626**	**0,860**
0,2	-0,023	0,166	0,000	0,083	-0,334	0,004	0,003	**0,062**	**-0,975**	**-0,402**	**0,653**	**0,816**
0,3	-0,068	0,251	0,000	0,071	-0,106	0,006	0,008	**0,036**	**-1,220**	**-0,212**	**0,384**	**0,870**
0,4	-0,090	0,027	0,000	0,096	-0,028	0,004	0,003	**0,027**	**-0,780**	**-0,167**	**0,876**	**0,689**
0,5	-0,237	0,083	0,000	0,004	-0,063	0,004	0,006	**0,036**	**-1,057**	**-0,302**	**0,744**	**0,912**

0,6	-0,143	0,022	0,000	0,139	-0,073	0,003	-0,007	**0,020**	**-0,676**	**-0,133**	**0,902**	**0,736**
0,7	-0,166	-0,030	0,000	0,142	-0,099	0,003	-0,010	**0,019**	**-0,513**	**-0,116**	**0,812**	**0,790**
0,8	-0,439	0,131	0,000	-0,040	-0,451	0,004	-0,006	**0,039**	**-0,166**	**-0,165**	**0,667**	**0,849**
0,9	-0,488	0,104	0,000	-0,105	-0,407	0,004	0,000	**0,044**	**-0,397**	**-0,114**	**0,501**	**0,876**
Order	**p-values**											
0,1	0,45	0,26	0,377	0,162	0,019	0,099	0,34	**0**	**0,029**	**0,001**	**0,006**	**0**
0,2	0,489	0,179	0,336	0,199	0,036	0,094	0,225	**0**	**0,02**	**0,001**	**0,004**	**0**
0,3	0,408	0,06	0,288	0,281	0,206	0,025	0,096	**0,002**	**0,002**	**0,029**	**0,041**	**0**
0,4	0,346	0,396	0,372	0,203	0,376	0,068	0,274	**0,011**	**0,028**	**0,069**	**0,001**	**0**
0,5	0,193	0,306	0,182	0,475	0,4	0,107	0,193	**0,004**	**0,01**	**0,009**	**0,001**	**0**
0,6	0,235	0,408	0,078	0,111	0,332	0,172	0,128	**0,06**	**0,057**	**0,12**	**0,001**	**0**
0,7	0,183	0,444	0,05	0,105	0,28	0,167	0,044	**0,062**	**0,108**	**0,144**	**0,002**	**0**
0,8	0,022	0,155	0,127	0,272	0,008	0,138	0,149	**0,003**	**0,266**	**0,064**	**0,009**	**0**
0,9	0,022	0,254	0,187	0,119	0,019	0,128	0,491	**0,001**	**0,173**	**0,153**	**0,02**	**0**
Order	**Lover bounds of 90% confidence intervals**											
0,1	-0,526	-0,196	0,000	-0,073	-0,676	-0,002	-0,009	0,044	-1,660	-0,690	0,229	0,570
0,2	-0,463	-0,129	0,000	-0,096	-0,616	-0,002	-0,006	0,043	-1,810	-0,638	0,255	0,524
0,3	-0,509	-0,014	0,000	-0,124	-0,388	0,001	-0,002	0,016	-1,988	-0,405	0,031	0,647
0,4	-0,511	-0,230	0,000	-0,098	-0,307	0,000	-0,006	0,008	-1,546	-0,386	0,421	0,431
0,5	-0,653	-0,196	0,000	-0,182	-0,304	-0,001	-0,006	0,016	-1,757	-0,505	0,299	0,713
0,6	-0,633	-0,217	0,000	-0,067	-0,341	-0,002	-0,018	-0,002	-1,388	-0,335	0,404	0,485
0,7	-0,643	-0,262	0,000	-0,060	-0,368	-0,002	-0,021	-0,001	-1,155	-0,311	0,355	0,564
0,8	-0,994	-0,122	0,000	-0,302	-0,627	-0,002	-0,015	0,018	-0,848	-0,358	0,311	0,650
0,9	-1,025	-0,163	0,000	-0,430	-0,614	-0,002	-0,011	0,021	-1,084	-0,287	0,132	0,652
Order	**Upper bounds of 90% confidence intervals**											
0,1	0,505	0,378	0,000	0,341	-0,087	0,011	0,014	0,090	-0,114	-0,221	1,202	0,990
0,2	0,507	0,421	0,000	0,314	-0,053	0,010	0,017	0,085	-0,232	-0,195	1,211	0,990
0,3	0,363	0,503	0,000	0,286	0,167	0,012	0,021	0,057	-0,544	-0,027	0,856	0,990
0,4	0,310	0,291	0,000	0,297	0,250	0,009	0,015	0,047	-0,146	0,017	1,383	0,971
0,5	0,186	0,368	0,000	0,226	0,247	0,009	0,018	0,057	-0,332	-0,116	1,225	0,990
0,6	0,241	0,290	0,001	0,327	0,181	0,008	0,003	0,038	0,021	0,060	1,377	0,982
0,7	0,187	0,236	0,001	0,337	0,195	0,008	0,000	0,038	0,194	0,064	1,256	0,990
0,8	-0,108	0,384	0,000	0,117	-0,182	0,009	0,005	0,057	0,467	0,017	1,131	0,990
0,9	-0,126	0,342	0,000	0,058	-0,093	0,009	0,010	0,063	0,387	0,084	0,983	0,990

The bolded estimates are statistically significant ($\alpha = 0{,}05$) for at least two quantiles.

Source: Own calculation.

The results show that the following variables were not statistically significant for any of the studied quantiles: air pollution **(AP)**, education **(E)**, **GDP**, material deprivation (**MD)**, social protection (**SP)**, population density **(PD)** and beds in hospitals (**BH)**. The last variable belongs to the healthcare factors, while the remaining to the socio-economic factors. Therefore, these variables were excluded and the model was respecified in the following form:

$$\mathbf{HLY}_M = \rho^{(\tau)}\mathbf{W}\,\mathbf{HLY}_M + \beta_0^{(\tau)} + \beta_1^{(\tau)}\mathbf{D} + \beta_2^{(\tau)}\mathbf{AC} + \beta_3^{(\tau)}\mathbf{C} + \beta_4^{(\tau)}\mathbf{OP} + \boldsymbol{\varepsilon}^{(\tau)} \quad (14)$$

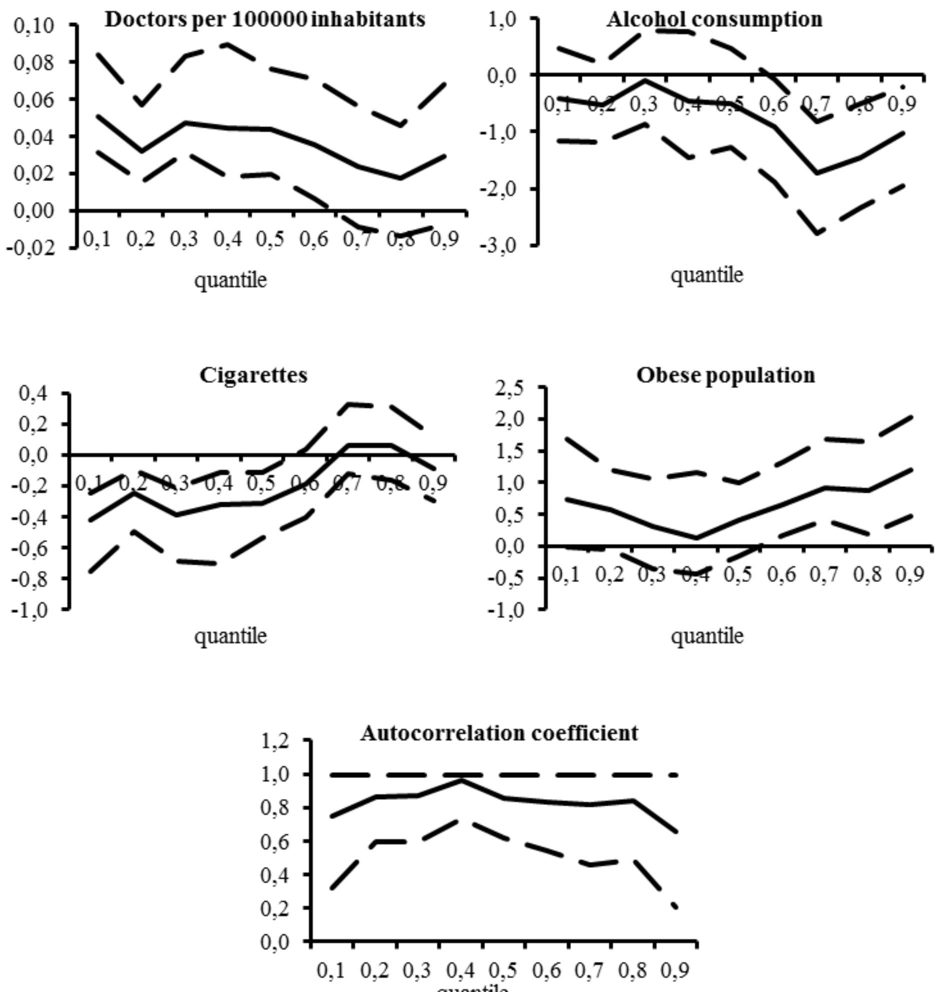
The point estimates of the model (14) with the associated p-values are reported in table 4. The 90% confidence intervals are depicted on figure 4.

Table 4. Estimates for HLY_M for different quantiles, confidence intervals and p-values for model (14)

	Doctors	Alcohol consumption	Cigarettes	Obese population	Autocorrelation ceofficient
Abbreviation	D	AC	C	OP	
Quantile	β_1	β_2	β_3	β_4	ρ
0,1	0,051	-0,408	-0,420	0,739	0,751
0,2	0,032	-0,518	-0,249	0,572	0,860
0,3	0,047	-0,102	-0,385	0,315	0,867
0,4	0,045	-0,447	-0,324	0,133	0,960
0,5	0,044	-0,493	-0,309	0,406	0,858
0,6	0,036	-0,902	-0,190	0,659	0,831
0,7	0,023	-1,720	0,062	0,917	0,813
0,8	0,017	-1,443	0,061	0,866	0,840
0,9	0,029	-1,025	-0,092	1,203	0,658
Quantile	**p-values**				
0,1	0	0,26	0	0,052	0,005
0,2	0,005	0,137	0,007	0,064	0
0,3	0	0,484	0	0,236	0
0,4	0,008	0,273	0,013	0,263	0
0,5	0,001	0,196	0,009	0,14	0
0,6	0,02	0,041	0,086	0,013	0,001
0,7	0,146	0,002	0,223	0,004	0
0,8	0,24	0,006	0,283	0,015	0
0,9	0,111	0,016	0,26	0,003	0,004

Source: Own calculation.

Figure 4. Parameter estimates with 90% confidence intervals for model (14)



Source: Own calculation.

Like for women, most of the socio-economic indicators are found not to be significantly associated with the HLY for men for all the studied quantiles (that is, for all the countries considered). This group include: Air pollution (AP), Education (E), GDP, Material deprivation (MD), Social protection (SP), Population density (PD) and Beds in hospitals (BH). On the other hand, most of the health determinants and healthcare factors do affect the response, although their impact varies with the analyzed quantiles of the HLY distribution. Specifically:

- The **number of doctors** is positively and significantly correlated with the HLY of men in countries with at most moderate level of the dependent variable (quantiles 0,2–0,6), but does not affect it in countries where the dependent variable takes the highest values (Malta, Iceland, Norway, Sweden, Ireland and Spain).
- The measure of **consumption of alcohol** is significantly associated with the HLY only for the higher quantiles 0,6–0,9, which means that it negatively affects the HLY in the countries with the longest disability-free life.
- On the other hand, the variable **smoking** is statistically significant for all the quantiles but the highest ones mentioned in the previous point. The impact of the variable on the HLY is negative.
- The measure of **obesity** significantly affects the HLY for the quantiles 0,1 and 0,6–0,9 which means that obesity has the highest impact on the HLY in the countries with the shortest (Slovakia, Latvia and Estonia), moderately high and the highest length of healthy life.
- The spatial **autocorrelation parameter** is statistically significant for all the studied quantiles. Its high values confirm the need to include the spatial autocorrelation effects into the models of the HLY of men.

5. Conclusion

The correct identification and quantification of factors associated with the length of disability-free life is indispensable for conducting successful health and social policy in the EU. We investigated several socio-economic indicators and health determinants as potential factors explaining the observed differences in the HLY between European countries. Because these differences are significant (the range is equal to 20 years for men and 18,5 years for women) we believe that the relationship between the HLY and the factors may vary across the countries with different lengths of disability-free life. As a result, we used the approach that blends the spatial autoregression model with quantile regression to analyze the factors associated with the HLY.

The estimated structural and spatial autocorrelation parameters do vary across the different quantiles. For example for the median regression (quantile 0,5), some factors turned out to be insignificant, while considering other quantiles we observed significant relationships between the HLY and those factors. Our results show that the health determinants related to lifestyle are primarily associated with the observed differences in the HLY, while the impact of the considered socio-economic factors is small and in most cases statistically insignificant regardless of sex and the considered quantiles.

It should be noted that we did not find a significant relationship between the HLY and GDP. At first sight, it might be surprising but accounting for the low level of the HLY in such rich countries like Germany, Austria or Finland it is no longer unexpected. Additionally, the health determinants which in most cases were found significantly associated with the HLY are also correlated with GDP.

We also found that the impact of the health determinants on the HLY varies across the countries. More specifically, population density affects the HLY of women only in the countries with the moderate length of disability-free life but it is not significantly correlated with the HLY in the countries with the lowest levels of the HLY. The same variable seems not to affect the HLY of men at all. Consumption of alcohol has the most negative impact on the healthy life length of men only in the countries with the longest disability-free life. We did not find such a relationship for the other countries as well as women. Smoking shortens healthy life of women the most in the countries where the HLY is the lowest as well as high. For men, the negative impact is observed in all the countries but these with the highest HLY. The rest of the considered health determinants, namely obesity and the number of doctors per 10000 inhabitants, are significantly associated with the HLY regardless of sex and for almost all the quantile studied.

These results strongly support our verified hypothesis that the spatial autoregression models for the healthy life years should be estimated and interpreted using quantile regression approach.

References

Anselin L. (1988), *Spatial Econometrics: Methods and Models*, Springer, Berlin-Heidelberg.

Chernozhukov V., Chansen C. (2006), *Instrumental Quantile Regression Inference for Structural And Treatment Effect Models*, 'Journal of Econometrics', Elsevier, Vol. 127.

Gromulska L., Wysocki M., Goryński P. (2008), *Lata przeżyte w zdrowiu (Healthy Life Years, HLY) – zalecany przez Unię Europejską syntetyczny wskaźnik sytuacji zdrowotnej ludności*, 'Przegląd Epidemiologiczny', PZH, Vol. 62(4).

Kim T. H., Muller C. (2004), *Two-stage quantile regression when the first stage is based on quantile regression*, 'Econometrics Journal', Wiley, Vol. 7.

Koenker R., Bassett B. (1978), *Regression Quantiles*, 'Econometrica', Wiley, Vol. 46.

CSO (2015), *Life Expectancy Tables of Poland in 2014*, Central Statistical Office, Warsaw.

Kostov P. (2009), *A spatial quantile regression hedonic model of agricultural land prices*, 'Spatial Economic Analysis', Taylor & Francis, Vol. 4(1).

Lee L. F. (2002), *Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models*, 'Econometric Theory', Cambridge University Press, Vol. 18(2).

Lee L. F. (2007), *GMM and 2SLS estimation of mixed regressive, spatial autoregressive models*, 'Journal of Econometrics', Elsevier, Vol. 137(2).

LeSage J. P., Pace R. K. (2009), Introduction to spatial econometrics, CRC Press, Boca Raton.

LeSage J. (1997), *Bayesian estimation of spatial autoregressive models*, 'International Regional Science Review', SAGE, Vol. 20(1–2).

Lum K., Gelfand A. (2012), *Spatial quantile multiple regression using the asymmetric Laplace process*, 'Bayesian Analysis', International Society for Bayesian Analysis, Vol. 7(2).

Orwat-Acedańska A., Trzpiot G. (2011), *The classification of Polish mutual balanced funds on the management style – quantile regression approach*, 'Theory and applications of quantitative methods. Econometrics', University of Economics in Wrocław, Vol. 31(194).

Orwat-Acedańska A., Trzpiot G. (2016), *Spatial quantile regression in analysis of mortality*, 'Folia Oeconomica', University of Łódź, in print.

Portnoy S., Koenker R. (1997), *The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error Versus Absolute-Error Estimators*, 'Statistical Science', Institute of Mathematical Statistics, Vol. 12.

Robine J-M., Jagger C., Egidi V. (2000), *Selection of a Coherent Set of Health Indicators. A First Step Towards A User's Guide to Health Expectancies for the European Union,* Euro-REVES, Montpellier.

Suchecki B. (2010), *Spatial Econometrics*, Beck, Warszawa.

Trzpiot G. (2008), *The implementation of quantile regression methodology in VaR estimation (in polish),* 'Studies and researches of faculty of economics and management' University of Szczecin, 316–323.

Trzpiot G. (2009a), *Quantile regression model versus factor model estimation,* 'Financial investments and insurances – world trends and polish market', University of Economics in Wrocław, Vol. 60.

Trzpiot G. (2009b), *Estimation methods for quantile regression*, 'Economic Studies', University of Economics in Katowice, Vol. 53.

Trzpiot G. (2010), *Quantile regression model of return rate relation – volatility for some Warsaw Stock Exchange indexes, (in polish),* 'Finances, financial markets and insurances. Capital market', University of Szczecin, Vol. 28.

Trzpiot G. (2012), *Spatial Quantile Regression*, 'Comparative Economic Research. Central and Eastern Europe', University of Łódź, Vol. 15(4).

Wróblewska W. (2008), *Sumaryczne miary stanu zdrowia populacji*, 'Studia Demograficzne', PAN, Vol. 153–154(1–2).

Streszczenie

PRZESTRZENNA REGRESJA KWANTYLOWA W ANALIZIE DŁUGOŚCI ŻYCIA W KRAJACH UNII EUROPEJSKIEJ

Celem pracy jest badanie wpływu wybranych czynników na średnią długość życia z zdrowiu kobiet i mężczyzn w krajach UE. Ze względu na fakt, że kraje Unii Europejskiej charakteryzuje silne zróżnicowanie pod względem średniej długości życia w zdrowiu oraz jakości życia obywateli, stosujemy w pracy modele wielorakiej kwantylowej autoregresji przestrzennej. Regresja kwantylowa umożliwia analizę zależności pomiędzy zmiennymi w różnych kwantylach rozkładu zmiennej niezależnej. Ponadto narzędzie to jest odporne na założenie klasycznej regresji dotyczące postaci wielowymiarowego rozkładu składnika losowego. Estymacji punktowej parametrów modeli dokonano przy użyciu zmiennych instrumentalnych (Kim, Muller 2004), natomiast do estymacji przedziałowej i weryfikacji hipotezy istotności parametrów wykorzystano metodę bootstrap.

Słowa kluczowe: *regresja kwantylowa, wieloraka kwantylowa autoregresja przestrzenna, analiza przestrzenna, długość życia w zdrowiu*