SEM-Tree hybrid models in the preferences analysis of the members of Polish households

Adam Sagan¹ · Mariusz Łapczyński¹

Received: 27 June 2019 / Revised: 13 July 2020 / Accepted: 31 July 2020 / Published online: 10 August 2020
© The Author(s) 2020

Abstract

The purpose of the paper is to identify the dimensions of the strategy of resources allocation of Polish households members and test the hypothesis concerning risky shift effect in the relationship between strategy of family decision making and trade-off in family scarce resources allocation. These dimensions were identified on the basis of nationwide empirical data gathered on a representative sample of 1020 respondents nested in 410 households. SEM-Tree hybrid models are used in the analysis of the results, which combine the confirmatory structural equation models with exploratory and predictive classification and regression trees. This allows to apply structural modeling for the study of heterogeneous populations and to assess the hierarchical impact of exogenous predictors on the identification of segments with separate and unique model structural parameters. The approach combines the advantages of a model approach (at the stage of constructing hypotheses on structural relationships and specifications of measurement models) and exploration-based data (at the stage of recursive division of the sample).

Keywords Customers attitudes · Structural equation models · Decision trees · Hybrid model

Mathematics Subject Classification 91B42 · 91B10 · 91B32

The paper was prepared on the basis of the Research Grant of the National Science Centre No DEC-2011/01/B/HS4/04812.

✉ Adam Sagan
sagana@uek.krakow.pl

Mariusz Łapczyński
lapczynm@uek.krakow.pl

¹ Cracow University of Economics, Rakowicka 27, Cracow, Poland

1 Introduction

The methods used for data analysis are often divided into confirmatory methods, applicable to well-structured research problems and used to verify (falsify) theoretical models, and exploration methods that are related to solving unstructured problems and formulating theoretical proposals that emerge from the analyzed data. This division is supplemented with hybrid approaches that constitute the submission of methods, often belonging to different analytical traditions, using other methodological assumptions.

These types of hybrid methods are used both in such confirmatory approaches as structural modeling with latent variables (Structural Equation Modeling—SEM), as well as in exploratory approaches, which include the decision trees method.

Examples of hybrid approaches in structural modeling are the so-called automated structural models that are a combination of structural models and heuristic numerical procedures. Within them, SEM models are combined with heuristic ant algorithms (ACO-SEM), genetic algorithms (GA-SEM), tabu search (TS-SEM), destruction and reproduction (ruin) and-recreate—R and R-SEM), or simulated annealing (SA-SEM) (Marcoulides and Ing 2012; Sagan and Perek-Białas 2016).

Their application is associated with attempts to heuristic search for model specification (model search), which results from the exploratory nature of the process of searching for specifications, complexity of the model, variables and possible combinations of the number of potential dependencies between them. Without taking into account any theoretical assumptions, the number of possible SEM models built on a given covariance matrix based on k variables is $n = 4^{\frac{k(k-1)}{2}}$.

The idea of hybridization decision trees with other algorithms is not new. The combination of decision trees (CHAID algorithm) with logistic regression carried out by Lindahl and Winship (1994) was probably the first attempt to build this kind of a hybrid model. Hybridization was based on the sequential use of these analytical tools. After the initial exploration of data set by using CHAID algorithm cases were divided into terminal nodes.

In the second step of the procedure a separate logistic regression model was built for each leaf. Another concept of hybridization was proposed a few years later by Steinberg and Cardell (1998). It combined CART (classification and regression trees) algorithm (Breiman et al. 1984) with logit models. This time it was also a two-step procedure, however, the set of independent variables in the logit model was supplemented with an additional variable whose categories informed about the terminal node to which the case was assigned.

After 1998 several researchers attempted to combine decision trees with logistic regression. They developed new hybrid approaches known as LOTUS (Logistic Tree with Unbiased Selection) (Chan and Loh 2004), LMT (Logistic Model Tree) (Landwehr et al. 2005), or PLUTO (Penalized, Logistic Regression, Unbiased Splitting, Tree Operator) (Zhang and Loh 2014).

Combining clustering with decision trees for building hybrid predictive models has long been of interest to many researchers as well. Some authors combined the results obtained from clustering algorithms (k-means, k-medoid, self-organizing maps (SOM), fuzzy c-means and Balanced Iterative Reducing and Clustering using Hierar-

chies (BIRCH)) with the results obtained from the decision tree (C5.0) with boosting. Some authors proposed a hybrid model to predict churning in the area of customer relationship management, where C5.0 decision trees and Growing Hierarchical Self-Organizing Map (GHSOM) were combined.

In general, hybrid models based on decision trees and clustering algorithms are referred to as cascade models, cross-algorithm ensembles, and two-stage classification (Łapczyński and Jefmański 2013; Łapczyński 2016).

One of the hybrid methods that allow combining the advantages of a confirmatory SEM approach and exploratory classification trees is the SEM-Tree model (Brandmaier et al. 2013b).

In the literature there are many papers on the allocation of resources in households, but they do not relate to relationships with altruistic and competitive decision-making strategies and preferences. Kazianga and Wahhaj (2017) investigated the relationship between the family ties of household members and the allocation of resources within that household. On the other hand, Quanjer and Kok (2019) examined how the distribution of the family budget influences the nutrition of children and their height in adulthood. Often the subject of research refers to the sex role in the household. The authors examine how the resource allocation is determined when the wife earns more than her husband (Commuri and Gentry 2005), when the wife spends money in the absence of her husband in a family of Mexican immigrants in the US (Antman 2015), when the expenses relate to healthcare (Onah and Horton 2018). The worse situation of women in the decision-making process on the allocation of household resources is confirmed by research results from almost all over the world: from Kenya (Marinda 2006), from Russia (Lacroix and Radtchenko 2011) or from India (Fuwa et al 2006).

The aim of the article is to identify the dimensions of shaping the strategy of scarce resources allocation used by the members of Polish households and to segment them. The hybrid approach allows the separation of household segments based on the inter-relationships between the decision-making strategies (altruistic—patriarchal person vs. competitive—rivalry strategy) and preferences in terms of consumer resources that are their cultural, human and financial capital (reputation, time and money). The applicative contribution of the paper is an identification of these relationships in a heterogeneous population using exploratory and hybrid approaches linking structural equation modeling and classification and regression trees in SEM-Tree model. Because the contemporary applications of SEM-Tree approach are relatively rare in the literature, several methods of SEM-Tree model growing methods and n-fold cross-validations were used and compared.

2 SEM-Tree approach

Structural Equation Model Trees (SEM Trees) is the method combining a confirmatory approach (SEM) and the exploratory approach (recursive partitioning known from decision trees) (Brandmaier et al. 2013a; Zeileis et al. 2008). The graphic form of the model resembles the structure of the decision tree, in which the nodes correspond to estimated structural equation models. The partition of nodes in decision tree is performed on the basis of covariates, which are not used while building structural

models. At each terminal node of the tree, the SEM model is built on the basis of cases (subset of data) that belong to it.

SEM-Tree is flexible and useful tool for assessing model invariance for either categorical or continuous partitioning variables. However, SEM Tree has a tendency to overfit the sample data so pruning is needed to obtain a more generalizable solution (Brandmaier et al. 2013a).

The elementary algorithm for the induction of SEM Trees is as follows:

1. Fitting a parametric model (so-called template model) to the whole data set.
2. Binary split of the data set in all possible ways using all covariates for this purpose. In child nodes of decision tree, build structural models and compare the fit of them (so called the compound model) with the fit of template model.
3. Choosing the best compound model, the one that best describes the data from the point of view of the chosen criterion. Repeating the procedure with the first step in case the best model fits significantly better than template model. Otherwise, the ending of procedure of growing SEM Trees.

The structural model $M(\theta)$ based on all cases (built in the root of the decision tree) is called either a template model, a pre-split model, or a base model. It is created with the use of the fit index, which is usually the maximum likelihood index of fit (Furnival 1961).

Data set D is represented by a $n \times (k + l)$ matrix, where n is the number of cases, k the number of observed variables and l the number of covariates not included in the structural model. The distinction between observed variables and covariates means that matrix D is divided into respectively: submatrix D_k and submatrix D_l .

Covariates are used to split nodes, and their values or categories decide to which child node the object from data set will be assigned to. In order to include continuous (quantitative), ordinal and nominal covariates in the analysis, all multi-valued and multi-categorical variables are converted into dichotomous ones. The way they are transformed depends on the type (measurement scale).

Let N denote the number of categories or values of covariates. According to the proposed binarization procedure, continuous and ordinal covariates can be transformed in $N - 1$ ways, and nominal covariates in $2^{(N-1)} - 1$ ways. For every possible split in child nodes, structural models are built, which are called submodels. They are also called “post-split model” or “compound model”.

The pre-split model and post-split model are algebraically nested models if the parameter estimates are obtained with maximum likelihood estimation. Because they are nested, a likelihood ratio test is used to determine whether the model can be sub-modelled. The likelihood ratio is asymptotically chi-square distributed with the null hypothesis, which states that the covariate does not influence the model (the template model is not significantly different than the compound model). At each level of the tree, a covariate with the highest value of log-likelihood ratio is selected. This procedure is continued recursively in the subsequent stages while splitting the tree.

There are natural stop criteria throughout the procedure:

- there is no covariate that could split tree nodes,
- the number of cases in the leaf (terminal node) is lower than the threshold determined by the researcher,

- the desired tree depth has been achieved,
- the best split of the node is not good enough. The fourth criterion is used to prevent overfitting.

The formal evaluation of the split of the node is as follows:

1. The template model $M(\hat{\theta}) \mid D$ is characterized by the likelihood function for the parameter set $\hat{\theta}$ with m free parameters and the data set D .
2. The parameter set $\hat{\theta}_F$ is obtained for the whole data set by minimization:

$$f(\hat{\theta}_F \mid D) \quad (1)$$

3. For a given potential split, it can be assumed that data set D will be divided into j disjoint subsets where $j = 1, 2, \dots, k$. The partition is based on covariate C_i , whose values are denoted by v_{ij} .
4. Since the subsets are disjoint, so the parameters of the structural models $\hat{\theta}_{vij}$ are estimated independently by minimizing likelihood function.
5. The compound model $M(\hat{\theta}_{vij}) \mid D_{vij}$ is now referred to as M_{SUB} . The template model M is nested within M_{SUB} , because M corresponds to the M_{SUB} with n additional linear constraints on the free parameters in $\hat{\theta}$.
6. Given this nested structure, a null hypothesis can be formulated, which says that model fit M does not differ significantly from model M_{SUB} , i.e. $H_0 : \hat{\theta}_1 = \hat{\theta}_2$.
7. The log likelihood ratio between M and M_{SUB} is expressed by the formula:

$$\Lambda_i = -2LL(M(\hat{\theta}) \mid D) + \sum_{j=1}^k 2LL(M(\hat{\theta}_{vij}) \mid D_{vij}) \quad (2)$$

8. It is known that Λ is chi-square distributed with $(k - 1)m$ degrees of freedom.

Taking into account all covariates, all the possible splits of the node are evaluated. The model with the highest increase in goodness of fit is compared with the previously selected threshold determined by the significance level α . If the split is statistically significant, the tree construction procedure is continued.

One of the first applications of the SEM Trees approach in the field of psychology concerned nonsuicidal self-injury (Ammerman et al 2016). The set of variables included, among others: suicidal thoughts and behavior, symptoms of depression, problems with emotions, anxiety, symptoms of eating disorders. The covariate, which was used to separate the homogeneous subsets of the examined people, was the number of self-mutilations.

Another example of the application of the SEM Trees approach was identification of factors affecting brain health, its cognitive and mental functions at various stages of people's life (Walhovda 2018). The research was carried out as part of the "Lifebrain" project, whose vision was to enable targeted prevention of problems related to brain health. Risk factors and protective factors related to sociodemographic variables, lifestyle variables, birth and early health data, and early intellectual function. The dependent variables were, among others: memory, depression, anxiety, while moderators were: age, sex, brain structure, health factors and brain functions.

The SEM Trees model was used in the study of cognitive decline among older people (Zelinski and Jacobucci 2017). Thanks to this approach, almost 21,000 patients were grouped into subsets using such covariates as: race, education, gender and Hispanic ethnicity. Separate structural models that were built for each of the subsets, showed different patterns of cognitive ageing.

The relationship between cognitive functions of the brain and the age of adults has also been the subject of other studies where SEM Trees models were used during data analysis (de Mooij et al 2018). The research involved the grey matter and the white matter in the brain as well as cognitive tasks, which included language, memory and fluid intelligence. Using this approach, more than 600 subjects were divided into subsets (tree nodes), due to the continuous covariant, which was the age of an adult.

In marketing research SEM-Tree approach is applied to identify the structure of preferences and values among Polish consumers. According to Authors' knowledge this is the first application of SEM-Tree approach to marketing data.

3 Data characteristics

The SEM-Tree model was built on the basis of data from nationwide survey concerning the analysis of the structure of preferences and values of Polish households in the allocation of resources for consumption, saving and investing. The research was carried out in 2013 on a random sample of 1020 respondents chosen from 410 households (interviews were conducted in families with a father, mother and the oldest child over 16 years old at home) (Sagan 2014).

The variables selected for the SEM model are indicators of three constructs related to attitudes towards resource allocation and decision making strategy. The first concerns the strategy of making allocation decisions in the family with respect to the altruistic dimension (that is the patriarchal person strategy of decision-making focused on caring for the common good of the family) and rivalry (having own, independent "budgets" by family members). Models of behaviours within a household: altruistic, cooperative (Bergstrom 1996), negotiating-competitive (Mc Elroy and Horney 1981; Donni and Chiappori 2011).

A female decision maker, a large number of children, and worse financial standing should be conducive to altruistic attitude and common ends of savings and consumption.

A male decision maker, smaller number of children, and better financial standing should promote more competitive attitudes, individual ends of consumptions and savings, giving preference to the role of the decision maker. Educational factor is related to other factors (the higher education, the higher income) and it helps to predict more altruistic attitudes.

The measurement was made using the 5-point Likert scales and the following statements regarding the altruistic dimension pole (A):

A—altruistic value

- *p*23—"The family should limit spending on individual needs in order to satisfy common ones"

Table 1 Basic statistics of the latent variable indicators.

Source: Printout of the *psych* library (R package)

Indicators	N	Mean	Median	SD
p23	1020	3.42	4	0.96
p24	1020	3.76	4	0.92
p25	1020	3.41	3	0.96
p26	1020	3.58	4	1.02
p51	1020	3.24	3	1.03
p52	1020	3.09	3	1.06
p54	1020	1.87	2	0.98
p55	1020	2.00	2	1.00

- *p24*—“The common good of the whole family is more important than a family member separately”
- *p25*—“The sense of accomplishment is provided by goods intended for the whole family”
- *p26*—“The joy of life draws more in the family from the goods that serve everyone”

The other two latent variables concerned compromises (trade offs) resulting from the relationship between reputation and income as well as leisure time and the pursuit of higher income and material security (the dilemma “reputation” and “money” and “time” and “money”). The most important asset of family members are: money (financial capital), leisure-time (human capital) and reputation (cultural capital). The research question therefore was what are the relationships between the strategies of income management and money (M)—leisure time (T) and money (M)—reputation (R) trade offs?

On the basis of research question the main hypothesis was formulated: the higher tendency to take patriarchal person strategy, the larger T–M and R–M trade off due to risky shift effect. The risky shift effect as a form of group polarization phenomenon, explains the tendency of group as a whole (or group leader) to reveal a more extreme course of action in comparison to the average of the individual judgments of group members (Myers and Lamm 1976; Shaw 1976).

The specific latent variables and its indicators (statements with 5-point scales Likert) are given below:

Y1—“time” and “money” trade-off:

- *p51*—“Earning and spending money is more important than free time”
- *p52*—“Earning a safe tomorrow is more important than free time”

Y2—“reputation” and “money” trade-off:

- *p54*—“The end justifies the means - money is more important than own reputation”
- *p55*—“You can lose your reputation as a safe financial future”

Basic statistics for the indicators are given in Table 1. Covariance matrix of the indicators are presented in Table 2.

Additional accompanying variables that participated in the construction of the SEM-Tree tree (predictors) were:

Table 2 Covariance matrix of the latent variable indicators. *Source:* Printout of the *stats* library (R package)

	p23	p24	p25	p26	p51	p52	p54	p55
p23	0.922							
p24	0.255	0.841						
p25	0.234	0.355	0.923					
p26	0.172	0.248	0.380	1.049				
p51	0.104	0.086	0.162	0.105	1.058			
p52	0.096	0.115	0.153	0.070	0.459	1.121		
p54	0.101	−0.122	0.098	−0.022	0.161	0.244	0.964	
p55	0.051	0.007	0.029	−0.045	0.126	0.164	0.658	0.994

- $m1$ —sex: male, female
- $m2$ —age: 18–24, 25–34, 35–49, 50–64 and 65+
- $m3$ —education: primary education, lower secondary, vocational education, secondary, higher education (1st level) and higher education (2nd level)
- $m5$ —subjective financial situation: very good, good, average, bad and very bad

4 Results

4.1 Template SEM model

The basic template SEM model is depicted in Fig. 1.

The fit of the model is acceptable. Model fit test Chi-Square statistic is 51.533 with 17 df and p -value = 0.000 (Chi-square/df = 3.03). The incremental fit indices Comparative Fit Index (CFI) = 0.975, Tucker-Lewis Index (TLI) = 0.960. Root Mean Square Error of Approximation (RMSEA) = 0.045 (90 percent confidence interval is between 0.031 and 0.059, with p -value RMSEA equal or less 0.05 is 0.716). Standardized Root Mean Square Residual (SRMR) = 0.031. Factor loadings, path coefficients and error variances of the template model are presented in Table 3.

The Cronbach's alpha (CA) and McDonald's Omega reliability coefficients of scales are acceptable. For Y1 scale, CA = 0.59 and Omega = 0.60. For Y2 scale CA = 0.80 and Omega = 0.83. For A scale CA = 0.62 and Omega = 0.63. The reliability of total model is 0.61 and 0.75 respectively. Average variance extracted (Fornell-Lacker AVE) for Y1 scale = 0.43, for Y2 scale = 0.72, for A = 0.31 and for total model = 0.45.

The parameters of structural part of the model are significant and shows that the higher altruistic attitude towards income allocation, the higher time-money trade off ($A-Y1 = 0.44$) and reputation - money trade off ($A-Y2 = 0.22$). It confirms the main research hypothesis concerning the risky shift effect. Also, the higher time-money trade off, the higher reputation—money trade off ($Y1-Y2 = 0.18$). Decomposition of mediation effect indicates that indirect effect ($ab = 0.039$, $p = 0.04$) significantly explains part of the relationships between altruistic attitude and reputation—money trade off.

Fig. 1 Template SEM model. A: altruistic value, Y₁: time-money trade off, Y₂: reputation-money trade off. Source: printout of the *lavaan* library (R package)

4.2 SEM-Tree model

The disadvantage of SEM models built in the leaves of the classification tree is the need to take into account the predefined dependent variable in the tree construction process.

In the SEM-Tree algorithm presented at the beginning, the splits cover all combinations of the binarized predictors without the need to consider the dependent variable. In the model building process several models, based on splitting methods and stop rules, are developed. The *semtree* package offers several tree growing methods:

1. “naive” splitting takes the best split value of all possible splits on each covariate.
2. “fair” selection tests all splits on half of the data, then tests the best split value for each covariate on the other half of the data.
3. “fair3” has an additional step of retesting all of the split values on the best covariate found in the second phase.
4. “crossvalidation” partitions the data for maximizing splits on each variable, then comparing maximum splits across each variable on the rest of the data

Additional criterion was the size of final leaf (25, 50, 100, 200 and 400). In sum, 20 models are developed and compared.

Model comparisons are based on deviance ($-2LL$ of leaf (terminal) nodes). Figure 2 presents the model comparisons. It should be noted that for $N > 25$ all method present

Table 3 Parameters of the template path model. *Source:* Printout of the *lavaan* library (R package)

	Estimate	Std. Err	Z-value	P-level
Loadings				
Y1–p51	1.000			
Y1–p52 (11)	1.260	0.173	7.280	0.000
Y2–p54	1.000			
Y2–p55 (12)	0.693	0.097	7.148	0.000
A–p23	1.000			
A–p24 (13)	1.379	0.151	9.106	0.000
A–p25 (14)	1.822	0.198	9.207	0.000
A–p26 (15)	1.353	0.155	8.699	0.000
Regression				
Y1–A	0.444	0.094	4.739	0.000
Y2–A	0.221	0.100	2.204	0.028
Y1–Y2	0.178	0.038	4.716	0.000
Variances				
p51	0.693	0.058	12.040	0.000
p52	0.542	0.081	6.700	0.000
p54	0.014	0.129	0.109	0.914
p55	0.538	0.066	8.122	0.000
p23	0.778	0.038	20.536	0.000
p24	0.566	0.034	16.707	0.000
p25	0.445	0.042	10.521	0.000
p26	0.784	0.042	18.702	0.000
Y1	0.301	0.049	6.102	0.000
Y2	0.942	0.136	6.947	0.000
A	0.144	0.027	5.333	0.000

approximate similar results. The most consistent results appear at $N=50$ (in general, the higher number of respondents in the final leaf, the higher deviance).

Additionally, several models based on n -fold cross validation and $N=50$ were developed. Figure 3 presents the results of that simulation in the range of 3-fold up to 15-fold cross validation. The simulation shows that the 7-fold cross validation performs the best. In the Authors' opinion, too many folds in cross validation procedure make subsets to test the model too small. As a consequence, the variance is high and the final model error is also high.

Figure 4 shows the final hybrid SEM-Tree model calculated using the *semtree* library of the R program. The model was created using the CV method with 7-fold cross validation.

The model shows that the most important predictors turned out to be: age, financial situation and gender. The structure of the segments are as follows:

Segment 1 refers to the young (18–34). The patriarchal person strategy is characterized by stronger relationships with R-M and weaker T-M trade off. People in this

Fig. 2 Comparison of SEM-Tree growing methods. CV: crossvalidation partitions for maximizing splits on each variable, fair: partitions based on all splits on half of the data, fair 3: partitions based on retesting all of the split values on the best covariate, naive: best split value of all possible splits

Fig. 3 SEM-Tree model cross-validation comparison

Fig. 4 SEM-Tree model of resource preferences. m1: sex (1—male, 2—female), m2: age (1—18–24, 2—25–34, 3—35–49, 4—50–64, 5—65+), m5: subjective financial situation (1—very good, 2—good, 3—average, 4—bad, 5—very bad), I1: Y1-p52 loading, I2: Y2-p55 loading, I3: A-p24 loading, I4: A-p25 loading, I5: A-p26 loading. *Source:* printout of the *semTree* library (R package)

segment possess relatively weak cultural capital, and they are able to sacrifice own reputation in order to reach a higher financial position. This segment consists of 439 respondents.

Segment 2 refers to mature (above 35) rich males. The patriarchal person strategy is characterized by strong relationships with both R–M and T–M trade off. People in this segment possess relatively strong financial but weak cultural and human capital, and they are able to sacrifice both own reputation and free time in order to reach a higher financial position. This segment consists of 130 respondents. The findings confirm the main hypothesis but it does not prove the patriarchal role of women.

Segment 3 refers to mature rich (very good and good financial situation) females. The patriarchal person strategy is characterized by very strong relationships with T–M trade off and negative R–M trade off. People in this segment possess relatively strong

financial and cultural capital but weak human capital, and they are able to sacrifice only free time in order to reach a higher financial position, keeping own reputation intact. This segment consists of 137 respondents. Within this segment the main hypothesis is confirmed and the results indicate the patriarchal role of women.

Segment 4 refers to the mature poor (average, bad and very bad financial situation). The patriarchal person strategy is characterized by very strong relationships with T–M trade off and insignificant R–M trade off. People in this segment possess relatively weak financial and cultural but strong human capital, and they are able to sacrifice only free time in order to reach a higher financial position, keeping own reputation intact. This segment consists of 314 respondents. The results enable to accept the research hypothesis and confirms the role of low income (regardless of gender).

The variable “education” did not appear in the tree, because other predictors reduced the variances in child nodes to a higher degree. Interestingly, all segments relate to the patriarchal person strategy. The patriarchal system means that men play the dominant role in the family and in society. This does not mean a lower status of women, but it is related to the fact that the man represents the family, he is to ensure its safety and financial resources. According to the research describing the social roles of men and women (Hofstede et al. 2010), society is called masculine when men have challenging work, have opportunities for high earnings, are assertive and focus on the material situation. An analysis of the masculinity index (MAS) values in European countries shows that in Poland it is quite high (64 points). This level is similar to the one in Germany and Great Britain (66 points), but quite distant from the one in Slovakia (110 points) or Sweden (5 points) and Norway (8 points).

5 Conclusions

The SEM-Tree hybrid models allow exploratory analysis of path relationships in a heterogeneous population. In contrast to classical decision trees with SEM models in leaves, they enable the extraction of predictors associated with the variation not only of the level of the dependent variable, but also of the path relations between variables. The model comparisons shows that 7-fold cross validation gives the best model performance.

Both types of models confirm the risk-shift hypothesis about the influence of decision-making strategies (altruistic) on the devotion of free time and reputation for obtaining higher income in the household.

Summing up the results of the model, it should be emphasized that the choice of an altruistic (patriarchal person) strategy (taking care of the “common good” with a strong control function of the family head) strongly encourages the members of the household to submit the value of money (“struggle for existence”), even at the expense of free time and loss of one’s reputation. Therefore, it can be assumed that the decentralization of budgets, greater autonomy and thus the sense of security of family members tend to reveal a higher preference for free time, taking care of their own reputation and thus a more harmonious individual life.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Antman FM (2015) Gender discrimination in the allocation of migrant household resources. *J Popul Econ* 28(3):565–592
- Ammerman BA et al (2016) Development and validation of empirically derived frequency criteria for NSSI disorder using exploratory data mining. *Psychol Assess*. <https://doi.org/10.1037/pas0000334>
- Bergstrom TC (1996) Economics in a family way. *J Econ Lit* 34(4):1903–1934
- Brandmaier AM, Oertzen T, McArdle JJ, Lindenberg U (2013) Structural equation model trees. *Psychol Methods* 18(1):71–86
- Brandmaier AM, Oertzen T, McArdle JJ, Lindenberg U (2013) Exploratory data mining with structural equation model trees. In: McArdle JJ, Ritschard G (eds) *Contemporary issues in exploratory data mining in the behavioral sciences*. Routledge, New York, pp 96–127
- Breiman L, Friedman J, Stone CJ, Olshen RA (1984) *Classification and regression trees*. Wadsworth, New York
- Chan K-Y, Loh W-Y (2004) LOTUS: an algorithm for building accurate and comprehensible logistic regression trees. *J Comput Graph Stat* 13:826–852
- Commuri S, Gentry JW (2005) Resource allocation in households with women as chief wage earners. *J Consum Res* 32:185–195
- de Mooij SMM et al (2018) Age differentiation within gray matter, white matter, and between memory and white matter in an adult life span cohort. *J Neurosci* 38(25):5826–5836
- Donni O, Chiappori A (2011) Nonunitary models of household behavior: a survey of the literature. In: Molina JA (ed) *Household economic behaviors*. Springer, London
- Furnival GM (1961) An index for comparing equations used in constructing volume tables. *For Sci* 7(4):337–341
- Fuwa N et al (2006) Gender discrimination, intrahousehold resource allocation, and importance of spouses' fathers: evidence on household expenditure from rural India. *Dev Econ* 44(4):398–439
- Hofstede G, Hofstede GJ, Minkov M (2010) *Cultures and organizations. Intercultural cooperation and its importance for survival*. McGraw-Hill, New York
- Kazianga H, Wahhaj Z (2017) Intra-household resource allocation and familial ties. *J Dev Econ* 127:109–132
- Lacroix G, Radtchenko N (2011) The changing intra-household resource allocation in Russia. *J Popul Econ* 24:85–106
- Landwehr N, Hall M, Frank E (2005) Logistic model trees. *Mach Learn* 59(1–2):161–205
- Łapczyński M (2016) Modele hybrydowe w marketingu relacji. Cracow University of Economics, Cracow
- Łapczyński M, Jefmański B (2013) Impact of cluster validity measures on performance of hybrid models based on k-means and decision trees. In: Perner P (ed) *Adv data min. ibai-publishing, Fockendorf*, pp 153–162
- Lindahl WE, Winship C (1994) A logit model with interactions for predicting major gift donors. *Res High Educ* 35(6):729–743
- Marcoulides GA, Ing M (2012) Automated structural equation modeling strategies. In: Hoyle RH (ed) *Handbook of structural equation modeling*. Guilford Press, New York, pp 690–704
- Marinda PA (2006) Child-mother nutrition and health status in rural Kenya: the role of intra-household resource allocation and education. *Int J Consum Stud* 30(4):327–336
- Mc Elroy M, Horney MJ (1981) Nash-bargained household decisions: toward a generalization of the theory of demand. *Int Econ Rev* 22:333–349
- Myers DG, Lamm H (1976) The group polarization phenomenon. *Psychol Bull* 83:602–627

- Onah MN, Horton S (2018) Male-female differences in households' resource allocation and decision to seek healthcare in south-eastern Nigeria: results from a mixed methods study. *Soc Sci Med* 204:84–91
- Quanjor B, Kok J (2019) Homemakers and heights. Intra-household resource allocation and male stature in the Netherlands 1860–1930. *Econ Hum Biol* 34:194–207
- Sagan A (ed) (2014) Values and preferences in income allocation of polish households. Cracow University of Economics, Cracow
- Sagan A, Perek-Białas J (2016) Eksploracyjne podejścia w budowie modeli strukturalnych. In: Grześkowiak A, Mazurek-Lopacińska K, Sobocińska M, Stanimir A (eds) *Metody badań marketingowych: modelowanie, technologia, wizualizacja*. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, pp 38–51
- Shaw ME (1976) *Group dynamics: the psychology of small group behaviour*. McGraw Hill, New York
- Steinberg D, Cardell NS (1998) The hybrid cart-logit model in classification and data mining. In: Eighth annual advanced research techniques forum. American Marketing Association, Salford Systems, pp 1–7
- Walhovda KB et al (2018) Healthy minds 0–100 years: optimising the use of European brain imaging cohorts ("Lifebrain"). *Eur Psychiatry* 15:47–56
- Zeileis A, Hothorn T, Hornik K (2008) Model-based recursive partitioning. *J Comput Graph Stat* 17(2):492–514
- Zelinski EM, Jacobucci R (2017) Heterogeneity in cognitive change trajectories of older adults observed from structural equation model trees. *Alzheimer Dement* 13(7):584–585
- Zhang W, Loh W-Y (2014) PLUTO: penalized unbiased logistic regression trees. [arXiv:1411.6948v1](https://arxiv.org/abs/1411.6948v1). Accessed 12 Feb 2017

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.