RT&A # 02 (29)
(Vol.8) 2013, June

MODEL OF MULTILEVEL STOCHASTIC ANALYSIS OF ROAD SAFETY ON REGIONAL LEVEL

J.Wachnicka

•

Gdansk University of Technology, Poland
e-mail: joanna.wachnicka@pg.gda.pl

L. Smolarek

•

Gdynia Maritime University, Poland
e-mail: leszsmol@am.gdynia.pl

ABSTRACT

In this paper multilevel approach to the issue of road safety level on the road network of European regions, classified as NUTS 2 in statistical databases of the European Union, has been presented. Following the pattern of many publications on road safety it has been assumed that the risk calculated as the number of death casualties in road accidents per 100,000 inhabitants of a given region has Poisson distribution. Therefore, generalized Poisson model has been assumed in the modelling process. Multilevel stochastic analysis was performed for the studied factor. Then a model was created that took into account the impact of different characteristics available on different level of aggregation, which may be helpful in the actions aimed at improvement of road safety in respective regions.

Key words: road safety, factors, modelling, Europe, regions

1 INTRODUCTION

In 1771 the first accident involving motor vehicle, a steam powered one, was reported. Since then several hundred million accidents have occurred, in which over 60 million people have died. Despite the activities being carried out with an aim to improve safety, over 1.2m people die on roads each year, and even up to 50 million are injured [1]. This is then a global issue of an epidemiologic nature. Scientists in Western Europe and United States for quite a long time have been searching for the cause of this situation. The issue is complex enough to be addressed by scientists from different fields: economy, mathematics, transport or medicine. However, so far they have focused on researches covering data for respective countries, without going deeper into differences between respective regions of a given country. Most frequently analysed were the figures of changes in number of casualties over time, by means of time series [2]. In researches aimed at finding factors that could influence fatality on roads national product per capita [3] and transport activity [4] have been indicated. Unfortunately, transport activity is unavailable in regional databases. Therefore, the scientists often point to population density as a good substitute index, which may replace transport activity [5]. Literature studies showed that the researchers focus either on national characteristics or on regional characteristics alone, and do not combine both. In this paper the combination of national and regional characteristics in one model has been presented.

2 MULTILEVEL MODEL – METHODOLOGY OF THE APPROACH

In order to create model that combines national and regional characteristics, data concerning the number of death casualties in a given region have been collected as well as additional characteristics that describe regions and countries. The reason for this approach was the fact that there were characteristics available on national level that could effectively differentiate safety levels in the regions of respective countries, though unfortunately they were unavailable on regional level of aggregation. On the other hand, respective regions of a given country differ among each other in terms of population density or road network concentration and these elements are worth considering in the model. Since significant dispersion of fatality rate has been observed, it was decided to model demographic index of fatalities on roads (FATALR) calculated as the number of killed per 100,000 inhabitants. The assumption was that the model should have the following formula:

$$FATALR = \alpha \cdot MODEL_{NATIONAL}^{\beta_1} \cdot MODEL_{REGION}^{\beta_2} \cdot NPPC_{NATIONAL}^{\beta_3} \qquad (1)$$

where:
FATALR – demographic rate of fatalities in road accidents in a given region [fat./100 thou. inhab.]
$MODEL_{NATIONAL}$ *- model for national data*
$MODEL_{REGION}$ *- model for regional data*
$NPPC_{NATIONAL}$ *- model describing changes of average national product per capita*
α, β_1,β_2,β_3 *- estimated parameters*

FATALR estimation is based on the assumption that this parameter has Poisson distribution. National models were created on the basis of data from 11 European countries, whereas in the case of regional models in this paper the focus has been on two European countries that substantially differ in terms of road safety level: Great Britain, where the actions for improvement of road safety had a long tradition, and Poland, where the average fatality rate is more than double British figure, likely attributed to cultural, political and economic differences. Histograms of FATALR value in regions of comparable countries in the analysed period of 1999-2008 presented on Fig.1 show that there are no grounds for rejecting the hypothesis of Poisson distribution, frequently assumed in the analyses of safety level [6]. In further analyses, according to this assumption, FATALR index will be alternatively referred to as λ.

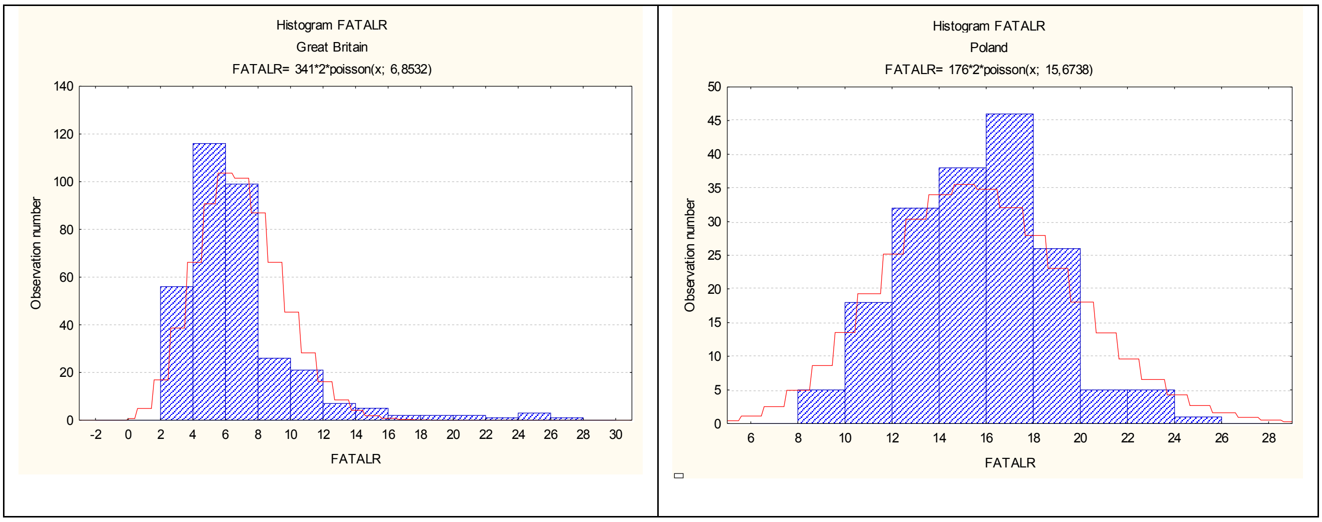

Figure 1. Histograms of analysed FATALR indices in the regions of Great Britain and Poland.

3 NATIONAL MODEL

Taking into account the assumptions, λ parameter has been evaluated with 95% probability for the set of data from comparable countries. Then the set of independent variables, available only on national level of aggregation, which characterize the country, used for development of descriptive models for evaluated λ parameters ($MODEL_{NATIONAL}$), has been created. In this paper the impact of such factors were analysed as: corruption index - COR (the higher the value, the better a country is perceived, i.e. as less corrupted), percentage of passenger cars older than 10 years – OLD in the total fleet of the country, calculated as average from 10 years.

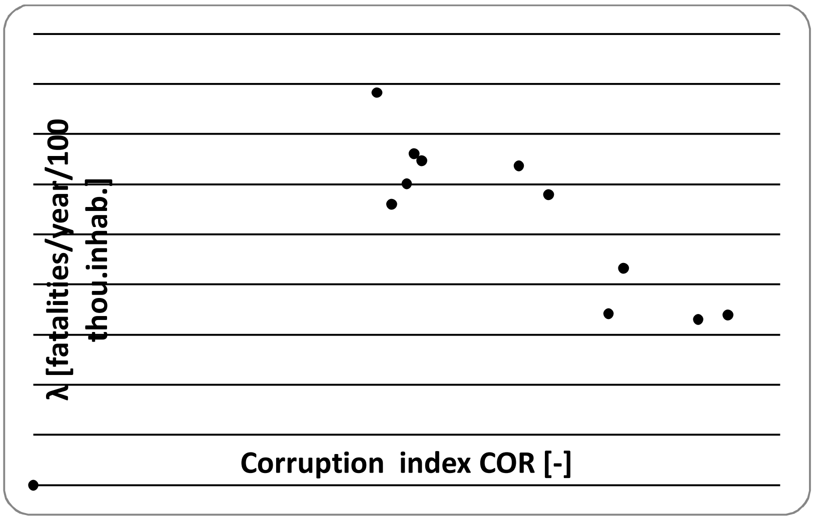

Figure 2. Graph of dependence of λ on the value of corruption index in a given country

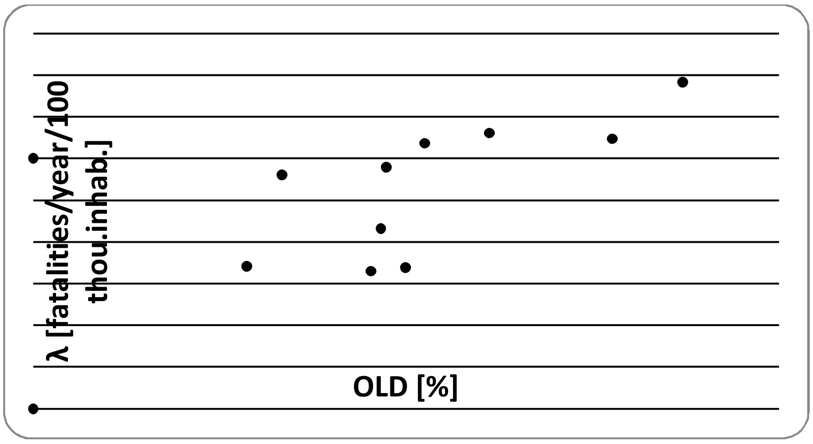

Figure 3. Graph of dependence of λ on the percentage of passenger cars older than 10-years old

As Figures 2 and 3 show these indices may impact the studied parameter. The tendency of falling λ parameter with the increase of corruption index has been observed, whereas the reverse correlation has been seen in the case of the percentage of old passenger cars. The developed model has a shape of linear model:

$$\lambda = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 \quad (2)$$

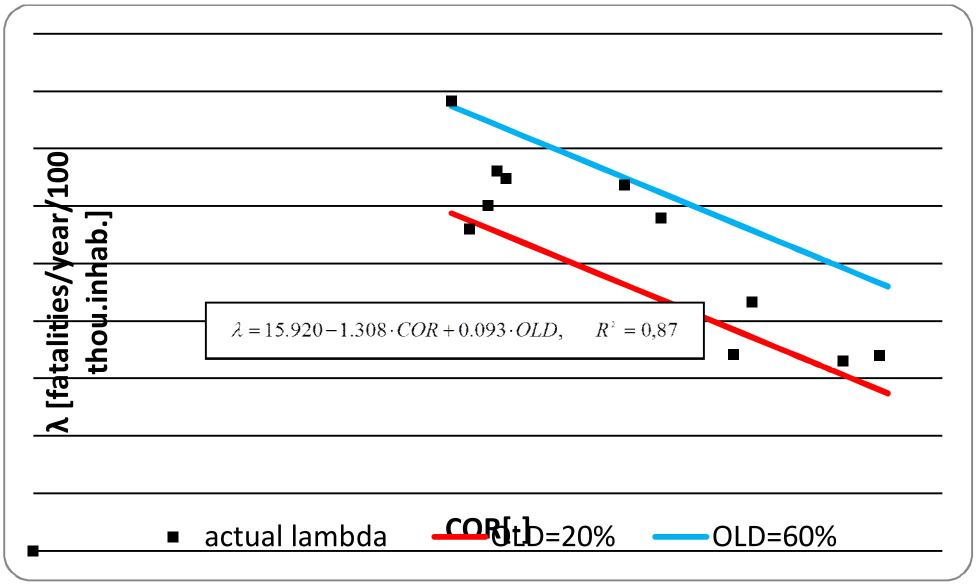

Figure 4. Illustration of linear model that takes into account corruption index and the percentage of cars older than 10 years old in the total fleet

Q factor for the model equals 0.87. Together with the increase of corruption index – COR, λ parameter is falling, however if there is a large percentage of old cars in total fleet of passenger cars, then λ parameter will be higher.

4 REGIONAL MODEL

The next step was development of models describing the impact of regional characteristics in respective countries on FATALR values in these regions. For this purpose, separated base of regional data was created for each country and the attempts were made to develop a model of impact of respective variables on modelled dependent variable. In the case of all countries one type of the model was checked, which was initially prepared based on joined database from all European regions. General shape of this model has been presented below (3), whereas in table 1, calculated indices in the model in analysed countries have been listed. Cluster analysis allowed specification of classes of correlated variables. In individual models the impact of respective classes have been taken into account through selection of their representatives.

$$MODEL_{REGION} = \alpha \cdot (lnNPPC)^{\beta} \cdot NPPC^{\gamma_6} \cdot DPOP^{\gamma_1} \cdot VEHD^{\gamma_2} \cdot e^{(\gamma_3 \cdot CARP + \gamma_4 \cdot ROADC + \gamma_5 \cdot UNEMP)} \quad (3)$$

where:

$MODEL_{REGION}$ - model for regional data
NPPC - average national product per capita in a given region, in a given year [thou. EUR]
DPOP - population density [no. of people/km^2]
VEHD – total vehicle density [veh./km^2]
CARP – percentage of passenger cars in total fleet of cars [%]
ROADC – intensity of the total number of roads per one inhabitant [km/person]
UNEMP – unemployment index
α, β, γ_1, γ_2, γ_3, γ_4, γ_5, γ_6 - estimated parameters

Table 1. List of parameters of regional models for selected countries

Model	R^2	α	β	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
Great Britain	0,7 1	0,012	32,3 18	- 0,41 3	0,26 6		0,36 4	- 0,03 9	- 9,50 1
Poland	0,5 4	154,3 61	- 0,17 9	0,12 3	- 0,14 9	- 0,03 1	0,35 8		

On Figures 5 and 6 the regional model has been shown

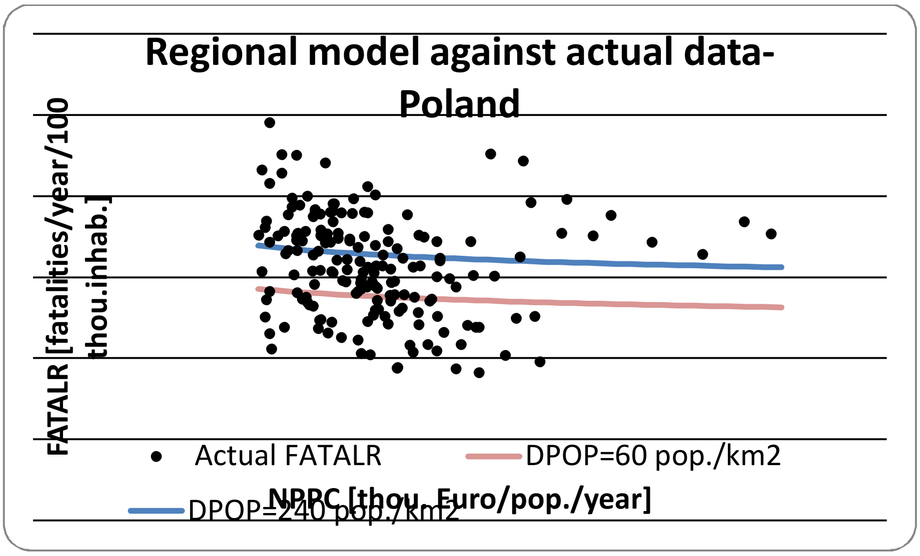

Figure 5. Graph of prepared regional FATALR model in relation to population density DPOP against the actual data for Poland - remaining variables from the model assumed as average.

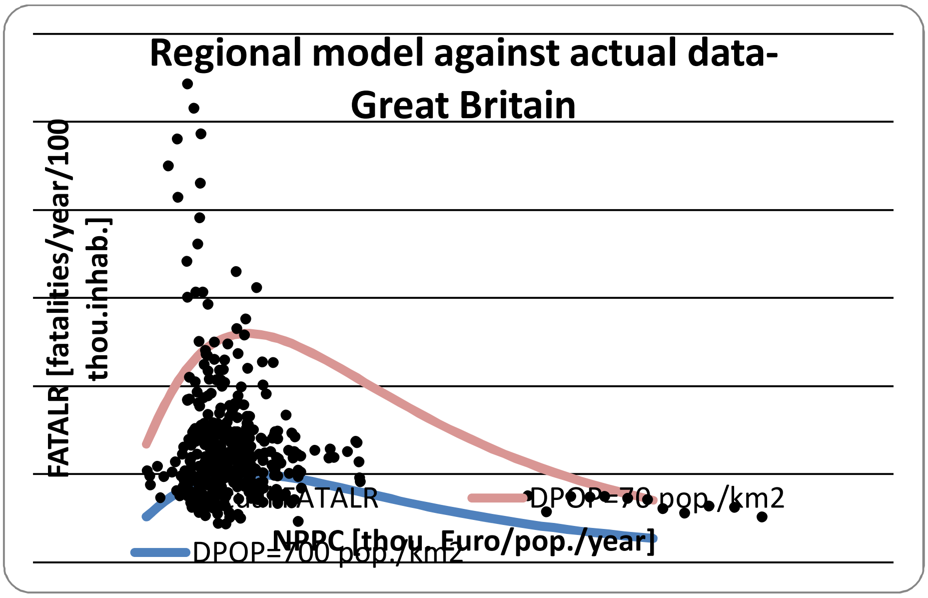

Figure 6. Graph of prepared regional FATALR model in relation to population density DPOP against the actual data for Great Britain - remaining variables from the model assumed as average.

5 MODEL FOR TREND OF NATIONAL PRODUCT PER CAPITA VARIATION IN TIME

Global national and local regional models are the ones, in which the majority of independent variables are characterized by slight variability in time. The only variable dynamically changing in time, and at the same time occurring in almost all models, is national product per capita. Analysis of variations of average national product per capita NPPC over time proved that analysed countries are characterized by two types of NPPC change trends in time.

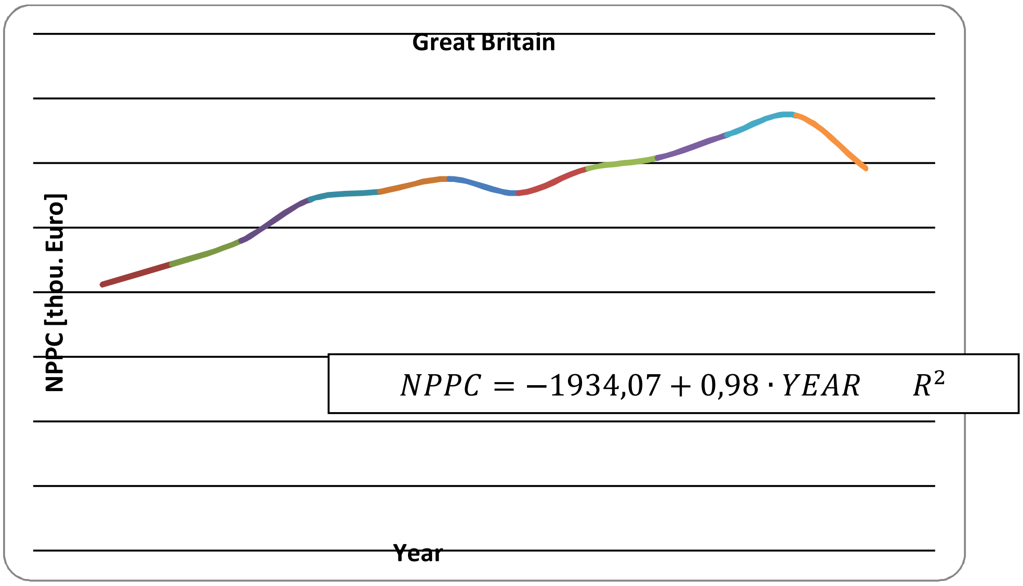

Figure 7. Graph of dependence of NPPC changes in Great Britain in the analysed years

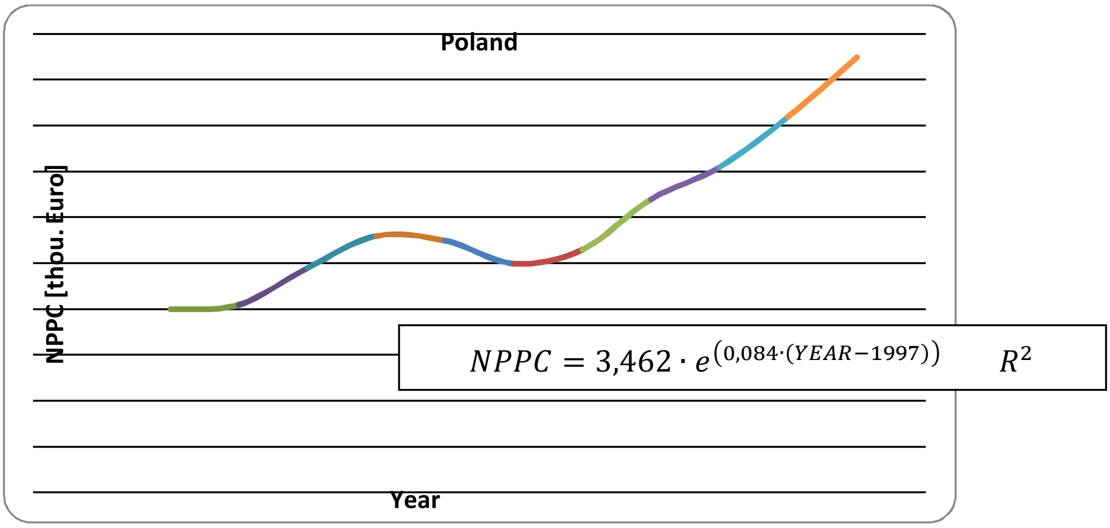

Figure 8. Graph of dependence of NPPC changes in Poland in the analysed years

First one is a linear, which characterizes mainly the countries of the “Old Europe”, where the economic situation is stabilized and standard of living is improving. The second one, however, is nonlinear trend, probably resulting from dynamic changes occurring in these countries after their access to the European Union, Fig. 7 and Fig. 8.

6 Aggregated model

After preparation of partial models, the model comprising all partial models has been created and it has the following formula:

$$ATALR = 5{,}353 \cdot MODEL_{NATIONAL}^{-0{,}285} \cdot MODEL_{REGION}^{0{,}818} \cdot JPKB_{NATIONAL}^{-0{,}212} \quad (4)$$

It is a multiplicative model, elements of which have Q factors ranging between 0.89 and 0.54, while resultant model has Q factor equal to 0.76. This is a result of using average annual data as input data, in order to eliminate momentary fluctuations that could obscure the character of effects of respective influences. Received results may be considered satisfactory and visualisation of model adjustment to the actual data has been presented in Fig. 9 and 10.

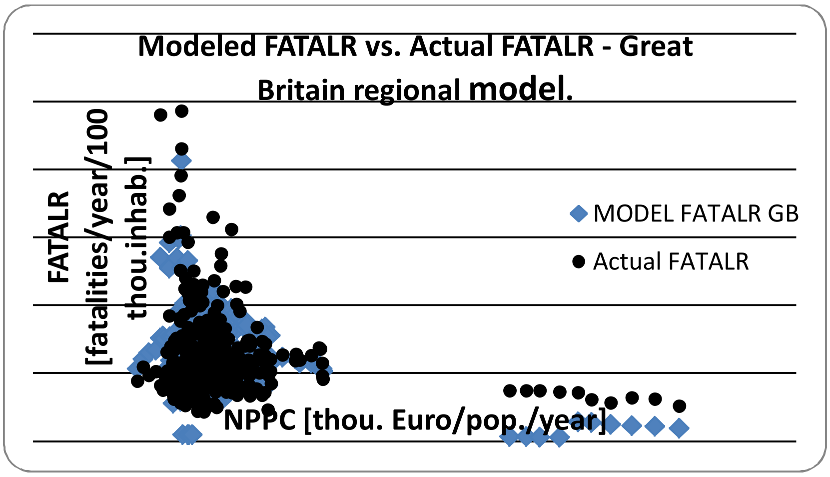

Figure 9. Graph of the actual and modelled data in relation to NPPC in a given region for regions in Great Britain.

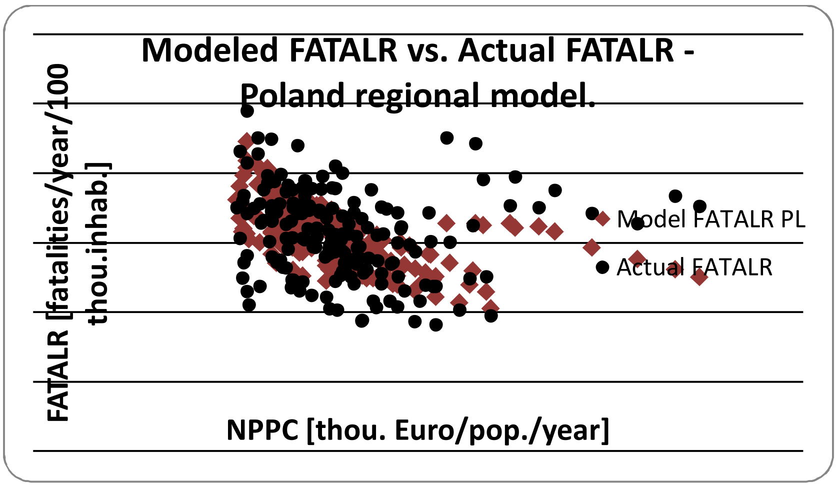

Figure 10. Graph of the actual and modelled data in relation to NPPC in a given region for regions of Poland.

7 ANALYSIS OF THE IMPACT OF RESPECTIVE VARIABLES ON FATALR MODEL

Analysis of the model's sensitivity is based on analysing changes in FATALR occurring as a result of changes in the values of variables being the arguments of model function, based on the observation of this function's derivatives against respective variables.

Furthermore, significant information on the way of potential impact on FATALR are obtained from the information on the values of sensitivity coefficients and the direction as well as size of variations of these coefficients caused by the changes of function's arguments. Sensitivity analysis is one of the basic instruments of risk assessment in decision-making process. Sensitivity analysis allows defining accuracy level, which is necessary at estimation of respective parameters for the model to be precise enough.

Since hierarchic model of FATALR is a multiplicative model, in which respective factors are raised to different powers, then the impact of each factor, estimated by elasticity function, is constant and equal to exponent of the power [7]. Hence performing sensitivity analysis for $\text{Model}_{\text{REGION}}$ is particularly important, especially that it takes into account local specificity as concerns variables existing in the model. Depending on the number of changed factors, the impact of which is taken into consideration, univariate and multivariate analysis is distinguished. In the case of univariate sensitivity analysis, the reaction of the model to a change of one of the factors is studied, while assuming constant level of the other ones.

Most frequently used sensitivity measures of gain include measures based on the concept of point elasticity of function towards respective variables. This approach yields correct results, when we consider slight change of a given factor.

Similarly to referenced article [7], by calculation of elasticity of regional model for respective variables we receive:

$$E_M(CARP) = \gamma_3 \cdot CARP, E_M(ROADC) = \gamma_4 \cdot ROADC, E_M(UNEMP) = \gamma_5 \cdot UNEMP$$
$$E_M(DPOP) = \gamma_1, E_M(VEHD) = \gamma_2$$
$$E_M(NPPC) = \frac{\beta + lnNPPC \cdot \gamma_6}{lnNPPC} = \frac{\beta}{lnNPPC} + \gamma_6.$$

Set of independent variables may then be divided into three groups of variables:

- fixed level of elasticity, DPOP and VEHD variables;
- linearly dependent elasticity level, CARP, ROADC, UNEMP variables;
- inversely proportional elasticity level, NPPC variable.

Variables of the first two groups are characterized by relatively small dynamics of change, as in a period of about 10 years it is difficult to observe significant changes in demography of a given region, and similarly in the total vehicle fleet or total length of roads. It is different in case of unemployment index, although this variable is present only in one of the models prepared and turned out to be useful in the case of British regions. The most interesting proved the impact of national product per capita in respective regions. Its changes turned out to have the biggest impact on the modelled FATALR values in the regions.

When building the model special attention must be paid to separable homogenous groups of elements, especially in respect of their variability.

The purpose of sensitivity analysis of FATALR models is checking how the value of the model changes as a result of combined changes of factors that impact the model. In multivariate sensitivity analysis the impact of combined changes of several factors is taken into account. Therefore, it is necessary to apply differential for the assessment of the total change of the model's value. By calculating partial derivatives for respective variables, we receive

$$\frac{\partial Model_{Region}}{\partial CARP} = \gamma_3 \cdot \alpha \cdot (lnNPPC)^{\beta} \cdot NPPC^{\gamma_6} \cdot DPOP^{\gamma_1} \cdot VEHD^{\gamma_2} e^{(\gamma_3 \cdot CARP + \gamma_4 \cdot ROADC + \gamma_5 \cdot UNEMP)}$$

$$\frac{\partial Model_{Region}}{\partial ROADC} = \gamma_4 \cdot \alpha \cdot (lnNPPC)^{\beta} \cdot NPPC^{\gamma_6} \cdot DPOP^{\gamma_1} \cdot VEHD^{\gamma_2} e^{(\gamma_3 \cdot CARP + \gamma_4 \cdot ROADC + \gamma_5 \cdot UNEMP)}$$

$$\frac{\partial Model_{Region}}{\partial UNEMP} = \gamma_5 \cdot \alpha \cdot (lnNPPC)^{\beta} \cdot NPPC^{\gamma_6} \cdot DPOP^{\gamma_1} \cdot VEHD^{\gamma_2} e^{(\gamma_3 \cdot CARP + \gamma_4 \cdot ROADC + \gamma_5 \cdot UNEMP)}$$

$$\frac{\partial Model_{Region}}{\partial DPOP} = \gamma_1 \cdot \alpha \cdot (lnNPPC)^{\beta} \cdot NPPC^{\gamma_6} \cdot DPOP^{\gamma_1 - 1} \cdot VEHD^{\gamma_2} e^{(\gamma_3 \cdot CARP + \gamma_4 \cdot ROADC + \gamma_5 \cdot UNEMP)}$$

$$\frac{\partial Model_{Region}}{\partial VEHD} = \gamma_2 \cdot \alpha \cdot (lnNPPC)^{\beta} \cdot NPPC^{\gamma_6} \cdot DPOP^{\gamma_1} \cdot VEHD^{\gamma_2 - 1} e^{(\gamma_3 \cdot CARP + \gamma_4 \cdot ROADC + \gamma_5 \cdot UNEMP)}$$

$$\frac{\partial Model_{Region}}{\partial NPPC} = [\beta + lnNPPC \cdot \gamma_6](lnNPPC)^{\beta - 1} \cdot NPPC^{\gamma_6 - 1} \cdot \alpha \cdot DPOP^{\gamma_1} \cdot VEHD^{\gamma_2} e^{(\gamma_3 \cdot CARP + \gamma_4 \cdot ROADC + \gamma_5 \cdot UNEMP)}$$

By calculating differential we receive a formula, which describes joined impact of independent variables on the value of $Modelu_{Region}$:

$$dModel_{Region} = \alpha\{[\beta + lnNPPC \cdot \gamma_6] DPOP \cdot VEHD dNPPC + lnNPPC \cdot NPPC \cdot [DPOPVEHD(\gamma_3 dCARP + \gamma_4 dROADC + \gamma_5 dUNEMP) + VEHD\gamma_1 dDPOP + DPOP\gamma_2 dVEHD]\} \cdot \cdot (lnNPPC)^{\beta - 1} \cdot NPPC^{\gamma_6 - 1} \cdot DPOP^{\gamma_1 - 1} \cdot VEHD^{\gamma_2 - 1} e^{(\gamma_3 \cdot CARP + \gamma_4 \cdot ROADC + \gamma_5 \cdot UNEMP)}$$

8 SUMMARY

Presented heuristic and statistical analysis points to advisability of using local characteristics for the assessment of safety in road transport. Due to differences in availability of registered data, classes of “close”, in terms of cluster analysis, variables and their representatives for specified groups of effects (impacts) have been distinguished.

λ parameter in national model is inversely proportional to corruption index and directly proportional to percentage of old passenger cars. These mutually competing effects (impacts) may be used to specify a strategy of actions aimed at reducing road accident fatalities index.

Models describing the impact of regional characteristics in respective countries on the FATALR values in such regions, and analysis of sensitivity, point to NPPC variable as dynamic and controlled element that drives FATALR changes.

In multivariate analysis of sensitivity the impact of concurrent change of several independent factors is analysed. That is why it is necessary to check correlation of variables used in aggregated model as well as in partial models. It will allow development of more effective mechanisms of influencing FATALR index.

Acknowledgement

The paper has been funded by grant-in-aid for carrying out scientific research or development work and related tasks, aimed at the progress of young scientists and doctoral students.

9 REFERENCES

[1] "Global status report on road safety:time for action," Geneva, 2009.

[2] J. J. F. Commandeur, F. D. Bijleveld, R. Bergel-Hayat, C. Antoniou, G. Yannis, and E. Papadimitriou, "On statistical inference in time series analysis of the evolution of road safety.," *Accident; analysis and prevention*, pp. 1–11, Dec. 2012.

[3] A. E. Van Beeck, E., Mackenbach, J. P., Looman, C. W., & Kunst, "Determinants of Traffic Accident Mortality in the Netherlands: A Geographical Analysis.," *International Journal of Epidemiology*, vol. 20, no. 3, pp. 698–706, 1991.

[4] K. Jamroz, "The impact of road network structure and mobility on the national traffic fatality rate," no. Jamroz 2011, pp. 1–7.

[5] L. Fridstrøm, J. Ifver, S. Ingebrigtsen, R. Kulmala, and L. K. Thomsen, "Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts.," *Accident; analysis and prevention*, vol. 27, no. 1, pp. 1–20, Feb. 1995.

[6] Firth, D. (1991) Generalized linear models. In: Hinkley V, Reid N, Snell EJ, editors. *Statistical theory and modelling: in honour of Sir David Cox, FRS*. London: Chapman and Hall; p. 66-93.

[7] Wachnicka J.,Smolarek L, " The Multivariate Multilevel Analysis Of Different Regional Factors Impact on Road Safety in European Country Regions" Journal of KONBIN, No 4 (24) 2012, pp. 141-148, ISSN 1895-8281