Warsaw School of Economics

Collegium of Economic Analysis

mgr Kamil Gala

Abstract of PhD thesis titled

Spatial aspects of insurance risk in automobile insurance

PhD thesis
prepared under the guidance of
dr hab. Wojciech Bijak, prof. SGH

Warsaw 2020

1. Subject, purpose and scope of work

The subject of this dissertation is the spatial distribution of insurance risk on the Polish motor insurance market, including motor third-party liability insurance for motor vehicle owners (MTPL) and motor own damage insurance (MOD or *autocasco*). In view of the ambiguity and complexity of the concept of insurance risk (Ronka-Chmielowiec, 2013) in this work it is defined as the risk associated with concluded insurance contracts, which takes the form of insurance accidents and related claims and benefits. This definition emphasizes the likelihood of an insurance accident (one or more) and the amount of loss resulting from it.

The literature on the subject (e.g. Taylor, 2001; Brouhns et al., 2002; Gala, 2017) indicates that insurance risk (measured e.g. by claim frequency) shows noticeable spatial heterogeneity. The main goal of the dissertation is to assess the significance of the geographical dimension as a factor enabling better assessment of insurance risk in motor insurance. To this end, tools combining the experience of spatial econometrics and actuarial ratemaking have been developed and used.

A priori ratemaking is part of the insurance risk assessment process in which the premium is determined at the time the contract is concluded on the basis of the observable characteristics of the insured and the subject of insurance (e.g. vehicle or real estate). MTPL and MOD insurance constitute a significant part of the Polish property insurance market, which is due to the large number of motor vehicles in Poland and the obligation to have MTPL coverage of each vehicle. Therefore, statistical methods are the fundamental tool for determining insurance premiums in motor insurance.

Over the years, both specific actuarial methods (taking into account, for example, asymmetrical distributions of the dependent variables) and a set of factors considered to be good predictors of insurance losses (such as the age of the insured, the engine capacity of the vehicle or the region of its use) have been established. At the same time, spatial econometrics is developing largely without connection to actuarial statistics. These methods are used to study phenomena located in geographical space, most often characterized by spatial diversity and the relationship between processes occurring in neighboring areas (so-called spatial autocorrelation). One can indicate (see e.g. Lee and Pace, 2009) many real-world situations that lead to the observation of spatial autocorrelation - both as a result of real interactions between regions (so-called *contagion effect*), and as a result of measurement errors or omission of relevant variables of spatial nature. Especially the second effect may be significant in motor insurance, taking into account that the boundaries of administrative division units do not reflect

the specificity of road traffic and weather phenomena, which factors affect the risk of claim occurence. Regardless of the actual source of autocorrelation, omitting it during the construction of the model may cause that the obtained parameter estimators will not have the desired properties, such as consistency, unbiasedness or efficiency.

The dissertation covers theoretical and empirical analysis of statistical models used for *a priori* ratemaking taking into account the spatial nature of insurance data. The first chapter is devoted to *a priori* ratemaking. The second chapter presents issues related to spatial data and describes the basic models used in spatial econometry. The third chapter formulates a generalized spatial linear model, then presents models with a multi-level factor as its specifc case and discusses methods of spatial segmentation. In the last chapter, theoretical considerations were illustrated by empirical analysis of data on the Polish motor insurance market, which included a description of the spatial distribution of insurance risk, the estimation of parameters of predictive models and the assessment of the effectiveness of the methods and models under consideration.

2. Reasons for undertaking the research topic - its theoretical and practical significance

In accordance with art. 33 of the Act of 11 September 2015 on insurance and reinsurance activity (Dz.U. 2020 poz. 895 as amended), the insurance undertaking is required to determine the the insurance premium after assessing the insurance risk (section 1). This premium should be set in an amount ensuring at least the fulfillment of all obligations resulting from insurance contracts and coverage of the costs of carrying out the activity of the insurance undertaking (section 2), and the undertaking collects relevant statistical data to determine the premium based on them (section 3).

Competition on the Polish motor insurance market, increasing customer awareness and legal regulations force insurance companies to improve their risk assessment methods. Consequently, there is a growing interest in both new data about the vehicle owner, such as his credit standing (Gala and Kolak, 2015), penalty points and traffic offenses (Pinquet et al., 2011), as well as data on the number of kilometers driven per year and the style of driving (Lemaire et al., 2016). It should be noted, however, that the use of these possibilities may be limited due to high costs, legal regulations or practical considerations (e.g. the need to obtain client's consent to access data from an external source). In this situation, the alternative may be to use more efficiently the data that is already in the possession of the insurance company. Identyfying of

appropriate tariff variables and determining their impact allows not only to better match the premium to the risk associated with the insurance contract, but also to actively target the insurance undertaking's sales activities towards groups of clients with low claim frequency. It is also worth noting that, in a broader sense, the pricing policy of insurance companies may provide a financial incentive for specific behaviors that affect road safety (e.g. encourage drivers to choose cars with a lower engine power).

The importance of spatial aspects in the study of economic phenomena was emphasized in the work of Anselin (1988), who as the main issues indicates spatial heterogeneity and spatial autocorrelation. At the same time, he draws attention to the need to take into account spatial effects for the purpose of proper specification of the model, estimation of its parameters and verification by means of statistical tests. The author lists two main sources of spatial autocorrelation - errors related to the phenomenon measurement and data aggregation (territorial observation units do not reflect the spatial distribution of phenomena), as well as real processes related to the spread of phenomena and interactions between various areas in space (e.g. migration of people or transport of goods).

Both of these sources can be found in motor insurance. First of all, traditional tariffs at the level of a geographical unit (postal code, municipality, district) may be affected by an error resulting from the fact that the boundaries of these units often do not coincide with the occurrence of large urban agglomerations or communication routes (e.g. highways). Secondly, there may be interactions between different areas, e.g. in the form of heavy traffic between areas (e.g. commuting). In both cases, spatial autocorrelation will be observed at the level of cross-sectional data, which should be included in the specification of the tariff model. It is also worth paying attention to the fact that while more detailed spatial division could allow to reflect spatial nature of the studied phenomenon closely, the consequence of this approach is the small number of observations within territorial units or the lack of observations for some of them. In this situation, obtaining reliable results is difficult or even impossible.

There is extensive scientific literature both in the area of actuarial risk assessment in motor insurance and in the area of spatial data analysis. Issues related to actuarial risk assessment in motor insurance, both *a priori* and *a posteriori*, have been extensively described, including in the works of Denuit et al. (2007) and Lemaire (1995). In turn, issues related to spatial econometrics were developed, among others in the works of Anselin (1988) and LeSage and Pace (2001), while in Polish literature one can indicate a work edited by Suchecki (2010). It should be noted, however, that in actuarial literature there are relatively few papers in which the spatial aspects of insurance risk in motor insurance are analyzed in detail, and none of these

works comprehensively covers the issue. In the paper by Boskov and Verrall (1994), the authors make assumptions about the spatial structure of random effects, but do not provide a method for their extraction from the systematic component of the model. In the work of Brouhns et al. (2002) the use of the Boskov and Verrall model for the analysis of data from the Belgian market is broadly described and the procedure of estimating its parameters is fully described, while the question of assessing the predictive abilities of the model is omitted. In turn, the works of Taylor (1989; 2001) use the method of spatial interpolation of geographical effects determined as residues from a generalized linear model, while no assumptions regarding spatial autocorrelation or spatial heterogeneity of the studied phenomenon are given. The importance of spatial differentiation of insurance risk was also indicated in the studies of Raichle (1997), Brubaker and Bylls (1997) and Lee and Lee (2016), although more attention was focused on the risks associated with property insurance (including fire, flood and natural disasters), and the issue of motor insurance has not been broadly addressed. However, this work indicates the practical importance of geographical analysis in the field of insurance.

Therefore, this dissertation aims to fill the gap existing between actuarial statistics and spatial econometrics, which have so far developed largely independently of each other. The main focus here is to increase the effectiveness of the insurance risk assessment process in motor insurance (understood as the best possible match between the premium and the risk associated with the insurance contract) by taking into account the spatial aspects of this phenomenon. This involves formulation and empirical verification of predictive models that, in addition to standard tariff variables (e.g. vehicle or vehicle owner characteristics) take into account the spatial structure of data and potential interactions between different locations. It is worth noting that the approach adopted gives the opportunity to optimize the risk assessment process by making better use of data held by the insurance undertaking and regional data that are publicly available.

An important element of the work is the empirical analysis of individual, detailed data on the entire motor insurance market in Poland, so far not present in Polish actuarial literature. The use of market-wide data makes it possible to adopt a level of detail that cannot be achieved at the level of an individual insurance undertaking, which for some regions may not have sufficient (if any) number of observations.

To sum up, from the theoretical point of view, the work can contribute to the systematization and development of *a priori* ratemaking methods that combine the tools of classical actuarial statistics and spatial econometrics. The practical benefit will be a better

understanding of the processes taking place on the Polish motor insurance market and identification of the pricing approach appropriate to the specifics of this market.

3. Identifying basic research questions, research goal, main thesis and detailed hypotheses

The main purpose of this dissertation is to assess the significance of spatial aspects of insurance risk in motor insurance in the process of determining the insurance premium and to indicate the best method of using the spatial nature of insurance data. It should be emphasized that one of the criteria for assessing *a priori* pricing methods is also their complexity and applicability in practice.

In addition to the main objective, specific objectives have also been formulated, which can be divided into three categories:

1. Cognitive goals:
    - C1. Presentation of the importance of spatial data in analyses related to insurance risk in motor insurance and indication of directions for further development of research in this area.
    - C2. Identification and description of spatial distribution and spatial autocorrelation of the phenomenon under study on the basis of real data.
2. Methodological goals:
    - C3. Creating a common conceptual framework allowing the analysis of different variants of spatial models.
    - C4. Specification of different models allowing for the analysis of spatial effects.
3. Practical purpose:
    - C5. Estimation of parameters of the spatial models under consideration, assessment of their tariff efficiency and identification of the best approach from a practical point of view.

The following research hypotheses were formulated in the thesis:

H1. Insurance risk in motor insurance is spatially heterogenous and there is a correlation between the level of risk in different areas. Areas which are close to each other are more similar in terms of insurance risk than areas far away from each other.

H2. The precision of statistical inference about insurance risk in the studied area can be increased by using data for other areas.

H3. Including the spatial dimension in the *a priori* pricing model allows to increase its predictive capabilities and efficiency of the pricing process.

H4. The spatial aspects of insurance risk are important from the point of view of its proper assessment and determining the premium appropriate to the risk incurred by the insurance company.

4. Research methodology

The research objectives and hypotheses formulated in the thesis justify the classification of research methods into two groups - theoretical and empirical.

Theoretical methods include the development of tariff models which take into account the spatial dimension of the data and analysis of properties of these models. In this part of the dissertation, a generalized spatial linear model has been formulated, while its specific case – a multi-level factor model (MLF - see Ohlsson and Johansson, 2010) has been subjected to in-depth analysis. In addition to the standard MLF model, its hierarchical version (HMLF) is also described, as well as the spatial MLF model (SMLF). The SMLF model is based on the author's generalization of the classic Bühlmann-Straub model consisting in removing the assumption about the independence of latent risk parameters. Moreover, presentation includes methods of spatial segmentation that can be used to construct tariff variables of a geographical nature, thus constituting an alternative to more complex spatial models, including the author's own method based on the algorithm of detecting communities in non-directed graphs.

Empirical methods include statistical analysis of data from the market-wide database of the Insurance Guarantee Fund Information Center (OI UFG). This database contains information about concluded MTPL and MOD contracts, accidents and claims or benefits paid or refused. The data is mandatorily sent to OI UFG by insurance companies, while article 102 of the Act of May 22, 2003 on compulsory insurance, UFG and PBUK (Journal of Laws 2019 item 2214) defines the scope of data collected in the database. In turn, article 102 section 7 indicates that UFG may process data for purposes other than those enumerated in the Act, after modification that will not allow establishing the identity of the data subject. This allows the Insurance Guarantee Fund to act as a center for advanced analytics for the Polish motor insurance market. As a results, the goal of this dissertation is not only to accomplish certain

research objectives, but also to assess the analytical potential of OI UFG and to support (theoretically and empirically) the development of analytical services provided by UFG to the insurance market.

The research covers MTPL and MOD contracts concluded in 2017. Dataset on MTPL insurance contained about 16.5 million records, while in the case of MOD it was 2.8 million records. Both sets were randomly divided into a training set (70% of observations) and a validation set (30% of observations). The scope of the data used in the study included basic data on the vehicle (type, brand) and data on vehicle owners (age, gender, place of residence), as well as data on the insurance history of the vehicle and its owners.

The empirical analysis was divided into three part, which were further subdivided into stages:

- **Part 1: exploratory analysis:**
  - **stage 1:** preparation of learning and validation datasets,
  - **stage 2:** statistical description of the spatial distribution of the loss statistics,
  - **stage 3:** visualization of loss statistics using cartograms,
  - **stage 4:** calculation of measures describing the spatial distribution of insurance risk, including the spatial autocorrelation coefficients of Moran, Geary and Getis-Ord,
  - **stage 5:** initial selection of explanatory variables based on their relationship with the dependent variable and construction of the best model without geographical variables (referred to as the "base model").
- **Part 2: estimation of parameters of spatial models:**
  - **stage 1:** analysis of the results of the estimation of the base model parameters,
  - **stage 2:** estimation of MLF model parameters with various definitions of regions,
  - **stage 3:** estimation of parameters of HMLF models with various definitions of regions,
  - **stage 4:** estimation of parameters of SMLF models,
  - **stage 5:** carrying out spatial segmentation and analysis of models taking into account the resulting tariff variables.
- **Part 3: comparing the effectiveness of models with spatial effects:**
  - **stage 1:** assessment of the effectiveness of models with a multi-level factor using adopted criteria,

- **stage 2:** assessment of the effectiveness of models including tariff variables obtained as a result of spatial segmentation.

5. Results and conclusions

The results of the descriptive analysis lead to the conclusion that the insurance risk is geographically heterogenous (as evidenced by the spatial distribution of the loss statistics under examination). The claim frequency distribution, which is the main subject of this study, is presented in Fig. 1. The data are presented according to the address of the oldest vehicle owner, and in addition they were presented in the form of an index, i.e. the original value for the region was divided by the statistics value for the entire market.

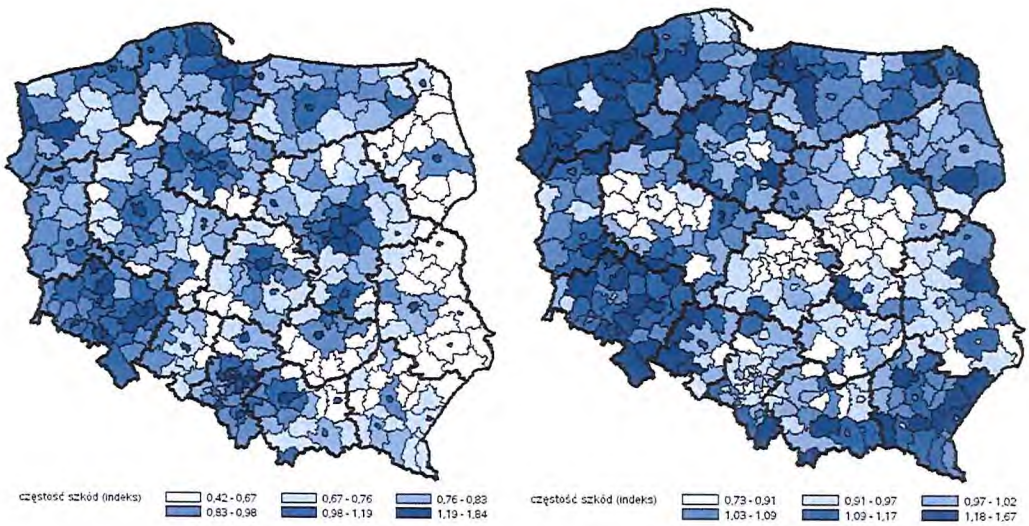

Fig. 1. Index of claim frequency by district for MTPL (left) i MOD (right).

In the case of MTPL there is a significant impact of urban hubs on the insurance risk, while in the case of MOD the western voivodships stand out. Insurance risk is also characterized by spatial autocorrelation, as evidenced by the high values of Moran statistics (over 0.4 in the case of claim frequency). It is also worth emphasizing that at a high level of detail in the geographical breakdown (municipalities, postcodes), it is becoming apparent that most insurance companies do not have data covering all regions, and some regions have a small sample size. This is the justification for considering Bayesian methods and models with random effects.

The results of predictive modeling indicate that the base model includes a wide range of data, including both the basic data of the vehicle and its owner, as well as data on the insurance history, including the length of the claim-free period and the number and frequency of claims. However, the base model is not able to fully explain the observed spatial diversity of data, as evidenced by the distribution of random effects estimated using MLF models at the district, municipality and postal code level. In turn, the HMLF and SMLF models indicate that the correlation between the number of claims arising from concluded contracts occurs not only within the region (which is inferred from the significance of the region as a tariff variable), but also between regions. HMLF models indicate a significant relationship between insurance risk in regions belonging to the same parent region (in particular dsitricts within the same voivodship), while the SMLF model indicates a positive autocorrelation between random effects at the district level. Finally, the results of spatial segmentation indicate the possibility of constructing tariff variables that, when included in the base model, improve its fit to the data, while the quality of such a model is comparable to the base model with an attached voivodship.

The efficiency criteria of the tariff model used in this dissertation allow us to state that in most cases taking into account spatial effects allows for a better risk assessment in comparison to the base model. It should be emphasized, however, that the results obtained do not allow the selection of a model that would be the best in terms of each criterion. The increase in efficiency achieved due to taking into account spatial effects using the considered models varies significantly depending on the type of contract and the criterion under investigation. As for models with a multilevel factor, in the case of MOD contracts good results are obtained by the HMLF model with grouping of postal codes based on the first two digits (portfolio and sub-portfolio level), as well as MLF and HMLF models based on the municipality (individual level). In the case of MTPL it is worth highlighting models at the municipality level - HMLF at the sub-portfolio level and MLF at the individual level. MLF model for postal codes also gives comparable results. In turn, the most complex SMLF model in terms of stochastic structure proved to be the best only in the context of forecasting the number of damages for the entire portfolio, while with more detailed criteria for assessing the benefits of its use were not apparent. In the case of spatial segmentation, all considered methods are characterized by similar effectiveness. Both the extended base model as well as the graph method and the k-means method give similar results when it comes to the criteria for assessing the models adopted in this work. Their use allows to improve the accuracy of the forecast in geographical terms, while it has no significant impact on the forecast in cross-sections related to the vehicle and its owners.

The results of the study confirm the research hypotheses put forward in the work. The H1 hypothesis was confirmed by analyzing the geographical distribution of loss statistics (claim frequency, average loss, pure premium), including spatial autocorrelation coefficients. The H2 hypothesis was confirmed by analyzing the properties of MLF, HMLF and SMLF models. Estimation results indicate that in most cases the reliability of data for individual regions is too low and it is advisable to estimate spatial effects taking into account data for the entire market. The H3 hypothesis was confirmed by validating models on a data set that was not used to estimate the parameters of these models. While it is not possible to identify one model that would be the best according to all of the criteria under consideration, most of the presented spatial models allow to obtain better results than those obtained using the base model. In turn, the H4 hypothesis was confirmed by positive verification of the H1 - H3 hypotheses. The results of estimation of model parameters confirm the occurrence of spatial relationships, e.g. by significant influence of hierarchy in HMLF models or positive spatial autocorrelation between hidden risk factors in the SMLF model. In turn, the results of model validation indicate the advantages of including the geographical dimension in the analysis, which are associated with the improvement in the predictive capacity of the tariff model and can be transformed by the insurance company into measurable financial benefits. Bearing in mind the results of theoretical and empirical analyzes carried out, it should also be acknowledged that the main objective and specific objectives set out in this dissertation have been achieved.

The contribution of this work to the development of science is twofold. First, the methodological value of the work results from the definition of a spatial generalized linear model that can act as a common conceptual framework for a wide family of spatial models. A generalization of the classic Bühlmann-Straub model was also made to take into account the correlation between hidden risk factors, interpreted as spatial autocorrelation in this paper, and the properties of this model were examined and an approach to estimating its parameters was proposed. To the author's knowledge, such a model has not yet been described in actuarial literature in the context of insurance risk assessment, similarly the proposed method of spatial segmentation using a graph algorithm. It is worth emphasizing that the presented spatial models can be used to analyze any regional data, not only in the field of insurance. Secondly, the cognitive value results from the presentation of the results of analysis of data from the Insurance Guarantee Fund, which as the only entity has data on the entire Polish insurance market. The large volume of data results in greater reliability of results, while justifying the need for Bayesian methods, because even in the scale of the entire market, data on some regions cannot be considered sufficiently reliable from a statistical point of view.

Despite its wide scope, this dissertation does not exhaust the potential associated with the use of spatial data in actuarial analyzes, and it is possible to identify potential directions for further research. In terms of methodology, consideration can be given to considering GLM spatial variants based on the Markov spatial process (see Boskov and Verrall, 1994) - the estimation of these models is more complicated compared to models from the MLF family due to the need to use simulation methods. Furthermore, the SMLF model is open to various specifications of the variance-covariance matrix of unobserved risk factors and alternative methods for calibrating its parameters. The spatial segmentation methods also open a field for further analysis, as the segmentation process may take into account the different features of regions and relationship between them. An interesting direction may also be the analysis of point data regarding the place of occurrence of accidents. On the other hand, it is possible to extend the analysis to other loss statistics, in particular average loss and pure premium - in the first case the distribution of the dependent variable in the generalized linear model can be gamma or inverse Gaussian, while in the second - Tweedie's models (see Kaas et al., 2001). While the general procedure would be similar to claim frequency modeling, the variable selection algorithm itself and the criteria for assessing the effectiveness of models should be modified to take into account the continuous (in the case of average loss) or mixed discrete-continuous (in the case of pure premium) nature of the dependent variable. Carrying out such an analysis would be consistent with a practical approach involving simultaneous modeling of the number and amount of claims, and then combining the results into one pricing model. However, this would significantly extend the scope of this dissertation, without changing the main conclusions of the analysis.

6. Bibliography

1. Anselin L., 1988, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.
2. Anselin L., 1995, *Local Indicators of Spatial Association – LISA*, "Geographical Analysis", Vol. 27, No. 2.
3. Anselin, L., 1996, *The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association*, [w:] "Fischer, M., Scholten, H. and Unwin, D., Eds., Spatial Analytical Perspectives on GIS", Taylor and Francis, London, pp. 111-125.
4. Bailey R.A., Simon L., 1960, *Two studies in Automobile Insurance Ratemaking*, „ASTIN Bulletin", Vol. 1, pp. 192-217.
5. Bailey, R.A., 1963, *Insurance rates with minimum bias*, „Proceedings of the Casualty Actuarial Society" L, pp. 4–13.
6. Blondel V.D, Guillaume J.L, Lambiotte R., Lefebvre E., 2008, *Fast unfolding of communities in large networks*, „Journal of Statistical Mechanics: Theory and Experiment", nr 10.
7. Boskov, M., Verrall, R.J., 1994, *Premium rating by geographical area using spatial models*, „ASTIN Bulletin" 24, pp. 131-143.
8. Brouhns, N., Denuit, M., Masuy, B., Verrall, R., 2002, *Ratemaking by geographical area: A case study using the Boskov and Verrall model*, Discussion paper 0202, Publications of the Institut de statistique, Louvain-la-Neuve, pp. 1-26.
9. Brubaker R., Bylls R., 1997, *Geocoding: Description and Uses*, Health Care Issues for Property/Casualty Insurers 1997 Discussion Paper Program, Casualty Actuarial Society, pp. 153-164, online: https://www.casact.org/pubs/dpp/dpp97/97dpp.pdf [access: 31.05.2020].
10. Bühlmann H., Gisler A., 2005, *A Course in Credibility Theory and its Applications*, Springer-Verlag Berlin Heidelberg.
11. Cieślik B., 2017, *Zastosowania telematyki w ubezpieczeniach komunikacyjnych*, „Roczniki Kolegium Analiz Ekonomicznych", zeszyt 45/2017, pp. 79-92.
12. Denuit M., Dhaene J., Goovaerts M., Kaas R., 2005, *Actuarial Theory for Dependent Risks: Measures, Orders and Models*, John Wiley & Sons Ltd.
13. Denuit M., Maréchal X., Pitrebois S., Walhin J., 2007, *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, Wiley, New York.
14. Dionne G., Vanasse Ch., 1989, *A Generalization of Automobile Insurance Rating*

Models: The Negative Binomial Distribution with a Regression Component, „ASTIN Bulletin", Vol. 19, No. 2, pp. 199-212.

15. Finger, R.J., 2001, *Risk Classification*, [w:] "Foundations of Casualty Actuarial Science" 4th ed., Arlington, VA: Casualty Actuarial Society.
16. Francis L., 2014, *Unsupervised Learning*, [w:] „Predictive Modeling Applications in Actuarial Science, Volume I: Predictive Modeling Techniques" (red. Frees E., Derrig R., Meyers G.), Cambridge University Press, New York., pp. 280-314.
17. Frątczak, E. (red.), 2012, *Zaawansowane metody analiz statystycznych*, Szkoła Główna Handlowa, Warszawa.
18. Gala K., Kolak K., 2015, *Wykorzystanie informacji kredytowej w procesie oceny ryzyka ubezpieczeniowego w ubezpieczeniach komunikacyjnych*, „Roczniki Kolegium Analiz Ekonomicznych", zeszyt 37/2015, pp. 73-87.
19. Gala K., 2016, *Lojalność klientów w ubezpieczeniach komunikacyjnych w świetle danych UFG*, „Studia Ekonomiczne, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach", nr 301, red. E. Dziwok, K. Sawicz, A. Sroczyńska-Baron, pp. 43-62.
20. Gala K., 2017, *Taryfikacja a priori z uwzględnieniem efektów przestrzennych*, „Śląski Przegląd Statystyczny", nr 15/2017, pp. 99-114.
21. Gala K., Bobrowski M., 2017, *Lojalność klientów w ubezpieczeniach autocasco*, „Śląski Przegląd Statystyczny", nr 15/2017, pp. 81-98.
22. Gala K., 2018, *Taryfikacja a priori w ubezpieczeniach komunikacyjnych z uwzględnieniem zależności przestrzennej*, „Roczniki Kolegium Analiz Ekonomicznych", zeszyt 51/2018, pp. 147-168.
23. Geary R.C., 1954, *The Contiguity Ratio and Statistical Mapping*, „The Incorporated Statistician", Vol. 5, No. 3, pp. 115-146.
24. Getis A., Ord J.K, 1992, *The analysis of spatial association by distance statistics*, „Geographical Analysis", nr 24.
25. Griffith D.A., 1983, *The Boundary value problem in spatial statistical analysis*, „Journal of Regional Science", vol. 23, no. 3, pp. 377–387, doi: https://doi.org/10.1111/j.1467-9787.1983.tb00996.x.
26. Griffith D.A., 1992, *What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics*. „Espace géographique", tome 21, no. 3, pp. 265-280; doi: https://doi.org/10.3406/spgeo.1992.3091
27. Hastie, T., Tibshirani R. Friedman J., 2009, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, New York, NY: Springer.

28. Jewell W.S., 1975, *The use of collateral data in credibility theory: a hierarchical model*, „Giornale dell'Istituto Italiano degli Attuari", no. 38, pp. 1-16.
29. Kaas R., Goovaerts M.J., Dhaene J., Denuit M., 2001, *Modern Actuarial Risk Theory*, Kluwer Academic Publishers, Dordrecht.
30. Larose D., 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley, Hoboken, NJ.
31. Lee J.S., Lee I.S., 2016, *The study of insurance premium rate GIS mapping considering the storm and flood hazard risks,* The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B8, 2016 XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic, online: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B8/95/2016/isprs-archives-XLI-B8-95-2016.pdf [access: 31.05.2020].
32. Leja F., 1976, *Rachunek różniczkowy i całkowy*. PWN, Warszawa.
33. Lemaire J., 1995, *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers, Boston-Dordrecht-London, ISBN 0-7923-9545-X.
34. Lemaire J., Park S.C., Wang K.C., 2016, *The use of annual mileage as a rating variable*, "ASTIN Bulletin", vol. 46, iss. 1, pp. 39-69.
35. LeSage J., 1999, *The Theory and Practice of Spatial Econometrics*, Department of Economics, University of Toledo, Toledo.
36. LeSage J., Pace R. K., 2009, *Introduction to Spatial Econometrics*, BocaRaton, FL: CRC Press.
37. Loimaranta K., 1972, *Some asymptotic properties of bonus systems*, „ASTIN Bulletin", Vol. VI. Part 3, pp. 233-245.
38. Maddala G.S., 2001, *Introduction to Econometrics*, John Wiley & Sons, Ltd.
39. Mildenhall S., 1999, *A Systematic Relationship between Minimum Bias and Generalized Linear Models*, „Proceedings of the Casualty Actuarial Society" LXXXVI, pp. 393-486.
40. Moran P.A.P., 1950, *Notes on Continuous Stochastic Phenomena*, "Biometrika", Vol. 37, No. 1/2, pp. 17-23.
41. Nelder J.A., Wedderburn R.W.M., 1972, *Generalized Linear Models*, "Journal of the Royal Statistical Society", Series A (General), Vol. 135, No. 3 (1972), pp. 370-384.
42. Ohlsson, E., 2005, *Simplified estimation of structure parameters in hierarchical credibility*, The 36th ASTIN Colloquium, Zürich 2005.
43. Ohlsson E., Johansson B., 2010, *Non-Life Insurance Pricing with Generalized Linear Models*, Springer-Verlag Berlin Heidelberg.

44. Ostasiewicz W. (red.), 2004, *Składki i ryzyko ubezpieczeniowe. Modelowanie stochastyczne*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
45. Paelinck J, Klaassen L., 1979, *Spatial econometrics*. Saxon House, Farnborough.
46. Pinquet J., Guillén M., Bolancé C., 2001, *Allowance for the age of claims in bonus-malus systems*, "ASTIN Bulletin", 31(2), pp. 337-348.
47. Pinquet J., Dionne G., Vanasse C., Mathieu M., 2011, *Incentive Mechanisms for Safe Driving: A Comparative Analysis with Dynamic Data*. „The Review of Economics and Statistics", 93 (1), pp. 218-227.
48. Raichle W. M., 1997, *Insurance Geographics*, Health Care Issues for Property/Casualty Insurers 1997 Discussion Paper Program, Casualty Actuarial Society, pp. 141-152, online: https://www.casact.org/pubs/dpp/dpp97/97dpp.pdf [access: 31.05.2020].
49. Ronka-Chmielowiec W. (red.), 2013, Metody *aktuarialne. Zastosowania matematyki w ubezpieczeniach*, Wydawnictwo Naukowe PWN, Warszawa.
50. Siddiqi N., 2006, *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, John Wile & Sons, Inc, Hoboken, New Jersey.
51. Sobiecki D., 2013, Dwustopniowe modelowanie składki za ubezpieczenie komunikacyjne OC, "Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu", nr 312, 2013, pp. 116 – 134.
52. Suchecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H. Beck, Warszawa.
53. Szymańska A., 2014, *Statystyczna analiza systemów bonus-malus w ubezpieczeniach komunikacyjnych*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
54. Śliwiński A., 2012, *Przestrzenne zróżnicowanie ryzyka ubezpieczeniowego a efektywność ubezpieczeń na życie*, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie, Warszawa.
55. Taylor, G., 1989, *Use of spline functions for premium rating by geographic area*. „ASTIN Bulletin", 19, pp. 91–122.
56. Taylor, G., 2001, *Geographic Premium Rating by Whittaker Spatial Smoothing*. „ASTIN Bulletin", 31(1), pp. 147–160. doi:10.2143/AST.31.1.999
57. Tobler W., 1970, *A computer model simulating urban growth in the Detroit region*, „Economic Geography", nr 46 (2).
58. Weibel E., Walsh J., 2008, *Territory Analysis with Mixed Models and Clustering*, Casualty Actuarial Society Discussion Paper Program.

59. Werner, G., Modlin, C., 2016, *Basic Ratemaking*, Casualty Actuarial Society, Fifth Edition, online: https://www.casact.org/library/studynotes/Werner_Modlin_Ratemaking.pdf [access: 31.05.2020]
60. Wolthuis, H., 2003, *Life Insurance Mathematics (The Markovian Model)*, IAE, Universiteit van Amsterdam.
61. Wolny-Dominiak A., Wanat S., Sobiecki D., 2018, *Selected Statistical Models in Non-Life Insurance with R*, LAP LAMBERT Academic Publishing.
62. Yao J., 2016, *Clustering in General Insurance Pricing*, [w:] „Predictive Modeling Applications in Actuarial Science, Volume II: Case Studies in Insurance" (red. Frees E., Derrig R., Meyers G.), Cambridge University Press, New York., pp. 159-179.

Web sources and legal acts:

1. *Biuletyn roczny. Rynek ubezpieczeń 2018*, Komisja Nadzoru Finansowego, https://www.knf.gov.pl/?articleId=67227&p_id=18 (access: 31.05.2020).
2. *Stanowisko Komisji Nadzoru Finansowego z dnia 28 września 2015 r. w sprawie kalkulacji wysokości składki ubezpieczeniowej,* Komisja Nadzoru Finansowego, https://www.knf.gov.pl/knf/pl/komponenty/img/kalkulacja_wysokosci_skladki_ubezpieczeniowej_28-09-15_42937.pdf (access: 31.05.2020)
3. *Słownik języka polskiego,* Polskie Wydawnictwo Naukowe, https://sjp.pwn.pl/sjp/efektywny;2556022.html (access: 31.05.2020)
4. Ustawa z dnia 23 kwietnia 1964 r. - Kodeks cywilny (*Act of 23 April 1964 – Civil code*), Dz.U. 2019 poz. 1145 as amended.
5. Ustawa z dnia 11 września 2015 r. o działalności ubezpieczeniowej i reasekuracyjnej (*Act of 11 September 2015 on insurance and reinsurance activity*), Dz.U. 2020 poz. 895 as amended.
6. Ustawa z dnia 22 maja 2003 r. o ubezpieczeniach obowiązkowych, Ubezpieczeniowym Funduszu Gwarancyjnym i Polskim Biurze Ubezpieczycieli Komunikacyjnych (*Act of 22 May 2003 on compulsory insurance, Insurance Guarantee Fund and Polish Motor Insurers' Bureau*), Dz.U. 2019 poz. 2214 as amended.

Kamil Gala