Przegląd Statystyczny. Statistical Review, 2023, vol. 70, 4, 73–85 https://doi.org/10.59139/ps.2023.04.3
Received: 25.06.2024, revised: 14.10.2024, accepted: 15.10.2024

Computational problems in variable selection for multilevel models using stepwise regression

Hanna Wdowicka[a]

Abstract. Multilevel modelling is a methodology that allows the consideration of variability in the level of the studied variables and the nature of the relationships between them, depending on the affiliation of study units to higher-level units (groups). Additionally, by dividing the studied population into groups, it is possible to explain part of the variability of the estimated characteristic using higher-level characteristics. The usefulness of multilevel modelling in estimating socio-economic characteristics was investigated in the author's previous works. However, with large populations characterised by a multilevel structure, a significant drawback of this approach is its high computational complexity, often resulting in unacceptably long computation times. The main objective of the article is to propose a simplification in the algorithm of forward stepwise multilevel regression, allowing a significant reduction in the time required for variable selection in the model. The considerations will be illustrated by constructing a multilevel model to examine the determinants of daily flows related to employment based on the matrix of employment-related population flows developed from the 2021 National Census of Population and Housing (NSP 2021).

Keywords: multilevel modelling, multilevel structure, random effects, cross-model, commuting to work

JEL: C51, C52, C55

1. Introduction

The idea of multilevel modelling emerged in the early 1970s, when attention was drawn to the fact students within the same school and students from different schools displayed different levels of academic achievement. D. Lindley and A. Smith developed general frameworks for studying nested data with complex structures of random errors (Lindley & Smith, 1972). Incorporating dependencies among units at the first level belonging to the same units at higher levels significantly improves estimation precision compared to classical linear regression, provided the estimated variable has a multilevel structure (Hox, 2010). The applicability of multilevel modelling to estimating socio-economic characteristics was analysed in works such as Gruchociak (2012b), Suchecka and Łaszkiewicz (2017) or Węziak (2007).

[a] Poznań University of Economics and Business, Department of Statistics, al. Niepodległości 10, 61-875 Poznań, Poland, e-mail: hanna.wdowicka@ue.poznan.pl, ORCID: https://orcid.org/0000-0003-2045-9831.

The aim of this paper is to present the computational challenges associated with constructing highly complex multilevel models, along with the proposed solutions. To achieve this, the first part of the article will introduce the empirical problem that underlies the construction of this kind of models. A brief overview of the concept of multilevel modelling will follow. Subsequently, we will explain why this method was chosen to address the specific empirical problem. We will then discuss the construction of appropriate models from a formal point of view. In the final part, we will show any computational difficulties we encountered and will propose relevant solutions.

The empirical aim of the study was to analsye the determinants of daily commuting flows and the relationships between them.

In Poland, the analysis of daily commutes to work is possible thanks to the results of the study carried out periodically by the City Statistics Center of the Statistical Office in Poznań. So far, the results of four editions of the study have been published, for the years: 2006, 2011, 2016 and 2021 (Filas-Przybył & Stachowiak, 2019; Kowalewski, 2014, 2024; Kruszka, 2010). The primary data sources for the first three editions were the Ministry of Finance's tax registers. Data from the National Census of Population and Housing in 2011 and from the Social Insurance Institution were also used (Filas-Przybył & Stachowiak, 2019; Kowalewski, 2024). The National Census of Population and Housing 2021 applied a mixed-method approach, combining data collected from respondents with administrative data (Łysoń, 2024), which was also the case regarding the latest edition of the commuting survey. All four editions of the survey present data at the gmina (the smallest administrative unit in Poland, alternatively referred to as a commune) level, and more specifically for pairs of gminas, with further breakdowns to urban and rural data in the case of urban-rural gminas.

In the current study, the latest edition of the above-mentioned survey was analysed, specifically the 'Commuting to work' survey, based on the results of the National Census of Population and Housing 2021 (Kowalewski, 2024).

2. Multilevel Modelling Methodology

The methodology of multilevel modelling allows the consideration of similarities among units at the first level of the analysis that belong to the same groups formed by a grouping variable at the second and higher levels (Bates, 2010; Biecek, 2011; Raudenbush & Bryk, 2002). Unlike classical linear regression, multilevel modelling does not assume that all observations are independent; instead, it acknowledges the dependence among units at the first level that belong to the same units at higher levels (Twisk, 2010). Failing to account for such dependencies leads to underestimated standard errors (Hox, 2010; Klimanek, 2003). Thus, multilevel models capture two types of variability: differences among units at the first level belonging to the same units at higher levels, and differences among higher-level units themselves (Frątczak & Mianowska, 2012). Incorporating dependencies among units at the first level that belong to the same higher-level units significantly improves the precision of estimation compared to classical linear regression, provided the estimated variable has a multilevel structure. This improvement occurs regardless of whether the classical regression model disregards the division of first-level units entirely or treats observations belonging to the same groups as a whole (with estimates conducted for entire groups) (Twisk, 2010).

Additionally, by dividing the studied population into groups, it becomes possible to explain some of the variability of the estimated characteristic using the characteristics from the higher levels. The need for aggregating information available at different levels is also mentioned in the work of Bołt et al. (1985).

It should be emphasised that the use of multilevel modelling methodology is justified only for specific populations and variables. Both the population and the variable should have a multilevel structure. Regarding the population and the classical multilevel model, this means that the population can be divided into a finite number of distinct and collectively exhaustive groups that cover all units at the first level (alternatively referred to as units at the second level) (Goldstein, 2003; Hox, 2010; Łaszkiewicz, 2016).

In the case of a model with more than two levels, units at the second level can also be divided into distinct and collectively exhaustive groups that cover the entire population (further referred to as units at the third level), and so on.

If a model has two grouping criteria (known as a cross-level model), units at the second level are defined twice, with each first-level unit belonging to exactly one second-level unit defined by each of the two grouping criteria. For units at the second level defined by both criteria, higher-level units can also be defined (Bates, 2010; Biecek, 2011).

Using a multilevel model is justified when the estimated variable is of a multilevel structure. This means that its value should significantly differ between groups, i.e. units, at each of the higher levels. This variability can stem from a direct relationship between the variable of interest and the fact that the first-level unit belongs to higher-level units. A classic example from the literature illustrating such a scenario is the variability in academic achievement, caused both by individual students' abilities and predispositions (factors at the first, individual level) and teacher qualifications and teaching methods (factors at the second, group level) (Goldstein, 2003; Hox, 2010).

Another reason for the variability of the studied variable between groups can be the relationship between this variable and the division into groups based on a certain hidden, often unmeasurable, variable. An example of this is the relationship between the economic activity and the level of regional development (Gruchociak, 2012a, 2012b). Geographical distribution of factors determining labor demand, such as the presence of natural resources, industrial facilities, development of technical, communication, and educational infrastructure, are significant determinants influencing the economic activity of the population, alongside factors affecting the degree of entrepreneurship among individuals.

If the dependent variable has a multilevel structure, employing an appropriate multilevel model can significantly improve the quality of the analysis (Goldstein, 2003; Hox, 2010; Raudenbush & Bryk, 2002; Twisk, 2010).

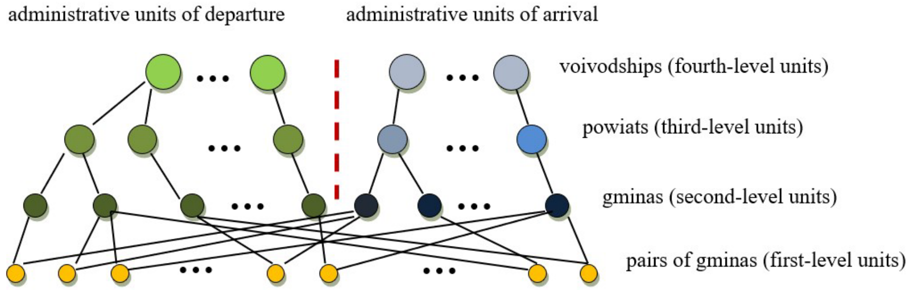
3. Empirical problem

3.1. Dependent variable and multilevel structure

The study focuses on daily commuting patterns between pairs of gminas. As mentioned earlier, the analysis of daily employee mobility is based on the results of the 'Commuting to work' survey using data from the 2021 National Census of Population and Housing (Kowalewski, 2024). Information regarding the number of individuals commuting to work is available at the gmina level, with a breakdown to an urban and a rural part in the case of urban-rural gminas. There were 95,890 pairs of gminas[1] with non-zero[2] employment-related flows, and these pairs are treated in our study as first-level units. The administrative divisions of the country naturally create a multilevel structure (voivodships consist of powiats, which further down consist of gminas). Given that the first-level units of the analysis are defined as pairs of gminas in this study, we are dealing with a cross-level multilevel structure with two grouping criteria defined at the second level. The second-level units can be identified as the gminas from which the commuting flows originate and the gminas of destination, thereby establishing two grouping criteria at the second level. Powiats and voivodships are considered as third- and fourth-level units respectively, due to both grouping criteria (see Figure 1).

[1] Throughout this article, a gmina is understood as a commune divided into urban and rural parts.
[2] Due to statistical confidentiality, the flows between communes with fewer than three people were disregarded.

Figure 1. Multilevel structure of administrative units in a cross-model with two grouping criteria defined the at the second level



Source: author's work.

The above-described structure of the studied community makes it possible to try to explain the number of individuals commuting between the gminas using a cross-level multilevel model with two grouping criteria (related to residential and workplace territorial units), and subsequent levels defined as successive territorial units (gmina, powiat, voivodship) determined by both grouping criteria. Obviously, the inclusion of the multilevel structure in the model will be preceded by verifying whether the studied characteristic of the commuting patterns has a multilevel structure, based on the described grouping criteria.

3.2. Explanatory variables

The distance between the place of residence and the place of work is regarded as one of the most crucial factors influencing the commuting intensity (Gumuła et al., 2007). Therefore, the distance between the gmina of residence and the gmina of workplace was adopted as the explanatory variable at the first level.

The number of persons commuting to work between gminas is naturally affected by the sizes of both gminas. Therefore, the characteristics related to these sizes constitute the second group of potential explanatory variables for the commuting intensity (see Table). The volume of industry within a gmina might also significantly impact commuting patterns, and thus the third group of potential explanatory variables could emerge. It is also plausible that the characteristics related to the attractiveness of a particular place and the relative ease of commuting from there would significantly influence the commuting intensity, so we put these factors together as the fourth group of potential explanatory variables (see Table).

Variations in the intensity of commuting to the workplace might also depend on the attractiveness of the employment conditions both in the 'departure' and the 'destination' gminas. According to the literature, a comprehensive set of characteristics illustrating the labour market situation should comprise variables representing both the labour demand and the supply (Gołata, 2004). Therefore, the set of potential explanatory variables has been expanded to include three additional groups describing the labour market, focusing in particular on the supply, demand and pricing characteristics (see Table).

Table. A set of potential explanatory variables for commuting to work between gminas

distance	
X	distance between gminas[3]
gmina size	
Cga1 ,Cgb1	area of the gmina in square kilometers
Cga2 ,Cgb2	number of people of working age
Cga3 ,Cgb3	number of working people
industry volume	
Ega1 ,Egb1	number of national economy entities from the public sector
Ega2 ,Egb2	number of national economy entities from the private sector
Ega3 ,Egb3	number of commercial companies with foreign capital
Ega4 ,Egb4	number of natural persons conducting economic activity
Ega5 ,Egb5	number of national economy entities
Epa1 ,Epb1	enterprises' investment outlays per person of working age
location	
Lga1 ,Lgb1	distance from the nearest voivodship capital
Lga2 ,Lgb2	distance from the nearest metropolis[4]
Lga3 ,Lgb3	number of parking lots in the Park & Ride system
Lpa1 ,Lpb1	investment outlays of enterprises in thousands of PLN
Lva1 ,Lvb1	number of railway lines

[3] Measured in a straight line between the centroids of gminas (Kopczewska, 2006).
[4] The metropolis were selected using the following procedure: all gminas were ranked in descending order according to the number of people employed. In the first step, the set of central centers was defined as the city with the largest number of employees, then the set was expanded by subsequent gminas ranked according to the number of employees. For the sets of large cities defined in this way in subsequent stages, the correlation between the distance and the intensity of trips to work was examined. Then we checked for which of the defined sets of large cities specified in individual steps the relationship between the distance and the intensity of trips to work was the strongest. In line with Thünen's theory that central centers stimulate the development of areas surrounding them, it was assumed that the distance from each gmina to the nearest central center should influence the intensity of trips to work. Ultimately, a set of large cities was selected for which the correlation relationship was the strongest. According to the above procedure, the following seven metropolis were identified: Warsaw, Krakow, Wroclaw, Poznan, Lodz, Gdansk and Katowice.

Table. A set of potential explanatory variables for commuting to work between gminas (cont.)

demand	
Dga1 ,Dgb1	ratio of working people to the working-age population
Dga2 ,Dgb2	number of working people per square kilometer
Dga3 ,Dgb3	number of people of working age per square kilometer
Dva1 ,Dvb1	share of graduates of public universities among the working-age population
Dva2 ,Dvb2	share of university graduates among the working-age population
supply	
Sga1 ,Sgb1	intensity[5] of national economy entities from the public sector
Sga2 ,Sgb2	intensity of national economy entities from the private sector
Sga3 ,Sgb3	intensity of commercial companies with foreign capital
Sga4 ,Sgb4	share of natural persons conducting economic activity in the working-age population
Sga5 ,Sgb5	intensity of national economy entities
Sga6 ,Sgb6	share of the unemployed among the working-age population
price	
Ppa1 ,Ppb1	average monthly gross salary

Note. The first letter of the subscript indicates the level at which this characteristic is available; g- gmina, d-powiat, v-voivodship. The second letter of the subscript determines the grouping criterion of a given territorial unit; a-territorial unit of residence, b-territorial unit of the workplace.

Source: author's work based on data sets published by GUS.

Among the potential determinants of commuting distances to work, various characteristics available at different levels of aggregation have been considered. Therefore, it seems particularly justified to attempt an analysis of the determinants of labour mobility using the modelling that takes into account the multilevel structure of the labour market.

4. Research procedure

To explain how the number of commutes between pairs of gminas is determined, a four-level cross model with two grouping criteria at the second level was constructed. Units at the first level were defined as pairs of gminas, while units at the second level were gminas from which commutes originated and the destination ones, thereby defining two grouping criteria at the second level. Powiats and voivodships were adopted as third- and fourth-level units taking into account both grouping criteria (see Figure 1). Thus, we considered the incorporation of up to six different grouping criteria defined by the gmina, powiat and voivodship of residence, as well as the gmina, powiat and voivodship the workplace. Additionally, a set of 53 potential explanatory variables defined at various levels (see Table),

[5] Understood here as the ratio of the number of national economy entities from the public sector to the number of working-age inhabitants; it is understood analogically elsewhere in Table.

interactions among these variables, and variations in the influence of individual explanatory variables on the number of commuters depending on their affiliation with higher-level grouping units were considered.

As we can see, designing an optimal multilevel model involves considering numerous potential model extensions for this research problem. Applying stepwise regression to all these extensions simultaneously is impossible due to the extensive computational time required.

This is a common problem in multilevel modelling, therefore the construction of a multilevel model is usually carried out in stages, which makes it possible to shorten the time of selecting extensions for the model (Bliese, 2022; Hox, 2010; Twisk, 2010). The analysis of determinants of commuting to work was conducted in the stages described below.

At first, expanding the fixed part of the model (i.e., without accounting for random components arising from the multilevel data structure) was carried out in five steps. In the following three stages, a random part of the model associated with grouping effects was included. This aimed to initially account for effects that can be measured (i.e. fixed effects). It appears that if the level of commuting varies across certain territorial units, but these differences can be explained by measurable explanatory variables, it should not be considered as a random factor, but it rather should be explained using these variables. Another advantage of this sequential model expansion is that after incorporating random components, the estimation time for the subsequent models with considered extensions significantly increases. Therefore, adding them in the final stages allows a substantial reduction in the total computation time.

The construction of the multilevel model was preceded by verifying the hypothesis of the multilevel structure of the estimated variable, conducted using an analysis-of-variance test (Krzyśko, 1996).

STAGE 1

In the first step, explanatory variables were introduced independently of the aggregation level for which they were defined. A total of 53 explanatory variables were analysed.

STAGE 2

In the second step, we introduced the interactions between the distance between the gminas of residence and of the workplace and the remaining potential explanatory variables defined for territorial units specified by both the place of residence and the workplace. This was done regardless of whether they were included or not in step 1. If the influence of a particular explanatory variable on commuting propensity varies depending on the commuting distance, including their interactions should improve the model fit. In such cases, we can conclude that the distance moderates the impact of an explanatory variable on the number of people commuting to work. A total of 52 interactions were analysed in this stage.

STAGE 3

In stage 3, the interaction between the characteristics of the gmina of residence and the gmina of work was taken into consideration. This enabled us to indicate in what ways the characteristics of the gmina of residence moderate the impact of the characteristics of the gmina of work on attracting commuters to it. Interactions between each pair of explanatory variables were considered, i.e. 20 x 20 = 400 potential extensions.

STAGE 4

Then, we were considering the inclusion of interactions between the characteristics of the same gmina (residence or work). This enabled us to indicate how certain characteristics of the gmina of residence affect the impact of other characteristics of the same gmina on the intensity of departures from this gmina, and how some characteristics of the gmina of the workplace affect the impact of other characteristics of the same gmina on the intensity of arrivals to this gmina. Interactions between each pair of explanatory variables were analysed, both for the gminas of residence and of the workplace, i.e. 2 x 20 x 19 / 2 = 380 potential extensions.

STAGE 5

Then we were considering taking into account the interactions between the characteristics of territorial units from different levels but according to the same grouping criterion (i.e. for territorial units of residence or of the workplace). This enabled us to determine in what ways the characteristics of higher-level residence units affect the impact of the characteristics of lower-level residence units on the intensity of trips to work, and how the characteristics of higher-level work units affects the impact of the characteristics of lower-level work units on the intensity of trips to work. Therefore, interactions between each pair of explanatory variables were analysed, both for territorial units of residence and of the workplace, i.e. 2 x (3 x 20 + 3 x 20 + 3 x 3) = 258 potential extensions.

STAGE 6

Then the construction of the random part of the model began. In stage 6, the impact of each grouping factor on the level of commuting to work, i.e. six potential random effects, were examined.

STAGE 7

In this stage, the differences in the impact of the commuting distance on the number of people commuting to work across the territorial units of residence and of the workplace of commuters were taken into account. Therefore, we were considering the introduction of six further random effects.

STAGE 8

In the last stage, we were contemplating taking into account the differences in the impact of other explanatory variables on the number of people commuting to work across higher-level territorial units created due to the same grouping criteria as the examined explanatory variable. Therefore, we were considering the introduction of another 2 x (2 x 20 + 3) = 86 random effects.

Within the individual stages, extensions were introduced according to forward stepwise regression procedure, with the result of the likelihood ratio test as the improvement criterion. In most of the stages (except stages 2, 3, 4 and 5), the improvement was considered statistically significant if the p-value did not exceed 0.05. The above-mentioned exceptions occurred because during including interactions between variables in the model, special caution was recommended, so for these stages the significance level was 0.01.

After having completed each stage, we checked whether any of the previously added (in this stage or any previous one) extensions still improves the quality of the model fit (measured using the likelihood ratio test, p-value = 0.05) in a statistically significant way. If any of the previously-added model extensions did not meet this condition, they were removed from the model in steps. Among other developmants, this approach allows, at least to some extent, the replacement of a model extension introduced in an earlier stage by the extension from a later stage that better explains a given part of the variability in commuting.

Before starting the calculations, all explanatory variables were centered, which is a standard procedure in multilevel modelling. In addition, variables measured on very different scales were rescaled to avoid convergence problems.

The calculations were performed in the R program using the lme4 library (Bates, 2013; Bates et al., 2015; Bates, 2024), in which the parameters of multilevel models are estimated using sparse matrices. The code of the described algorithm is available in the appendix of this article.

5. Calculational problems and proposed solutions

Despite the construction of the model divided into eight stages, starting with the construction of the less time-consuming constant part of the model and the use of the effective lme4 library of the R program, the number of extensions considered, especially the interactions between the potential explanatory variables, resulted in unacceptably long calculation times[6]. Therefore, we decided to use a simplified

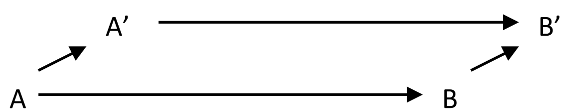
[6] The computation time has been estimated at many billions of years.

forward stepwise regression procedure in selected stages (those concerning the interactions between variables).

In order to shorten the calculation time, we tried to modify the forward stepwise regression procedure used. Namely, we assumed that if a certain extension of the model does not significantly improve the quality of the model fitting, then after enriching the model with other elements, adding the same extension as before, will not improve the quality of its fitting either.

The adopted simplification can also be written in a formal way. Let model B be an extension of model A, and models A' and B' be extensions of models A and B, respectively, with the same component. If the likelihood ratio test does not indicate a significant improvement of model A' in relation to model A, then model B' is not significantly better than model B according to this criterion (see Figure 2).

Figure 2. A diagram for the gradual expansion of multilevel models and the relationships between them



Source: author's work.

It should be noted that this simplification was adopted only in the stages in which interactions between variables were considered (i.e. stages 2, 3, 4, 5, 7 and 8), while in the remaining stages, the classic forward stepwise regression algorithm was used. It seems that the risk of disregarding an interaction between variables that would become significant only after taking into account another interaction at the same stage is relatively small. Additionally, it has to be remembered that no stepwise regression procedure can guarantee the selection of the optimal model; to be sure, models with all possible subsets of the considered extensions should be considered. Also, extending multilevel models within stages, although widely used (Bliese, 2022; Hox, 2010; Twisk, 2010), involves the risk of disregarding extensions that could become important at some point in the construction of the model.

It is worth remembering that thanks to the use of such a simplification, the number of models estimated successively during the forward stepwise regression procedure, e.g. in the third stage, decreased from 400! to less than a thousand, and in stage 8 instead of 86! multilevel models[7], it was enough to estimate just over 100

[7] Estimation of one multilevel model with a complexity level corresponding to stage 8 takes about 10 minutes, so it can be calculated that the estimation 86! such models would take many trillions of years.

such models, which allowed the calculations to be carried out in a finite time. In this context, the adoption of the above-mentioned simplification seems to be totally justified.

6. Conclusions

In conclusion, it can be said that the procedure of model construction that has been carefully thought over and the use of the proposed simplification in the forward stepwise regression procedure allowed the calculations to be carried out in an acceptable time. A discussion of the results obtained regarding the determinants of commuting between gminas will be discussed in detail in a separate article.

References

Bates, D. (2010). *lme4: Mixed-effects modeling with R*. Springer.

Bates, D. (2024). *Computational methods for mixed models*. http://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf.

Bates, D. (2013). *Linear mixed model implementation in lme4*.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Biecek, P. (2011). *Analiza danych z programem R. Modele liniowe z efektami stałymi, losowymi i mieszanymi*. Wydawnictwo Naukowe PWN.

Bliese, P. (2022). *Multilevel Modeling in R (2. 7). A Brief Introduction to R, the multilevel package and the nlme package*. https://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf.

Bołt, T. W., Krauze, K., & Kulawczuk, T. (1985). *Agregacja modeli ekonometrycznych*. Państwowe Wydawnictwo Ekonomiczne.

Filas-Przybył, S., & Stachowiak, D. (2019). *Przepływy ludności związane z zatrudnieniem w 2016 r.* Urząd Statystyczny w Poznaniu. https://stat.gov.pl/obszary-tematyczne/rynek-pracy/opracowania/przeplywy-ludnosci-zwiazane-z-zatrudnieniem-w-2016-r-,20,1.html.

Frątczak, E., & Mianowska, M. (2012). Modele mieszane. In E. Frątczak (Ed.), *Zaawansowane metody analiz statystycznych* (pp. 223–380). Oficyna Wydawnicza SGH.

Goldstein, H. (2003). *Multilevel Statistical Models* (3rd edition). Edward Arnold.

Gołata, E. (2004). *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*. Wydawnictwo Akademii Ekonomicznej w Poznaniu.

Gruchociak, H. (2012a). Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (242), 590–600.

Gruchociak, H. (2012b). Możliwości zastosowania modelowania dwupoziomowego w badaniach ekonomicznych. *Przegląd Statystyczny. Statistical Review*, *59*(4), 409–433. https://doi.org/10.59139/ps.2012.04.4.

Gumuła, W., Socha, J., & Wojciechowski, W. (2007). *Presja płacowa oraz niedopasowanie strukturalne na rynku pracy w świetle badań NBP*. Narodowy Bank Polski. https://static.nbp.pl/publikacje/materialy-i-studia/ms219.pdf.

Hox, J. J. (2010). *Multilevel Analysis. Techniques and Applications* (2nd edition). Routledge. https://doi.org/10.4324/9780203852279.

Klimanek, T. (2003). *Wielopoziomowa analiza struktury agrarnej gminy w systemie Geo-Info* [rozprawa doktorska, Akademia Ekonomiczna w Poznaniu]. https://bazawiedzy.ue.poznan.pl/info/phd/UEP844a7d3f538c48f4ba20dbebc9ee7803/.

Kopczewska, K. (2006). *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R-CRAN*. CeDeWu.

Kowalewski, J. (Ed.). (2014). *Dojazdy do pracy. Narodowy Spis Powszechny Ludności i Mieszkań 2011*. Główny Urząd Statystyczny. https://stat.gov.pl/spisy-powszechne/nsp-2021/nsp-2021-wyniki-ostateczne/dojazdy-do-pracy-nsp-2011,8,1.html.

Kowalewski, J. (Ed.). (2024). *Dojazdy do pracy w świetle wyników Narodowego Spisu Powszechnego Ludności i Mieszkań 2021*. Główny Urząd Statystyczny, Urząd Statystyczny w Poznaniu. https://stat.gov.pl/spisy-powszechne/nsp-2021/nsp-2021-wyniki-ostateczne/dojazdy-do-pracy-w-swietle-wynikow-nsp-2021,8,2.html.

Kruszka, K. (Ed.). (2010). *Dojazdy do pracy w Polsce. Terytorialna identyfikacja przepływów ludności związanych z zatrudnieniem*. Główny Urząd Statystyczny, Urząd Statystyczny w Poznaniu. https://stat.gov.pl/obszary-tematyczne/rynek-pracy/opracowania/dojazdy-do-pracy-w-polsce-2010-na-podstawie-danych-administracyjnych,8,1.html.

Krzyśko, M. (1996). *Statystyka matematyczna*. Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society, Series B (Methodological)*, *34*(1), 1–41.

Łaszkiewicz, E. (2016). *Ekonometria przestrzenna III. Modele wielopoziomowe – teoria i zastosowania*. Wydawnictwo C.H. Beck.

Łysoń, P. (Ed.). (2024). *Ludność na rynku pracy w świetle wyników Narodowego Spisu Powszechnego Ludności i Mieszkań 2021 – wersja uaktualniona*. Główny Urząd Statystyczny. https://stat.gov.pl/spisy-powszechne/nsp-2021/nsp-2021-wyniki-ostateczne/ludnosc-na-rynku-pracy-w-swietle-wynikow-narodowego-spisu-powszechnego-ludnosci-i-mieszkan-2021-wersja-uaktualniona,11,3.html.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and Data Analysis Methods* (2nd edition). Sage Publications.

Suchecka, J., & Łaszkiewicz, E. (2017). Wykorzystanie modelowania wielopoziomowego w analizie jakości życia. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (344), 133–144. https://www.ue.katowice.pl/fileadmin/user_upload/wydawnictwo/SE_Artyku%C5%82y_341_360/SE_344/08.pdf.

Twisk, J. W. R. (2010). *Analiza wielopoziomowa – przykłady zastosowań. Praktyczny podręcznik biostatyki i epidemiologii*. Oficyna Wydawnicza SGH.

Węziak, D. (2007). Wielopoziomowe modelowanie regresyjne w analizie danych. *Wiadomości Statystyczne*, *52*(9), 1–12.