

*Grażyna Trzpiot** *Agnieszka Orwat-Acedańska***

SPATIAL QUANTILE REGRESSION IN ANALYSIS OF MORTALITY

Abstract. The paper concerns mortality in Poland. The aim of the article is to identify factors determining average life length of men and women in 66 subregions of Poland. We use spatial quantile regression methodology which 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 multidimensional distribution of error term.

Keywords: quantile regression, analysis of mortality, spatial analysis.

JEL: J11, C14, C30.

1. INTRODUCTION

Since the early 1990 we have observed a decline in mortality rates in Poland coupled with a rise in the average life length. In 2014, men live 73.8 years on average and women – 81.6 years. The life length of men and women increased by 7.6 and 6.4 years, respectively, compared to 1990 and by 4.1 and 3.6 years (Central Statistical Office, 2015), compared to 2000. The observed life prolongation is fuelled by developments in medicine technology and the use of novel diagnostic tools. In Poland, as in other countries, the excessive mortality of men is observed, but its scale is considerably higher. Currently, the difference between the average life length of men and women in Poland is 7.8 years.

Additionally, significant spatial differences in the average life length in Poland are reported. Health inequalities seen from both social and spatial perspective constitute a severe and increasingly important problem. Finding solution to the problem is one of the priorities of public health programs. The goal of the National Health Program for the years 2016–2020 is to reduce the spatial inequalities in health so as to increase the life length and improve health.

From this perspective, the analysis of factors determining the average life length across spatial units emerges as a key problem. In other words, the

* University of Economics in Katowice, grazyna.trzpiot@ue.katowice.pl

** University of Economics in Katowice, agnieszka.orwat@ue.katowice.pl

mortality analysis in Poland should focus on the correct identification of the spatial socio-economic factors and quantification of risks affecting morbidity and mortality. Spatial autoregression model (SAR) is one of the statistical models allowing identification of these factors from the spatial perspective. The SAR model, similarly to the classical regression, is a conditional expected value model for a dependent variable. It describes a relationship between an outcome and covariates in the middle of the outcome distribution. As such, it disregards the dependencies in other parts of the distribution, particularly the tails, and gives only a limited perspective on the examined relationships. Because of the spatial differences in the average life length we extend the SAR model using quantile regression (QSAR). The quantile estimation of the spatial model put some more light on the spatial dependencies in different parts of the average life length distribution.

The aim of the paper is to identify the factors affecting the average life length in subregions of Poland using the quantile spatial autoregression model.

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. In section three we blend the two methodologies and introduce the QSAR model. The fourth section, which contains the empirical analysis, 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 MODEL¹

SAR model has the following form (Anselin 1988):

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

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 the observed value of a variable in one localization and its value in the other localization (region, for example). Spatial autoregression is represented by the

¹ In the literature, a few names of the model are used interchangeably. Using SAR term, we follow (LeSage, Pace 2009).

spatial lag term $\rho\mathbf{WY}$ of the dependent variable. Vector of the error terms has a multivariate normal distribution:²

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

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 (see Anselin 1988), generalized method of moments or two-stage least squares (see Lee 2007). For large-scale spatial models, Bayesian estimation is also employed (see LeSage 1997; 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 insight into dependencies in the other parts of the distribution. This is not the case for quantile regression which allows studying an impact of covariates on the outcome in any point of the outcome distribution.

3. QUANTILE REGRESSION MODEL – QR MODEL

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

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

$$Q_Y(\tau | \mathbf{X}) = F^{-1}(\tau | \mathbf{X}), \quad (3)$$

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

² In the paper, we do not consider the model with correlated error terms. The quantile versions of such a model can be estimated using Bayesian methods only.

$$Y_i = \mathbf{X}_i' \boldsymbol{\beta}^{(\tau)} + \varepsilon_i^{(\tau)}, \quad (5)$$

where $Y_i \equiv Q_{(\tau)}(Y_i | \mathbf{X}_i)$, $\boldsymbol{\beta}^{(\tau)} = (\beta_1^{(\tau)}, \beta_2^{(\tau)}, \dots, \beta_k^{(\tau)})'$ is vector of the sensitivity coefficients of the conditional quantile on the changes in values of covariates, and $Q_{\tau}(\varepsilon_i^{(\tau)} | \mathbf{x}_i) = 0$. 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 turns into to a model $E(Y_i | \mathbf{X}_i) = \mathbf{X}_i' \boldsymbol{\beta}$ 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's estimation stage³ is performed for a given quantile of order τ . Assuming that observations y_i , $i=1,2,\dots,n$ are treated as a random sample of the regression process $u_i = y_i - \mathbf{x}_i' \boldsymbol{\beta}$ with unknown distribution \mathfrak{D} , (Koenker, Basset 1978) defined a τ -th quantile regression estimator $\mathbf{b}^{(\tau)} = (b_1^{(\tau)}, b_2^{(\tau)}, \dots, b_k^{(\tau)})'$, which solves the following problem:

$$\min_{\mathbf{b} \in \mathfrak{R}^k} \left[\sum_{i \in \{i: y_i \geq \mathbf{x}_i' \mathbf{b}\}} \tau |y_i - \mathbf{x}_i' \mathbf{b}^{(\tau)}| + \sum_{i \in \{i: y_i < \mathbf{x}_i' \mathbf{b}\}} (1-\tau) |y_i - \mathbf{x}_i' \mathbf{b}^{(\tau)}| \right] \quad (6)$$

The problem (6) always has a solution and for continuous distributions, it is unique. Since the problem (6) 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 co-variates on various parts of the conditional distribution. Applications of the quantile regression method for the Polish capital market can be found among others in (Trzpiot 2008; Trzpiot 2009a; Trzpiot 2009b; Trzpiot 2010; Orwat-Acedańska, Trzpiot 2011).

³ 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.

4. QUANTILE SPATIAL AUTOREGRESSIVE MODEL – QSAR MODEL

QSAR model of order τ blends the two approaches mentioned above. It can be written as follows:

$$\mathbf{Y} = \rho^{(\tau)} \mathbf{WY} + \mathbf{X}\boldsymbol{\beta}^{(\tau)} + \boldsymbol{\varepsilon}^{(\tau)} \quad (7)$$

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

Because of the endogeneity problems in the models (7) 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:

1. Estimate the ordinary quantile regression model of order τ for \mathbf{WY} :

$$\mathbf{WY} = \mathbf{X}\boldsymbol{\beta}^{*(\tau)} + \mathbf{WX}\boldsymbol{\gamma}^{*(\tau)} + \boldsymbol{\varepsilon}^{*(\tau)} \quad (8)$$

2. Calculate the predicted values from (8):

$$\hat{\mathbf{WY}} = \mathbf{X}\hat{\boldsymbol{\beta}}^{*(\tau)} + \mathbf{WX}\hat{\boldsymbol{\gamma}}^{*(\tau)} \quad (9)$$

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

$$\mathbf{Y} = \rho^{(\tau)} \hat{\mathbf{WY}} + \mathbf{X}\boldsymbol{\beta}^{(\tau)} + \boldsymbol{\varepsilon}^{(\tau)} \quad (10)$$

and estimate its parameters using another ordinary quantile regression by solving the optimization problem (6).

5. EMPIRICAL ANALYSIS

5.1. Data and the empirical procedure

In the empirical part we try to identify factors affecting the average life length in $n = 66$ Polish subregions.⁴ We work with yearly data where most of the

⁴ The concept of a subregion relates to the NUTS classification. The subregions in Poland are defined by the resolution of the cabinet on 14.03.2002 (Dz. U. from 2002 r. nr 34, poz. 311). Every region (voivodeship) consists of a few subregions, which subsequently are formed from several counties.

series are from 2014. If unavailable, the previous year is used. The data are taken from Regional database of the Polish Main Statistical Office. Because of the differences in the average life length of men and women these two variables, denoted by Y_m and Y_w , respectively, are studied separately as the endogenous variables.

Selecting the potential exogenous variables, we focus not only on the diseases directly affecting the life length, but also take into account the socio-economic conditions. They should be considered because of the significant spatial differences in the life length in Polish subregions. Twelve exogenous variables are studied: density of population (DP), per capita earnings (E), unemployment rate (UR), gross value of fixed assets (GVA), the rate of urbanization (RU), the forestation rate (FR), per capita air pollutant emissions (APE), number of inhabitants per a hospital bed (IH), number of doctors per capita (ND), the fraction of deaths from cardiovascular diseases (CD), the fraction of deaths from cancers (C) and respiratory diseases (RD). Adding the constant term, matrix \mathbf{X} has $k = 13$ columns and 66 rows. All exogenous variables are standardized.

We have to discard the variables related to the lifestyle like consumption of fruits and vegetables, alcohol, tobacco or the fraction of obese persons, because such data are not spatially decomposed.

The parameters are estimated using QSAR model (7) for the following orders: $\tau = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. The instrumental variables procedure proposed by (Kim, Muller 2004) is employed.⁵ Confidence intervals 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.

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

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

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

5.2. Results

We start with the descriptive analysis of the dependent variables. The average life length of men and women in Polish subregions is depicted in Figures 1 and 2.

Generally, for both sexes, the longest life length is observed in the biggest cities and south-eastern subregions, while the shortest – in the northern and the north-western ones. More specifically, the longest life length of men is observed in Warsaw and Trójmiasto subregions (more than 71 years), as well as in the following subregions: łomżyński, sandomiersko-jędrzejowski, przemyski, rzeszowski, krośnieński, tarnowski, Kraków, Wrocław, Poznań, Szczecin (more than 70 years). On the other hand, the following subregions: ełcki, starogardzki, gdański, szczeciński and rybnicki are characterized by the shortest life length (less than 67 years). As far as women are concerned, the longest life length is observed in the following subregions: łomżyński, sandomiersko-jędrzejowski (more than 80 years), ostrołęcko-siedlecki, bialski, chełmsko-zamojski, puławski, radomski, rzeszowski, krośnieński, tarnowski, krakowski and siedlecki (more than 79 years). The subregions of the north-eastern Poland: starogardzki, gdański, szczeciński, as well as those located in the Upper Silesia: rybnicki, gliwicki, tyski, katowicki and bytomski are characterized by the shortest life length. Interestingly, the biggest cities are among the subregions with the longest life length of men, but not of women.

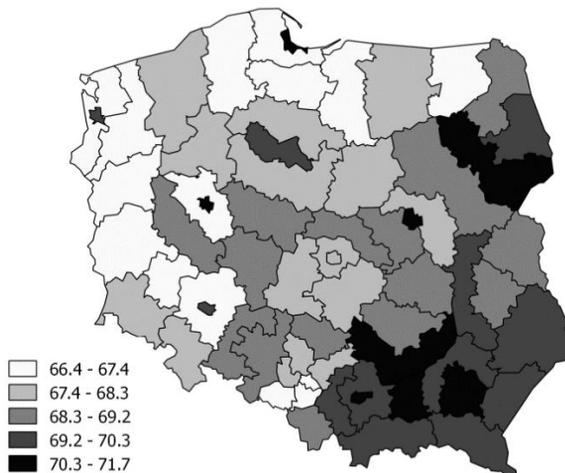


Figure 1. Average life length of men (Y_m)

Source: own calculation.

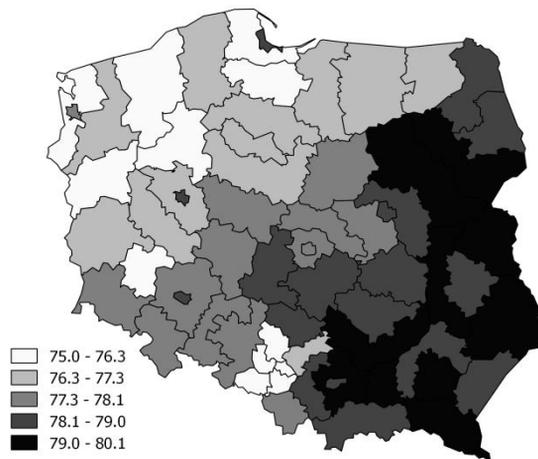


Figure 2. Average life length of women (Y_w)

Source: own calculation.

Next, we present the results of point and interval estimation of QSAR model for the selected quantiles:

$$\begin{aligned}
 Y = & \rho^{(\tau)} \mathbf{WY} + \beta_0^{(\tau)} + \beta_1^{(\tau)} \mathbf{DP} + \beta_2^{(\tau)} \mathbf{E} + \beta_3^{(\tau)} \mathbf{UR} + \beta_4^{(\tau)} \mathbf{GVA} + \\
 & + \beta_5^{(\tau)} \mathbf{RU} + \beta_6^{(\tau)} \mathbf{FR} + \beta_7^{(\tau)} \mathbf{APE} + \beta_8^{(\tau)} \mathbf{IH} + \beta_9^{(\tau)} \mathbf{ND} - \\
 & - \beta_{10}^{(\tau)} \mathbf{CD} + \beta_{11}^{(\tau)} \mathbf{C} + \beta_{12}^{(\tau)} \mathbf{RD} + \boldsymbol{\varepsilon}^{(\tau)}
 \end{aligned} \tag{11}$$

First, we consider the average life length of men. Below, we show only the estimates of $\beta^{(\tau)}$ that are significantly different from 0 for most of the quantiles. The solid lines represent the point estimates and the 90% confidence interval bounds are depicted by the dotted lines. In this case only 5 out of 12 variables significantly affect the life length: density of population (DP), unemployment rate (UR), the rate of urbanization (RU), number of inhabitants per a hospital bed (IH) and number of doctors per capita (ND). Impact of these variables differs with the quantile order, but three of them are statistically significant regardless of the quantile studied: density of population, the rate of urbanization and number of inhabitants per a hospital bed. Additionally, the unemployment rate significantly affects the life length in the subregions where the life length exceeds median. The number of doctors per capita plays a significant role only in the subregions where the life length is high (quantiles of order 0.7 and 0.8).

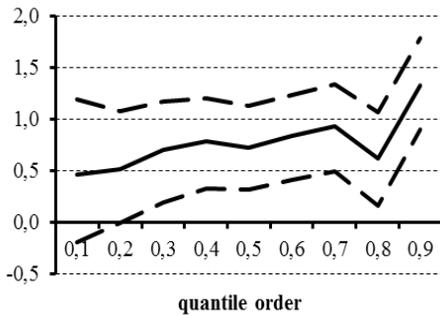


Figure 3. Parameter estimates for population density (DP) in the model (11)

Source: own calculation.

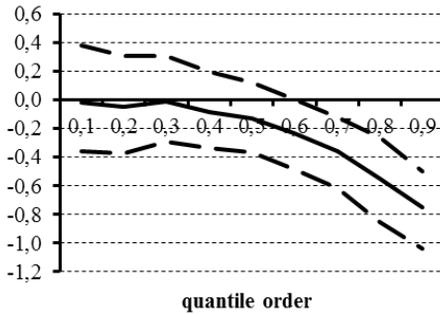


Figure 4. Parameter estimates for the unemployment rate (UR) in the model (11)

Source: own calculation.

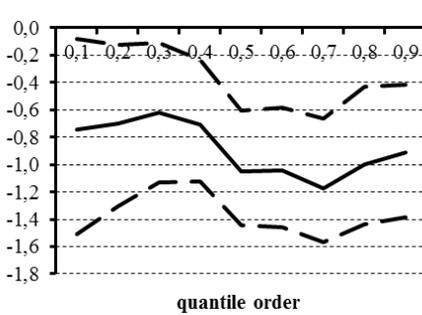


Figure 5. Parameter estimates for the urbanization rate (RU) in the model (11)

Source: own calculation.

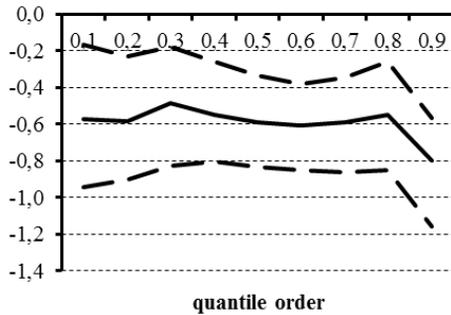


Figure 6. Parameter estimates for number of inhabitants per a hospital bed (IH) in the model (11)

Source: own calculation

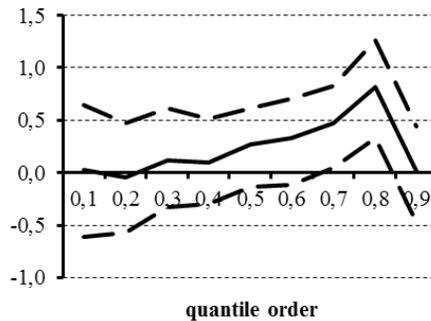


Figure 7. Parameter estimates for number of doctors per capita (ND) in the model (11)

Source: own calculation.

More detailed conclusions can be drawn after the model is respecified by excluding the insignificant variables. Now, the estimated model takes the following form:

$$Y = \rho^{(\tau)}WY + \beta_0^{(\tau)} + \beta_1^{(\tau)}DP + \beta_2^{(\tau)}UR + \beta_3^{(\tau)}PY + \beta_4^{(\tau)}IH + \beta_5^{(\tau)}ND + \epsilon^{(\tau)} \tag{12}$$

Figures 8–12 contain the plots of the new estimates.

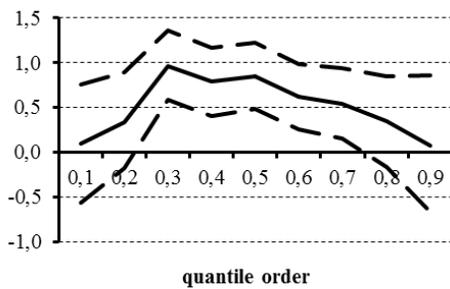


Figure 8. Parameter estimates for population density (DP) in the model (12)
Source: Own calculation.

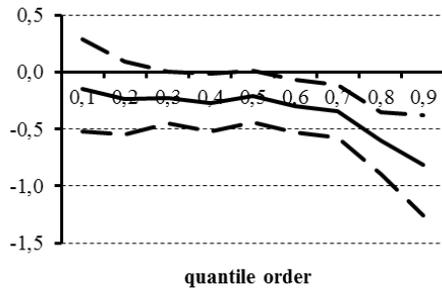


Figure 9. Parameter estimates for the unemployment rate (UR) in the model (12)
Source: own calculation.

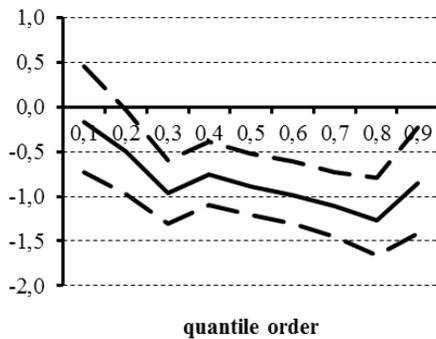


Figure 10. Parameter estimates for the urbanization rate (RU) in the model (12)
Source: own calculation.

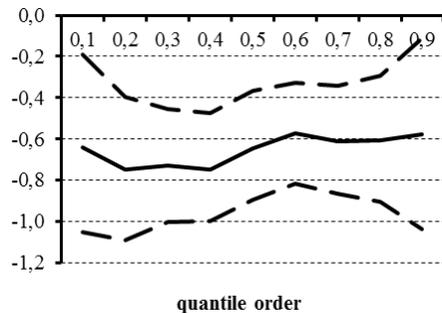


Figure 11. Parameter estimates for number of inhabitants per a hospital bed (IH) in the model (12)
Source: own calculation.

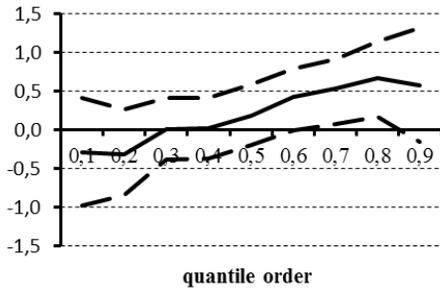


Figure 12. Parameter estimates for number of doctors per capita (ND) in the model (12)
Source: own calculation.

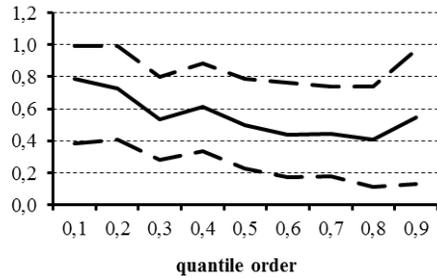


Figure 13. Parameter estimates of the spatial autoregression coefficient in the model (12)
Source: own calculation.

The impact magnitude of the variables on the studied outcome differs across the quantiles. For example for the quantiles $\tau = 0.2, 0.5$ and 0.8 , we have:

$$\hat{Y}_M^{(0.2)} = 0.72^{(0.2)} \hat{WY} + 18.04^{(0.2)} + 0.3^{(0.2)} DP - 0.2^{(0.2)} UR - 0.5^{(0.2)} RU +$$

$$- 0.7^{(0.2)} IH - 0.3^{(0.2)} ND$$

$$\hat{Y}_M^{(0.5)} = 0.5^{(0.5)} \hat{WY} + 34.2^{(0.5)} + 0.85^{(0.5)} DP - 0.21^{(0.5)} UR - 0.9^{(0.5)} RU +$$

$$- 0.64^{(0.5)} IH + 0.18^{(0.5)} ND$$

$$\hat{Y}_M^{(0.8)} = 0.4^{(0.8)} \hat{WY} + 41.19^{(0.8)} + 0.34^{(0.8)} DP - 0.6^{(0.8)} UR - 1.27^{(0.8)} RU +$$

$$- 0.6^{(0.8)} IH + 0.67^{(0.8)} ND$$

From Figure 8, we can conclude that the population density variable is insignificant for the quantiles of order 0.1 and 0.2 as well as 0.8 and 0.9. Therefore, the variable is positively associated with the average life length of men in all the subregions excluding those with the highest and the lowest life length. Similarly, the unemployment rate (Figure 9) and the urbanization rate (Figure 10) has a negative impact on the life length for all regions but those with the lowest life length. The number of inhabitants per a hospital bed (Figure 11) negatively affects the life length in all the subregions. These results show the importance of number, capacity and localization of hospitals in Poland in the public health management. Finally, the number of doctors has a significant, positive impact on the life length only in the subregions where the life length exceeds the median (Figure 12), namely the central and the south-western. Figure 13 presents the estimates of the spatial autoregression coefficient, which are relatively stable for the analysed quantiles.

We conduct a similar analysis for the average life length of women. The statistically significant estimates of the parameters of the model (11) are shown in Figures 14–18. We identify 5 variables that significantly affect our endogenous variable for most of the studied quantiles: density of population (DP), gross value of fixed assets (GVA), the rate of urbanization (RU), the fraction of deaths from cardiovascular diseases (CD) and the fraction of deaths from cancers (C). As for men, the impact of the variables differs across the quantiles, but the conclusions are drawn from the respecified model.

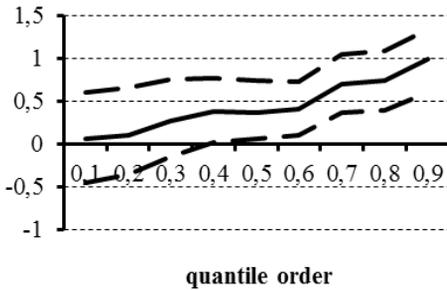


Figure 14. Parameter estimates for population density (DP) in the model (11) for women
Source: own calculation.

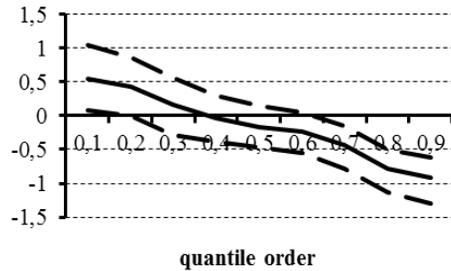


Figure 15. Parameter estimates for gross value of fixed assets (GVA) in the model (11) for women
Source: own calculation.

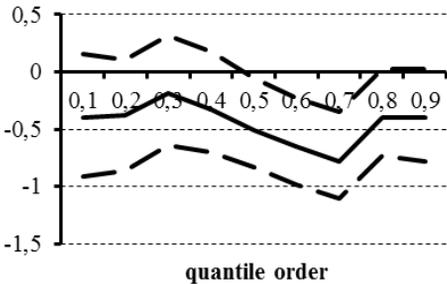


Figure 16. Parameter estimates for the urbanization rate (RU) in the model (11) for women
Source: own calculation.

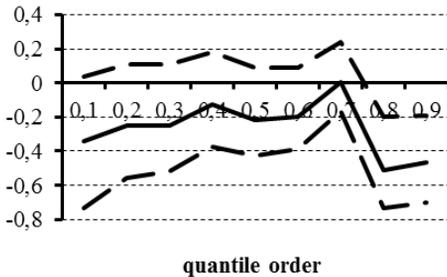


Figure 17. Parameter estimates for the fraction of death from cardiovascular diseases (CD) in the model (11) for women
Source: own calculation.

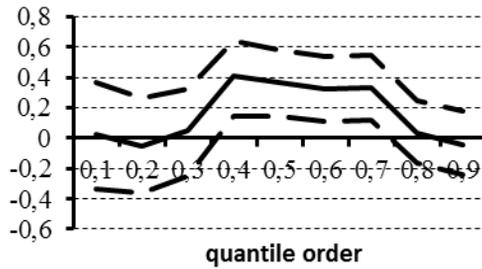


Figure 18. Parameter estimates for the fraction of deaths from cancers (C) in the model (11) for women

Source: own calculation.

The respecified QSAR model for women takes the following form:

$$\begin{aligned}
 \mathbf{Y} = & \rho^{(\tau)}\mathbf{WY} + \beta_0^{(\tau)} + \beta_1^{(\tau)}\mathbf{DP} + \beta_2^{(\tau)}\mathbf{GVA} + \beta_3^{(\tau)}\mathbf{RU} + \\
 & + \beta_4^{(\tau)}\mathbf{CD} + \beta_5^{(\tau)}\mathbf{C} + \boldsymbol{\varepsilon}^{(\tau)}
 \end{aligned}
 \tag{13}$$

The parameter estimates of that model are illustrated in Figures 19–23.

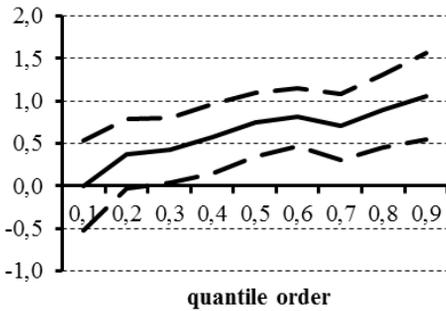


Figure 19. Parameter estimates for population density (DP) in the model (13) for women

Source: own calculation.

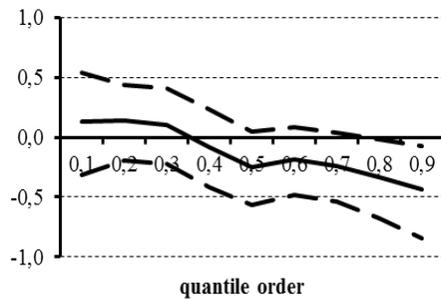


Figure 20. Parameter estimates for gross value of fixed assets (GVA) in the model (13) for women

Source: own calculation.

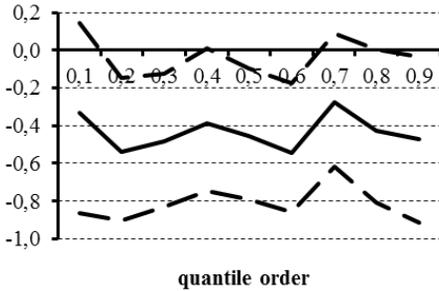


Figure 21. Parameter estimates for the urbanization rate (RU) in the model (13) for women

Source: own calculation.

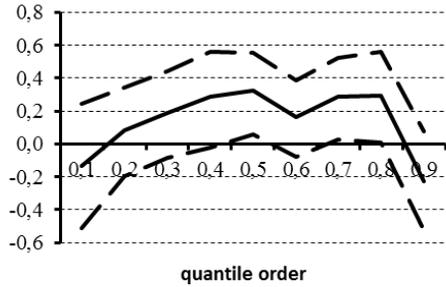


Figure 22. Parameter estimates for the fraction of death from cardiovascular diseases (CD) in the model (13) for women

Source: own calculation.

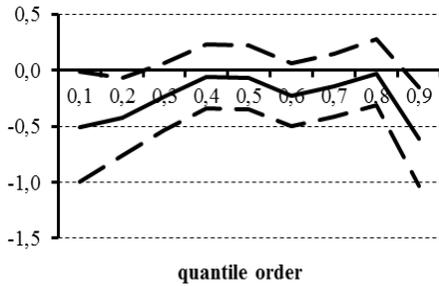


Figure 23. Parameter estimates for the fraction of deaths from cancers (C) in the model (13) for women

Source: own calculation.

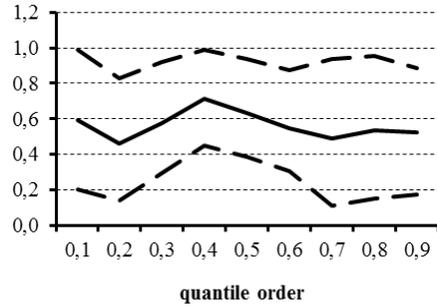


Figure 24. Estimates of the spatial autoregression parameter for the model (13) for women

Source: own calculation.

The impact magnitude of the variables on the studied outcome differs across the quantiles. For example for the quantiles $\tau = 0.2, 0.5$ and 0.8 , we have:

$$\hat{Y}_W^{(0.2)} = 0.46^{(0.2)} \hat{WY} + 41.48^{(0.2)} + 0.37^{(0.2)} DP + 0.14^{(0.2)} GVA - 0.53^{(0.2)} RU + 0.08^{(0.2)} CD - 0.42^{(0.2)} C$$

$$\hat{Y}_W^{(0.5)} = 0.63^{(0.5)} \hat{WY} + 28.75^{(0.5)} + 0.74^{(0.5)} DP - 0.25^{(0.5)} GVA - 0.45^{(0.5)} RU - 0.32^{(0.5)} CD - 0.06^{(0.5)} C$$

$$\hat{Y}_W^{(0.8)} = 0.52^{(0.8)} \hat{WY} + 37.85^{(0.8)} + 1.05^{(0.8)} DP - 0.44^{(0.8)} GVA - 0.47^{(0.8)} RU - 0.22^{(0.8)} CD - 0.61^{(0.8)} C$$

From Figure 19, we conclude that the population density variable is insignificant only for the quantile of order 0.1, which relates to the subregions

with the shortest life length (north-western and Upper Silesia). This is also the case for the urbanization rate, although now the relationship is negative (Figure 21). For the subregions where the average life length exceeds the median, the life length is inversely related to the gross value of fixed assets (Figure 20). Such a relationship is observed for the subregions with the moderate and long life length, namely the central ones and the south-western. The fraction of deaths from cancers (Figure 23) has a negative impact on the average life length only in subregions with the shortest and the longest life length. The opposite is true for the fraction of deaths from cardiovascular diseases (Figure 22), which means that it does not affect the average life length in the tails of its distribution. Figure 24 contains the estimates of the spatial autoregression coefficient that are of similar order as for men.

6. CONCLUSION

We study the health and socio-economic factors affecting the average life length of men and women in Poland. Their impact is generally different across space and sex. For men, the identified factors are the density of population, the unemployment rate, the rate of urbanization, number of inhabitants per a hospital bed and number of doctors per capita. For women, these are: the density of population, gross value of fixed assets, the rate of urbanization, the fraction of deaths from cardiovascular diseases and the fraction of deaths from cancers.

The results also show the advantage of the quantile regression approach which accounts for the regional differentiation of the studied variables. More specifically, quantile regression with the high quantile orders should be used for the mortality analysis in the south-western subregions, while the low quantiles are appropriate for the north-eastern subregions.

Our results can be utilized in the process of fund allocation for the health prevention to reduce the spatial health inequalities as postulated by the National Health Program for the years 2016–2020.

REFERENCES

- Anselin L. (1998), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Netherlands.
- Chernozhukov V., Chansen C. (2006), *Instrumental Quantile Regression Inference for Structural and Treatment Effect Models*, "Journal of Econometrics", vol. 127, no. 2, p. 491–525, <http://dx.doi.org/10.1016/j.jeconom.2005.02.009>.
- Kim T.H., Muller C. (2004), *Two-stage quantile regression when the first stage is based on quantile regression*, "Econometrics Journal", vol. 7, no. 1, p. 218–231, <http://dx.doi.org/10.1111/j.1368-423x.2004.00128.x>.

- Koenker R., Bassett B. (1978), *Regression Quantiles*, "Econometrica", vol. 46, no. 1, p. 33–50, <http://dx.doi.org/10.2307/1913643>.
- Life Expectancy Tables of Poland in 2014*, Central Statistical Office, Warsaw, 2015.
- Lee L.F. (2002), *Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models*, "Econometric Theory", vol. 18, no. 2, p. 252–277, <http://dx.doi.org/10.1017/s0266466602182028>.
- Lee L.F. (2007), *GMM and 2SLS estimation of mixed regressive, spatial autoregressive models*, "Journal of Econometrics", vol. 137, no. 2, p. 489–514, <http://dx.doi.org/10.1016/j.jeconom.2005.10.004>.
- LeSage J.P., Pace R.K. (2009), *Introduction to spatial econometrics*, CRC Book Press, Boca Raton.
- LeSage J.P. (1997), *Bayesian estimation of spatial autoregressive models*, "International Regional Science Review", vol. 20, no. 1–2, p. 113–129, <http://dx.doi.org/10.1177/016001769702000107>
- Lum K., Gelfand A. (2012), *Spatial quantile multiple regression using the asymmetric Laplace process*, "Bayesian Analysis", vol. 7, no. 2, p. 235–258, <http://dx.doi.org/10.1214/12-ba708>
- National Health Program for 2016–2020*, Appendix to the Resolution of the Cabinet on 15.05.2015, Warsaw.
- Orwat-Acedańska A., Trzpiot G. (2011), *The classification of Polish mutual balanced funds based on the management style – quantile regression approach*, "Research Papers of the University of Economics in Wrocław. Econometrics", vol. 31, no. 194, p. 9–23.
- Portnoy S., Koenker R. (1997), *The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error Versus Absolute-Error Estimators*, "Statistical Science", vol. 12, no. 4, p. 279–300, <http://dx.doi.org/10.1214/ss/1030037960>.
- Suhecki 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", vol. 9, p. 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, p. 469–479.
- Trzpiot G. (2009b), *Estimation methods for quantile regression*, "Economics Studies" vol. 53, Karol Adamiecki University of Economics in Katowice, p. 81–90.
- 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, p. 61–76.
- Trzpiot G. (2012), *Spatial Quantile Regression*, "Comparative Economic Research. Central and Eastern Europe", vol. 15, no. 4, p. 265–279.

Grażyna Trzpiot Agnieszka Orwat-Acedańska

PRZESTRZENNA REGRESJA KWANTYLOWA W ANALIZIE UMIERALNOŚCI

Streszczenie. Artykuł podejmuje problem analizy umieralności w Polsce. Celem pracy jest badanie wpływu wybranych czynników na średnią długość życia kobiet i mężczyzn w 66 podregionach Polski. Stosujemy modele kwantylowej autoregresji przestrzennej. W analizie wykorzystujemy metodologię wielorakiej regresji kwantylowej, która 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.

Słowa kluczowe: regresja kwantylowa, analiza umieralności, analiza przestrzenna.

JEL: J11, C14, C30.