Akademie věd České republiky
Ústav teorie informace a automatizace

Czech Academy of Sciences
Institute of Information Theory and Automation

RESEARCH REPORT

Jan ŠVORC, Jiří VOMLEL

Subjective well-being and the individual material situation in Central Europe: A Bayesian network approach.

No. 2387 December 2020

ÚTIA AV čR, P. O. Box 18, 182 08 Prague, Czech Republic
Fax: (+420)(2)688 4903
E-mail: utia@utia.cas.cz

SUBJECTIVE WELL-BEING AND INDIVIDUAL MATERIAL SITUATION IN CENTRAL EUROPE: A BAYESIAN NETWORK APPROACH

Abstract:

The objective of this paper is to explore the associations between the subjective well-being (SWB) and the subjective and objective measures of the individual material situation in the four post-communist countries of Central Europe (the Czech Republic; Hungary; Poland; and Slovakia). The material situation is measured by income; relative income compared to others; relative income compared to one's own past; perceived economic strain; financial problems; material deprivation; and housing problems. Our analysis is based on empirical data from the third wave of European Quality of Life Study conducted in 2011. Bayesian networks as a graphical representation of the relations between SWB and the material situation have been constructed in five versions. The models have been assessed using the Bayesian Information Criterion (BIC) and SWB prediction accuracy, and compared with Ordinal Logistic Regression (OLR). Expert knowledge, as well as three different algorithms (greedy, Gobnilp, and Tree-augmented Naïve Bayes) were used for learning the network structures. Network parameters were learned using the EM algorithm. Parameters based on OLR were learned for a version of the expert model. The Gobnilp model, the Markov equivalent to the greedy model, is BIC optimal. The OLR predicts SWB slightly better than the other models. We conclude that the objective material conditions' influence on SWB is rather indirect, through the subjective situational assessment of various aspects related to the individual material conditions.

Keywords:

Subjective Well-Being, Income, Economic Strain, Material Deprivation, Bayesian Networks, Central Europe

JEL Codes:

C11, I31

Acknowledgements:

This research was supported by the Czech Science Foundation (Project Nos. 17-08182S and 19-06569S).

1 Introduction

Various commonly used statistical methods are suitable for revealing the relationships among subjective well-being (hereinafter referred to as "SWB") and material living conditions[1] while simultaneously controlling the other variables. Yet the explanatory variables may be interrelated and their relationships with the dependent variable may not necessarily be direct. It can be intermediated by a set of other variables. To overcome this drawback we employ Bayesian networks as a modeling tool to graphically represent the interrelations between SWB and the variables of the material situation in this paper. The Bayesian networks help us explain and visualize the situation in its complexity and to understand the relationships of conditional independence between SWB and the other variables. Based on this representation, the interplay among the factors is clearly visible.

The research presented in this paper was undertaken as a part of the unpublished doctoral dissertation defended in 2019 (Švorc 2019).

2 Literature review

The research on SWB and its relationship with material circumstances were reviewed and examined in the author's earlier paper (Švorc 2018). In that paper, the pairwise associations between SWB and seven measures of the material situation[2] was confirmed for Central Europe as a whole[3]. In the new millennium, happiness research has been enriched with knowledge from poverty research; of particular interest is the concept of material deprivation.

Layte et al. (2001: 44) defined it as *"the enforced lack of possessions, activities or amenities through lack of resources".* For Ahrendt et al. (2015: 633) it *"concerns the inability to afford items that are considered essential"*. In the series of articles, Whelan and his colleagues talk about lifestyle deprivation (Whelan 1992; Whelan et al. 2001; Layte et al. 2001; Whelan and Maître 2007). Whelan (1992) distinguished primary (enforced absence of items or activities widely regarded as necessities); secondary (possessions and activities not socially sanctioned as necessities); and housing deprivation. Whelan et al. (2001) examined 24 household items which could serve as indicators of lifestyle deprivation and clustered them into five groups that are consistent across different countries: basic life-style deprivation

[1] We use the terms "material living conditions; "wealth"; "material conditions"; "material well-being"; "material circumstances"; etc. synonymously.

[2] Income; financial situation as compared to others and to own past; economic strain; material deprivation; financial problems; and housing defects.

[3] Few relationships remained unconfirmed in the cases of Slovakia and Hungary.

(essential items); secondary lifestyle deprivation (less essential items); housing facilities; housing deterioration; and environmental problems.

The relationship between material deprivation and SWB was investigated, for example, by Bellani and D'Ambrosio (2011). They reported that the association between SWB and deprivation is much stronger than between SWB and the individual's income[4]. Ervasti and Venetoklis (2010) empirically analyzed SWB and its determinants among the unemployed based on the European Social Survey (ESS) data from 21 countries and stressed the importance of the financial strain for the well-being of the unemployed. Based on the authors' claim, the deprivation theory tends to put the main emphasis on the psychological factors and ignores the financial strain aspect.

Whelan et al. (2001: 358) provide an overview of studies that found *"a substantial proportion of those on low incomes not to be suffering from deprivation while some households above income poverty lines do experience such deprivation"*. It hence seems that those having a low income and those scoring high in terms of material deprivation are not necessarily the same people (Berthoud and Bryan 2010). The authors found that the underlying link between long-term low income and long-term deprivation is close, but an increase in income is not necessarily accompanied with a proportionate decrease in deprivation. The authors concluded that the correlation between deprivation and income exists, but its strength is not that great.

Fahey (2007) raised the question of who is poor in the EU. He clustered 25 member states and three candidate states into four groups and found that, in each cluster, the lower the income the higher the proportion of deprived people. The same pattern was identified in the cases of the other two indicators: housing defects and financial problems. Those more deprived in objective terms are also the people who feel more deprived. Layte et al. (2001) focused on the relationship between income and lifestyle deprivation on an international scale. They suggested that this relationship varies across countries due to different welfare-state policies. Some welfare states allow the process the authors call "decommodification", which enables *"smooth income flows"* (Layte et al. 2001: 43). Basically, the social programs economically support people who have lost their income due to, for example, unemployment and this support, in turn, moderates the deprivation.

[4] The correlation between SWB and deprivation was 0.4 and the correlation between SWB and income was about 0.16 on average among nine EU countries (former EU15 without Austria, Finland, Luxembourg, Sweden, Germany, and the UK).

Whelan et al. (2001) identified the fact that the relationship between income and material deprivation is weak due to social and economic processes, such as existence of other resources; the variety of views of what is necessary; length of the period of lost income; etc. There was a weak relationship between income and housing, as well as environmental dimensions, whereas basic and secondary dimensions were impacted more strongly by income than were the other three dimensions. These conclusions can be generalized across the examined EU countries, but the relationship of income and lifestyle deprivation was weaker in the richer countries. The study also established that both income and deprivation have effects on the ability to make ends meet whereas the effect of the deprivation is stronger. Later, Whelan and Maître (2007) confirmed the relatively weak relationship between income and lifestyle deprivation and found that the effect of income on deprivation is stronger in the poorer EU regions. On the other hand, the strongest association between deprivation and perceived economic strain prevails in the richest EU countries. A weaker relationship was found in all other EU regions.

3 Bayesian networks

Bayesian networks were first introduced by Judea Pearl in the early 1980s (Pearl 1982). A Bayesian network (hereinafter referred also as "BN") is a compact graphical representation of a joint probability distribution over the variables of interest. It consists of a set of variables and a set of directed edges between the variables, where each of the variables has a finite set of mutually exclusive states (Jensen 2001). The variables together with the directed edges form a directed acyclic graph, often abbreviated as "DAG". The graphs are acyclic as they must not have any feedback loops[5]. The graphs are directed as the edges between the nodes are oriented. For each variable in the graph, a conditional probability distribution given the states of the parents of that variable is defined[6].

The directed edges between the variables need not represent the "cause-and-effect" relationships. They only define the relationships of the conditional dependencies and independencies among the variables. In other words, the structure induces d-separation properties of the network (Jensen 2001). Since we utilize the concept of d-separation as a criterion during the expert model construction, a brief introduction is useful. If there are two distinct variables A and B and an intermediate variable V distinct from A and B in a BN, the three possible connections between the nodes A and B through V can be drawn: serial from A to B

[5] Formally, a directed graph is acyclic if there is no directed path A_1 ->…->A_n so that $A_1 = A_n$.

[6] To each variable A with parents $B_1,…,B_n$ there is attached the potential table $P(A|B_1,…,B_n)$.

(or B to A) through V; diverging from V to A and to B; and converging from A and B to V. Two distinct variables A and B in the BN are d-separated given the intermediate variable V distinct from A and B if either the connection between A and B is serial or diverging and V is instantiated[7], or the connection between A and B is converging[8] and neither V nor any of V's descendants are instantiated. This definition is adapted from Jensen (2001). Basically, if the state of the variable V is known, it blocks the transmission of evidence[9] between A and B in the cases of serial and diverging connections. The situation of a converging connection is the only different one. In this situation, the communication is blocked in the case of the state of V (or one of its descendants) being unknown. A and B are d-separated if *"changes in the certainty of A have no impact on the certainty of B."* (Jensen 2001: 11). The definition formulated above could be generalized to more than one variable V.

The BN construction includes two main consecutive phases: determining the structure and learning the parameters[10]. Determining the structure includes definition of model variables and establishing directed links among the variables in a network. The structure can be determined based on the expert knowledge or learned from the available data using a machine learning algorithm. Learning the parameters means estimating the conditional probability distributions based on the training data. Again, the parameters can be either set by engineering considerations or estimated from the data.

From our perspective, the main advantages of the application of BNs in contemporary social research are the following.

1. They enable an efficient encoding of the domain knowledge and improve our understanding of complex problems.
2. They provide a compact representation of the joint probability distribution[11].
3. They enable exact probabilistic inference assuming the structure and the parameters are estimated correctly. Under these assumptions, the posterior probability distribution of any variable can be calculated.

[7] If the state of a variable is known, we say that this variable is instantiated (Jensen 2001).

[8] The converging connection is often referred also as a "head-to-head" connection or a v-structure. The diverging connection is also referred as a "tail-to-tail" connection; and the serial connection can also be referred to as a "head-to-tail" connection.

[9] Information that a given variable takes specific values.

[10] The parameters of a BN are the conditional probabilities that define the joint probability distribution of the network as their product.

[11] The structure defines the parents for every variable. Assuming binary random variables, a standard representation of the joint probability distribution has 2^n parameters whereas the joint probability distribution represented by a BN has at most $n2^k$ parameters (n = number of variables and k = maximum number of parents).

4. They help us answer queries under uncertainty. It is a method for decision-making applicable also in situations when either an insufficient amount of information is available or the information is vague.
5. As the software for learning the structure and parameters of BNs is available, complex situations can be modeled with the help of BNs.

On the other hand, the disadvantages of BNs are the following.

1. A large amount of data is needed to reliably learn the parameters in some cases.
2. The tendency to a cause-and-effect misinterpretation of the oriented edges in a graph.

So far, the application of BNs in contemporary happiness research has not been frequent. However, this paper is not the first attempt of this innovative approach. Recently, BNs have been used as an instrument for the analysis of multidimensional well-being in a study published by Ceriani and Gigliarano (2016), where sixteen variables were defined in ten dimensions of well-being together with two more control variables. A hillclimbing algorithm with K2, AIC and BIC scores is used for structure learning, whereas a set of edges was excluded and another set was forced to be included in the structure based on the expert knowledge. The accuracy of the networks was tested using BIC and the prediction accuracy of the four selected variables resulted from a cross-validation procedure in the study.

4 Data and measurement

The empirical data from the third survey of the European Quality of Life Study (hereinafter referred as EQLS), conducted in 2011,[12] is used for learning the structures and parameters of the models (Eurofound 2014). The survey covers all persons aged 18-plus whose usual place of residence (where the respondent normally sleeps) is in the territory of the surveyed countries at the time of the data collection. Only one interview per household was conducted, in which only an adult household member was eligible to answer questions.

The data of four post-communist central European countries (the Czech Republic, Hungary, Poland and Slovakia) are used for learning BNs. The total sample size is 5,298 respondents for all four countries, out of whom 3,259 (613 in the Czech Republic; 586 in Hungary; 1,428 in Poland; and 632 in Slovakia) complete data vectors are extracted by removing respondents having answered the relevant questions incompletely[13].

[12] The question on relative income compared to most people in the country was missing in the last survey made in 2016. We consider this variable too important to be omitted from the models, thus the data from the third survey is used.

[13] Omission of the respondents not having answered the relevant questions may not be a problem in the case of values missing at random (MAR). MAR is impossible to be verified statistically and is assumed for further analysis.

The data was collected from September to December 2011 in the four countries of Central Europe. The individuals for that survey were selected strictly randomly, so that all members of the survey population had a non-zero probability of being included in the sample. Because sufficient quality sampling frames (registers) covering at least 95% of persons were available in the Czech Republic, Hungary and Poland, a random probability sampling was made based on the registers. An enumerated random route was used for the selection process in Slovakia. The sample was stratified based on the regions according to NUTS2 or statutory districts and the degree of urbanization (metropolitan, urban and rural), and allocated proportionately to the population of the selected strata in all countries.

A stratified three-stage probability sample design was adopted in the Czech Republic, Poland and Slovakia. The three stages were comprised of a random selection of Primary Sampling Units (PSUs) in each stratum, random selection of addresses (households) from the available registers (or by the random route in case of Slovakia), and random selection of the respondents living in the households. The two-stage sample design was adopted in Hungary, because the Hungarian local register is name-based and not household based.

Given the sample size and the sampling procedures, the third wave EQLS samples can be considered representative of the total statistical population to have been covered in each surveyed country. EQLS Sampling report (Eurofound 2012a) and EQLS Technical report (Eurofound 2012b) are available for details.

Appendix 1 provides an overview of the variables together with the related EQLS questions and scales. There are two questions on two dimensions of SWB (happiness and life satisfaction) in the third EQLS survey. The sample of 5,262 respondents having stated both happiness and life-satisfaction is divided into quartiles based on the average score computed from the two ten-point scales. The internal consistency of the two measures of SWB has been checked using Cronbach's alpha (Cronbach 1951). The value of alpha is 0.774. This outcome is generally considered acceptable in social research (Nunnally 1978; Schmitt 1996). The two scales of happiness and life satisfaction can hence be considered internally consistent. Thus, the variable of SWB has four ordinal states based on quartiles and this node is abbreviated as "SWB" in our BNs.

Seven variables of the individual material situation are employed in the model. Three of them are based on a person's subjective perception (subjective variables): relative income compared to the past; relative income compared to others; and perceived economic strain. The other four variables are based on objective conditions (objective variables): income; material deprivation; financial problems; and housing problems.

The variable of income relative to one's own past has three ordinal states and the node is abbreviated as “PAST” in our BNs. The scale is not transformed in any way. In the case of the relative income being compared to others, the 5-point scale is transformed into a 3-point scale by merging the groups of “much worse” and “somewhat worse” into the single group “worse” and the groups “much better” and “somewhat better” into the single group “better”. The variable has three ordinal states and is abbreviated as “OTHR” in our BNs.

There are reasons for the reduction of the numbers of states of random variables. Firstly, the insufficient sample sizes in some of the groups were identified. Secondly, if the number of the states is lower, we are more likely to have a sufficient number of data vectors for learning conditional probabilities for all configurations of nodes in a network. This is particularly important if there are more complex configurations of the parents of a node. It is more likely that the respondents are not available (or just a few respondents are available) for learning all the parameters given the parent configuration in the case of more states being defined. The more data vectors available for a particular configuration in the dataset, the more reliable the estimation of the parameters is. The reduction of the states is hence a way to increase the reliability of parameter estimation. Thirdly, a distribution of five or more states is unnecessarily detailed for the purpose of modeling with the aid of BNs.

The 6-point scale of the variable of the perceived economic strain is transformed to a 4-point scale by merging the categories of those able to make ends meet very easily and easily into the single group “easily” and those able to make ends meet with difficulty and with great difficulty into the single group “with difficulty”. The variable of the perceived economic strain has, therefore, four ordinal states and the node is abbreviated as “STRS” in our BNs.

The respondents were asked to estimate their household’s total net income per month. The household income is equivalized and recalculated in EUR based on the purchasing power parity. The question on income has quite a high percentage of non-response (only 3,965 respondents out of the 5,298 declared their income). It is a continuous variable by its nature, but for purposes of the network construction it is transformed to have four ordinal states based on quartiles calculated from the sample of the 3,965 respondents who stated their income (all respondents stated both their levels of happiness and satisfaction)[14]. This node is abbreviated as "INC" in our BNs.

The EQLS question on material deprivation has six sub-questions. The resulting six binary variables (able to afford/unable to afford[15]) are transformed into a single ternary ordinal variable. The first group consists of the respondents who cannot afford 0-1 wanted items (the group of those not deprived). The respondents who cannot afford 2-4 wanted items (weakly deprived) fall into the second group; and the last group includes the respondents who cannot afford 5-6 wanted items (strongly deprived). Respondents who refused to answer any single sub-question are excluded from the analysis. The node is abbreviated as "DEPR" in our BNs.

The EQLS question on the financial problems has four sub-questions. They are transformed into the single ternary ordinal variable. The first is the group of respondents able to pay for all four items as scheduled (the group with no financial problems). The second is the group of the respondents unable to pay timely only for a single item out of the four listed ones (the group with minor financial problems). The third group consists of the respondents who are unable to timely pay for two or more items out of the four listed ones (the group with major financial problems). The respondents who refused to answer any single sub-question are excluded. The node is abbreviated as "FPRO" in our BNs.

The EQLS question on problems with housing has six sub-questions on six types of common housing deficiencies. The six binary variables are transformed into a single ordinal ternary variable. The first group includes those having no defects in housing; the respondents in the second group have only a single defect in housing; and those having two and more defects in housing fall into the last group. The respondents who refused to answer any single sub-question are excluded from the analysis. The node is abbreviated as "HOUS" in our BNs.

The last variable in our BNs represents the respondent's country. It has four states according to the examined countries. The node is abbreviated as "CRY" in our BNs.

5 Models

Five BNs are learned using different learning algorithms based on 3,259 complete data vectors. All model variables are observed and no hidden variables

[14] The income quartiles are the same across all four examined countries; they are not constructed separately for each country. The income is equivalized in order to take the household composition into consideration.

[15] There is neither the option to answer "We have the item" nor the option "We do not want/need this item at the moment". This may cause some confusion to the respondent, still, the question is, in essence, correct.

are considered. The models are summarized in Table 1. As a working progress for this article, the use of Bayesian networks was illustrated and compared to other methods in 22nd Czech-Japan Seminar on Data Analysis and Decision Making (Švorc and Vomlel 2019), where these models were briefly outlined.

Table 1: Bayesian networks.

	Model	**Structure learning algorithm**	**Scoring criteria**	**Parameters learning Algorithm**
1	Expert/EM	Expert	-	EM
2	Expert/OLR	Expert	-	Ordinal logistic regression
3	Greedy/EM	Greedy search and score	BIC	EM
4	Gobnilp/EM	Gobnilp	BIC	EM
5	TAN/EM	Tree augmented Naive Bayes	-	EM

The structure of the two expert models is established following the scheme of a PC algorithm (Spirtes et al. 2000). Instead of the data, expert knowledge is used to decide whether a conditional independence statement holds. The EM algorithm (Dempster et al. 1977) is used for learning the parameters of the Expert/EM model, whereas the parameters of the Expert/OLR model are learned based on the ordinal logistic regression model with SWB as the dependent variable[16]. The construction of the alternative expert model with the parameters based on ordinal logistic regression is reasoned by the fact that the EM parameters are not necessarily monotone, whereas the OLR parameters, by their definition, are. The parameters are learned by the analytical software Hugin (Hugin 2014) in all instances in which the EM algorithm is used[17]. Hugin is also employed for learning the structures of the Greedy/EM and TAN/EM models. The structure of the Gobnilp/EM model is established with the Gobnilp learning algorithm embedded in the Gobnilp software (Cussens and Bartlett 2015). This algorithm promises to learn the structure optimally in terms of BIC as the scoring criterion. The sixth model is the ordinal logistic regression model with SWB as the dependent variable, and all other variables included as explanatory variables. The calculation was carried out using R software[18] (R Core team 2014).

[16] Only the parameters of the income are learned using the EM algorithm because the country is not the ordinal variable and it is the only parent of the income in the expert model.

[17] As our dataset is complete (there are no missing values), the EM algorithm simply computes the relative frequencies from the data.

[18] The command „polr“ from the MASS package was used for that calculation.

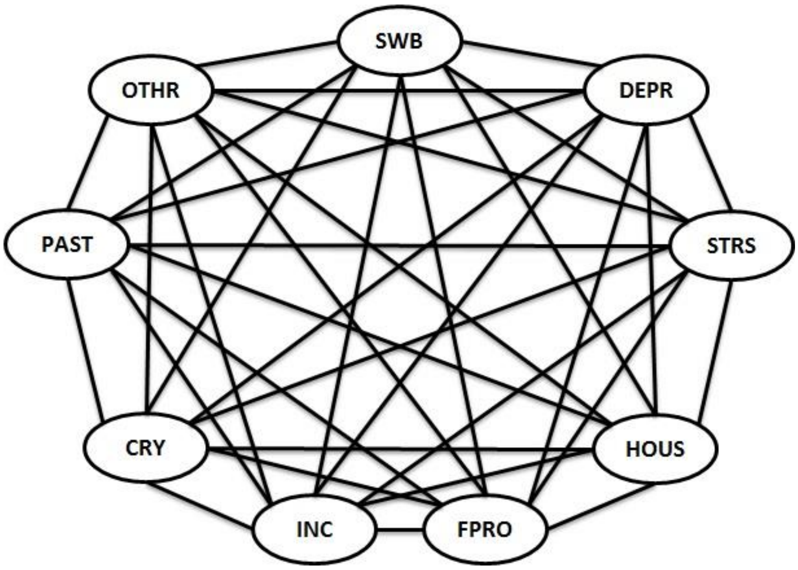
5.1 Expert model construction

We use the scheme of the PC algorithm in order to be systematic in building the structure of the expert model. The PC algorithm is a constraint-based learning method; based on the conditional independence tests, it establishes the independence relations as the constraints that should satisfy the d-separation criteria in a final graph[19]. The algorithm works in three basic consecutive stages: (1) determining the skeleton; (2) determining the v-structures; and (3) directing the remaining edges as far as possible.

Two sources of expert knowledge are considered. First is the knowledge gained from the review of the relevant literature. Unfortunately, the literature usually examines correlates and predictors of SWB and other factors, but it rarely provides us with a detailed analysis of the conditional independencies. Secondly, the knowledge based on the empirical individual real-life experience is applied. The use of this type of knowledge means that the hypotheses are formulated based on the common sense of an informed subject rather than on the exact scientific evidence. This second type of knowledge is used when sufficient research-based evidence is not available.

The first stage starts with the initial complete graph where all variables are pairwise linked by the undirected edges (Figure 1).

Figure 1: Complete undirected graph.



[19] Constraint-based learning algorithms assume that the d-separation encoded in the DAG is equivalent to the conditional independencies in the distribution. This is called a faithfulness assumption.

Step 1:

First, the pairs of variables (X,Y) are tested for conditional independence given the empty separating set of variables S (the size of the separating set is zero, k=0)[20]. We have previously established that the relationships between SWB and the seven variables of the material situation exist (no conditionality has been examined throughout the analysis so far). Also, the variables of the material situation are considered interdependent. We hence remove no edges in the first step. The graph remains the same after the first step.

Step 2:

For k=1 (there is only one variable in the separating subset) there is a set of 252 triples to be tested for conditional independencies. The scientific discussion continues over the relationships among income, country and SWB. The independence of SWB and economic growth can be supported, for example, by the conclusions made by Easterlin (1974; 1995; 2001; 2005; 2015). Easterlin proposes that the association between income and happiness is positive within a given country, but the level of happiness is not, on average, associated with national income per capita. The other authors collected the evidence not supporting Easterlin's conclusions and promoted the positive link between national income and the average level of SWB in a given country (Veenhoven 1991; Hagerty and Veenhoven 2003; Veenhoven and Hagerty 2006; Stevenson and Wolfers 2008; Diener and Biswas-Diener 2002). It can easily be argued that the level of a household's income in absolute terms mostly varies from country to country[21], whereas the material situation characteristics mostly depend on the household's income[22]. Hence, when the income is known, it is not necessary to know the household's country to draw a conclusion about its SWB and material situation. Thus, income seems to be the variable separating country from other variables in the model. More precisely, the node of the income seems to be the only node in the Markov blanket[23] of the country. The income d-separates the country from all the other variables in the model. As a result of such considerations, the income will remain the only node directly linked with the country in the expert model. The

[20] As the separation set is empty, the relationships between the pairs are actually unconditional in this case.

[21] The levels of disposable income are not too different in the four examined Central European countries, but the differences may be obvious when comparing rich and poor countries in a wider geographical context.

[22] The material situation may depend on material conditions in a wider sense, including the amount of savings, asset ownership, overall wealth, etc. Such variables are not embedded into the model.

[23] The Markov blanket of a variable A is the set consisting of the parents of A, the children of A, and the variables sharing a child with A. If all variables in the Markov blanket for A are instantiated, then A is d-separated from the rest of the network (Jensen 2001: 11).

formal notations of the conditional independencies for all steps of the analysis is in Appendix 2.

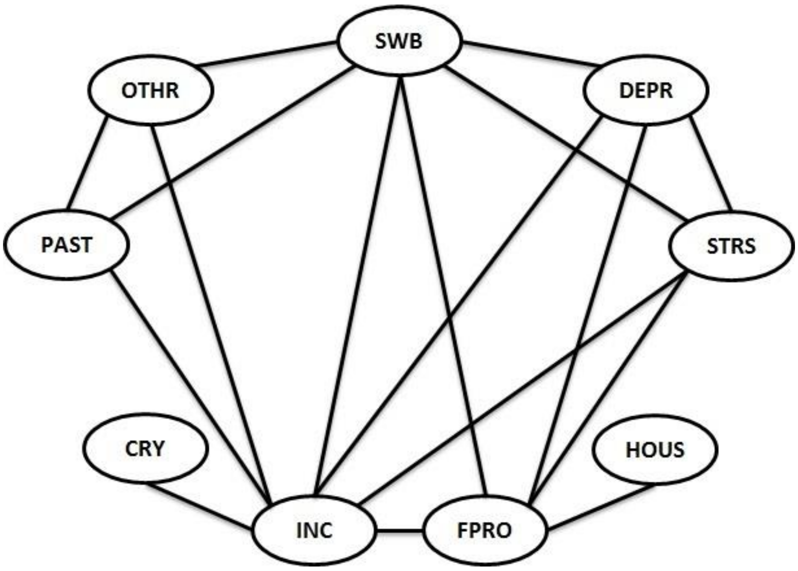
Apparently, the income is also important in terms of the relative income represented by the two nodes (PAST and OTHR). It is intuitively plausible that the measures of the relative income are linked to the absolute income unconditionally by the other measures in the model. A higher income will certainly increase the probability of a better relative income[24]. Similarly, when the income is low, the financial problems; housing problems; economic stress; and material deprivation may be more probable. The direct links between these variables and the income can be expected. On the other hand, it might be difficult, based on relative income alone, to directly draw a conclusion about material deprivation, financial and housing problems and economic stress.

The financial problems seem to represent an important variable separating the housing problems from the other seven variables in the model. The housing problems are mostly caused by the financial problems (caused, for example, by the low income). The conclusion on housing problems can be drawn directly from the household's income, but when the evidence of financial problems is known, the income will become conditionally independent. Given the financial problems, other variables also seem to be conditionally independent of the housing problems.

To conclude, 22 conditional independencies out of the 242 are derived in this step. Based on these identified conditional independencies, 19 edges are removed and the graph after this phase is shown in Figure 2.

[24] There are variables of the absolute income of neither others nor one's own past in the model. The own income in absolute terms is hence the only influencer.

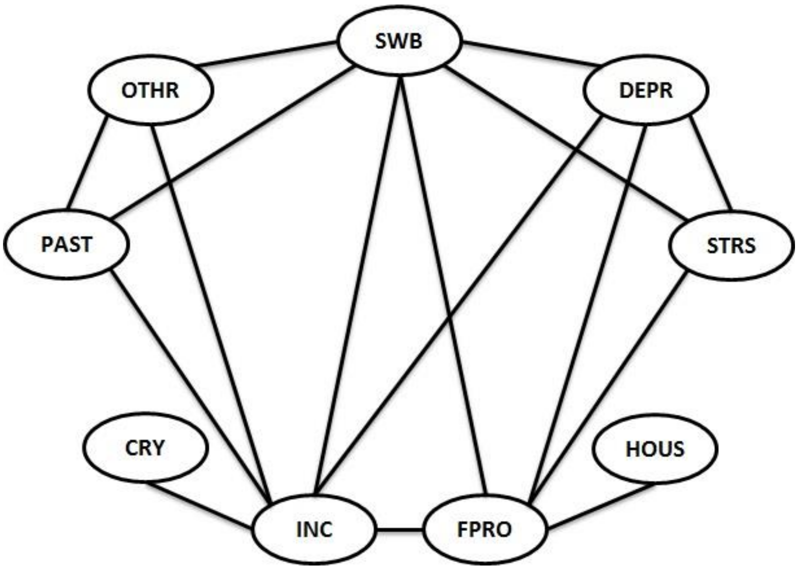
Figure 2: Graph after testing conditional independencies with k=1.



Step 3:

In the third step, the separating sets of the size k=2 (two separating variables) are considered. The conclusions on the perceived economic strain can obviously be drawn directly from the household's income, but they are conditionally independent, providing the material deprivation and the financial problems are known. People feel strained because of their objective financial problems and material deprivation (which may be influenced by the level of income) rather than by an insufficient level of the income itself. The single edge needs to be removed in this step of conditional independency testing. The graph resulting from this phase is shown in Figure 3.

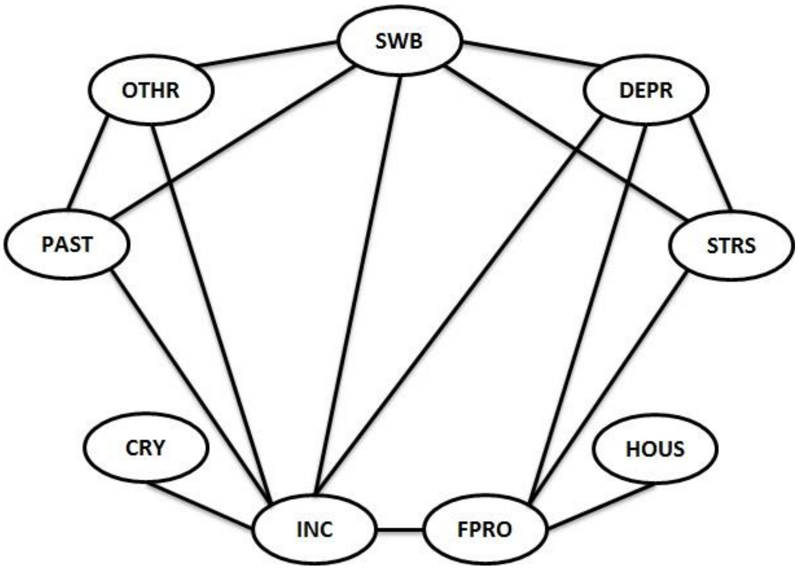
Figure 3: Graph after testing conditional independencies with k=2.



Step 4:

In this step, the separating sets of the size k=3 (three separating variables) are considered. It is difficult to find any exact evidence for such a complex relationship of conditional independence in the literature. SWB seems to be dependent on the financial problems (as a consequence of the low income), but only when no information about material deprivation, income and stress is available (all three together). The financial problems may or may not influence SWB. If we do not have another piece of information, we can hardly draw a conclusion of one from the other. But it seems to be possible to draw such a conclusion if we know the states of material deprivation, income and stress together. The edge linking the nodes INC and SWB needs to be removed and the graph resulting from this phase is shown in Figure 4.

Figure 4: Graph after testing conditional independencies with k=3.



Step 5:

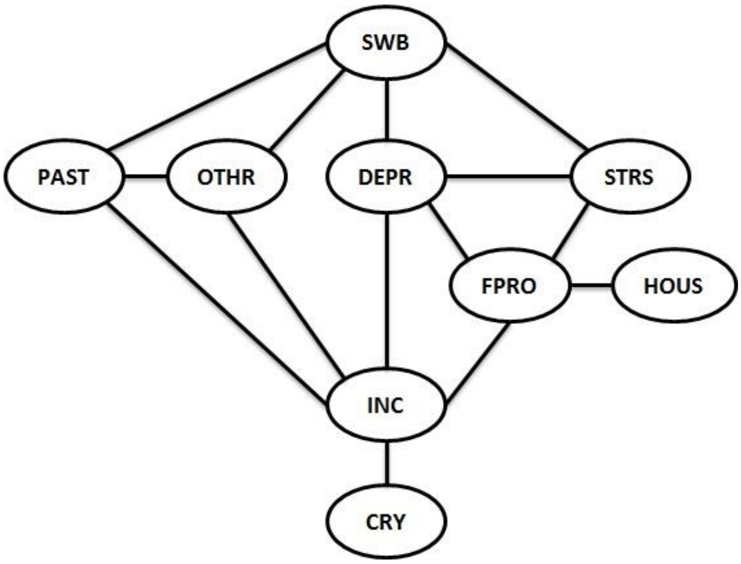
The separating set with four variables (k=4) is to be identified in this step. Although it is quite a complex relationship of conditional independence, the literature review comes up with helpful evidence on the relationship between the income and SWB. The newer research tends to understand the income rather as an indirect measure in terms of SWB (e.g., Christoph 2010). As seen before, the direct link between SWB and relative income can be traced to the literature, for example Clark and Oswald (1996); Ferrer-i-Carbonell (2005); Luttmer (2005); Clark et al. (2008); Dittmann and Goebel (2010); etc. Simply said, people are unhappy and unsatisfied when feeling their material situation has gotten worse comparing to either what it was before or what others have. The measures of relative income hence seem to stand in between the income and SWB.

The perceived economic strain as well as the material deprivation were examined in the EU-wide context by Fahey (2007), who suggests that these measures rather than the income should be employed as indicators for certain purposes as they provide better information than pure income thresholds in situations of inequality in the income between the EU member states. The evidence for the direct link between SWB and the perceived economic strain can be found, for example, in Mills et al. (1992) and Ervasti and Venetoklis (2010). Common sense reasoning is that SWB is reduced in the case of a household being unable to make ends meet. The link between SWB and material deprivation can be supported by, for example, Bellani and D'Ambrosio (2011). Basically, SWB drops if one yearns after something that cannot be afforded. In his paper on poverty Ringen (1988) discusses the measurement of poverty directly in terms of consumption (poverty as a low standard of consumption) and indirectly in terms of income (poverty as a low income). The consumption could be one of the variables standing in between SWB and income as an indirect measure.

In other words, the evidence is available for the four variables of material deprivation, economic stress and two measures of the relative income intermediating the variables of SWB and income. As a consequence, the edge between the nodes INC and SWB is removed. The testing can be stopped at the set of four separating variables (k=4) as the maximum adjacencies for all pairs (X,Y) is smaller than k+1 (which is five). The obtained undirected graph is the skeleton of a DAG. A planar version[25] of the skeleton is shown in Figure 5.

[25] A planar graph is a graph which can be plotted so that there are no edges crossing each other.

Figure 5: Skeleton - planar version.



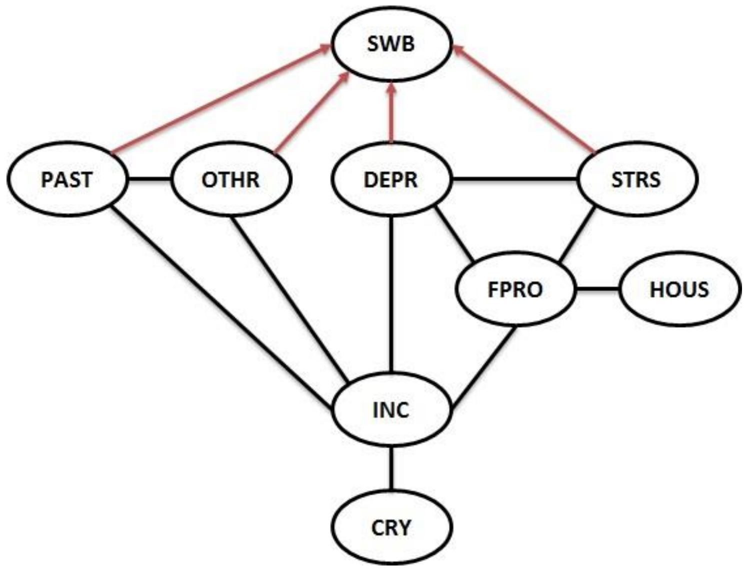
Step 6:

When the skeleton has been prepared, the orientation of the edges is the next task. The v-structures need to be identified in the next step of the PC algorithm. The following conditional independencies found in the first step of the PC algorithm indicate the v-structures.

- PAST⊥DEPR | INC
- PAST⊥STRS | INC
- OTHR⊥DEPR | INC
- OTHR⊥STRS | INC

In the first case, material deprivation (DEPR) and relative income as compared to one's own past (PAST) are conditionally independent given income (INC). Still, there is one more way between DEPR and PAST through SWB, whereas SWB∉S(DEPR,PAST). DEPR and PAST are independent if we know the income, but we must not know the state of SWB at the same time. The node of SWB is hence a collider on the way between DEPR and PAST. A v-structure hence must appear here. The explanations of the other three cases are analogous. Completed Partially Directed Acyclic Graph (hereinafter abbreviated as “CPDAG”), also called “essential graph” is obtained when the relevant edges are oriented according to the v-structures identified. It is shown in Figure 6 and the directed edges are drawn red.

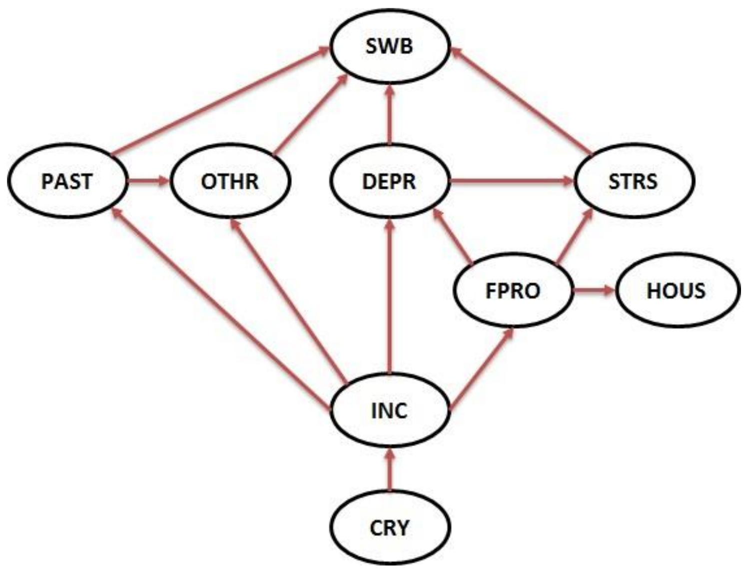
Figure 6: CPDAG (expert version).



Step 7:

The PC algorithm only produces a unique CPDAG, but it is unable to uniquely determine the directions of all of the edges. The remaining edges can be directed randomly so that no other v-structure is defined. We decided to orient the remaining edges from the bottom node of country towards the node of SWB. This last model is shown in Figure 7.

Figure 7: Bayesian network structure (expert version).

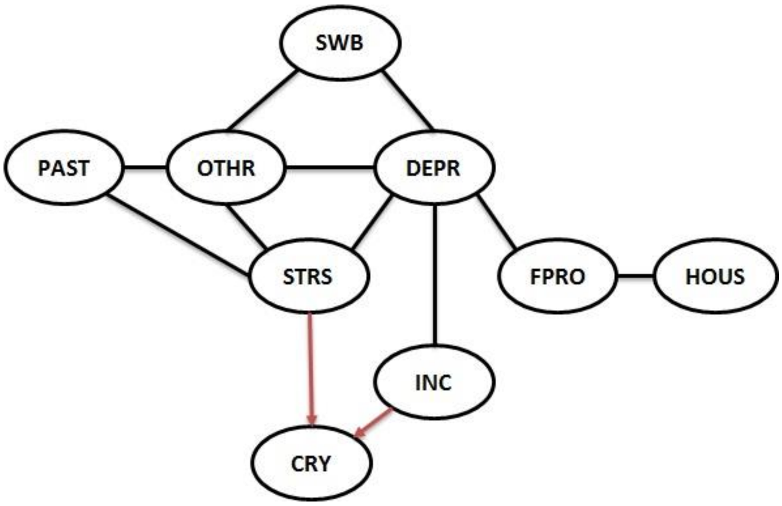


5.2 Greedy/EM and Gobnilp/EM models

Greedy (Chickering 2002) and Gobnilp (Cussens and Bartlett 2015) algorithms fall into the class of the search and score learning methods. They search for the structure maximizing a score function (BIC in our case).

The Greedy/EM model and the Gobnilp/EM model appeared to be in the same class of Markov equivalence because they have the same essential graphs (CPDAGs). These two models are said to be Markov equivalent. The essential graph is shown in Figure 8. The undirected edges in CPDAG can be oriented randomly.

Figure 8: CPDAG (Greedy/EM and Gobnilp/EM models)



The equivalent graphs imply the same set of conditional independence relationships via d-separation (they have the same d-separation properties). As the Gobnilp algorithm guarantees the optimal structure with respect to BIC, the Greedy/EM model is BIC optimal as well. These two equivalent models together are also referred to as "BIC optimal" hereinafter.

5.3 TAN/EM model

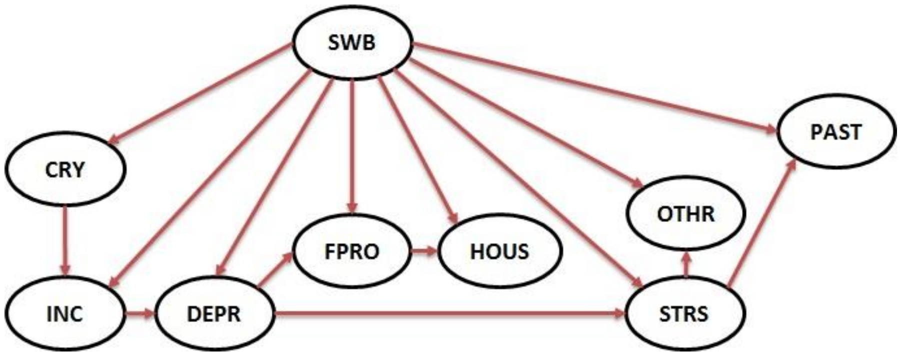
Unlike the previously discussed structure learning algorithms, the Tree Augmented Naïve Bayes (Friedman et al. 1997) is a method often used for classification problems[26]. This algorithm is an extension of the Naïve Bayes classifier[27]. The Tree Augmented Naïve Bayes relaxes the assumptions of independence of attributes given the class variable (Friedman et al. 1997). Hence,

[26] Learning a BN optimized with the intention to predict one of the variables.

[27] The Naive Bayes classifier assumes that the observed feature variables (attributes) are conditionally independent given the class variable (a variable to be classified based on the feature variables). If represented as a BN, the class variable has no parents and attributes have the class variable as the only parent.

each of the variables has the class variable and at most one other attribute variable as its parents. The attribute variables form a tree to augment the Naïve Bayes. The variable of SWB is the class one and, as such, it is the parent of all other variables – attributes in our model. The attributes are linked in the tree structure from CRY to PAST with the two tree branches[28]. This property should be clear from Figure 9, where the TAN/EM model is presented.

Figure 9: Bayesian network structure (TAN/EM version).



5.4 Ordinal logistic regression

The ordinal logistic regression (hereinafter referred also as "OLR") is employed in order to have another benchmark model. SWB is the response variable and the other eight variables are independent explanatory variables in our regression model. The explanatory variables are ordinal except for the country. Country as a multi-nominal variable is transformed to the following three dichotomous variables taking CRY=1 as the reference value in the following way:

- CRY2 (if CRY=Hungary then CRY2=1)
- CRY3 (if CRY=Poland then CRY3=1)
- CRY4 (if CRY=Slovakia then CRY4=1)
- CRY = the Czech Republic if all three variables are zero.

The OLR is presented in Table 2[29] and Table 3. The calculations are made in R software (R Core team 2014).

[28] The Hugin software enables us to define the root variable – the variable where the tree is starting. Following the expert model, country was defined as a starting point in this model. When the income is set as the starter, the Markov equivalent model is obtained – only the edge between CRY and INC is directed the other way around.

[29] Some negative coefficients are caused by the reversed coding of the ordered states.

Table 2: Ordinal logistic regression – coefficients.

Variable	Value	Std. Error	t value	p value
CRY2	0.38340221	0.11237612	3.411777	0.001
CRY3	0.78396892	0.09466560	8.281455	0.000
CRY4	0.13670987	0.10812207	1.264403	0.206
INC	0.08697156	0.03607650	2.410754	0.016
OTHR	0.46596625	0.06174160	7.547039	0.000
PAST	-0.27090906	0.05832283	-4.644991	0.000
STRS	-0.27411935	0.04513885	-6.072803	0.000
DEPR	-0.49603267	0.06095403	-8.137816	0.000
FPRO	-0.10474871	0.04964784	-2.109834	0.035
HOUS	-0.20441594	0.04658649	-4.387881	0.000

Table 3: Ordinal logistic regression – thresholds.

	Value	Std. Error	t value	p value
1|2	-2.32518538	0.29770178	-7.810452	0.000
2|3	-0.86866908	0.29486580	-2.945981	0.003
3|4	0.40559976	0.29490313	1.375366	0.169

We can observe that the probability of high SWB is especially increased by the higher personal income relative to others (OTHR) and the lower material deprivation (DEPR). Higher SWB is also implied by the low financial strain (STRS), few housing problems (HOUS) and income higher when compared to the past income (PAST). On the other hand, income (INC) and financial problems (FPRO) are both insignificant predictors on a 1% significance level (but are significant on a 5% level). Both of these variables are separated by the other variables in the expert version of the corresponding BN. Variable CRY4 is statistically insignificant on a 5% significance level (the difference between the Czech Republic and Slovakia is not statistically significant). Otherwise, the country is an important variable for SWB. The model is unable to statistically distinguish between states 3 and 4 of SWB on a 5% significance level. That might indicate that the difference between high and very high SWB levels cannot be explained using the given explanatory variables.

6 Testing and discussion

In order to select the most appropriate model, the outlined models are tested using the following criteria.

- The Bayesian Information Criterion (BIC).
- The prediction accuracy of SWB.

BIC (Schwartz, 1978), has become one of the most commonly used scoring criteria to search over possible networks. It is based on the likelihood function from which a penalty term is subtracted in order to prefer models with fewer parameters. The penalty is proportional to the number of independent parameters present in the model. A model with the highest BIC is selected from a finite set of models.

The likelihood of the four BNs measured by Log-likelihood (LL) and BIC is presented in Table 4. The measures are calculated in the Hugin software (Hugin 2014) based on the entire dataset of 3,259 data vectors.

Table 4: Models' likelihoods.

	Model	**LL**	**BIC**
1	Expert/EM	-29,579	-31,342
2	Expert/OLR	-29,938	-30,112
3-4	BIC optimal	-29,195	-29,822
5	TAN/EM	-29,285	-30,268

The BIC optimal model outperforms the other models in terms of BIC, as the optimality of BIC is ensured by the Gobnilp algorithm. The Expert/OLR is better in terms of BIC than the TAN/EM and Expert/EM, but it is the worst of all in terms of the log-likelihood. That behavior occurs because the Expert/OLR is penalized much less by BIC than the other models because its number of free parameters is only 43. The conditional probability tables of the Expert/OLR are restricted to the parameters of the ordered regression alone. Compared to that feature, the Expert/EM model has 436 free parameters. Its penalty is hence reduced and BIC is increased from -31,702 to -30,112 in the case of Expert/OLR.

The prediction accuracy can be defined as the number of correct predictions divided by the total number of predictions. It is important to point out that the classification of any variable in the model is not our primary goal. The main goal is to establish the best possible structure that explains the relationships of conditional independencies among SWB and the variables of the material situation. If SWB is selected to be predicted, still, the states of any other model variable could be predicted in the same way.

The 10-fold cross-validation is used for assessing the prediction accuracy. This method randomly splits the data into 10 equal-sized subsets. Nine subsets are used as training data to learn the parameters of the models whereas the obtained predictions of SWB are compared with the remaining single subset. This process is repeated ten times and each subset is used exactly once for the validation. The results are summarized for BNs as well as for the OLR in Table 5. The tests were done in the R software (R Core team 2014).

Table 5: SWB prediction accuracy.

	Model	**Accuracy**	**95% confidence interval**		**P-Value**
1	Expert/EM	41.58%	39.88%	43.29%	0.00
2	Expert/OLR	41.06%	39.36%	42.77%	0.00
3-4	BIC optimal	39.55%	37.87%	41.25%	0.00
5	TAN/EM	41.82%	40.12%	43.54%	0.00
-	OLR	42.13%	40.43%	43.85%	0.00

The visualization of results in the variable classification is often provided by a confusion matrix. The matrices are shown in Appendix 3[30] for all models. Several observations can be summarized.

- The four BNs and the OLR provide quite similar results in terms of SWB prediction. The 95% confidence intervals are largely overlapping.
- However, the OLR yields the best prediction of all. Thus, we cannot confirm that the suggested BNs provide better accuracy in terms of the prediction than widely used statistical methods based on regression.
- An expert model with the OLR parameters predicts slightly worse than the expert model with the EM parameters, but OLR parameters secure the monotonic classification whereas the EM parameters do not.
- The TAN/EM model is slightly better than the expert models as the algorithm is designed for the prediction of a selected variable, which is SWB in our case.
- The both expert models are better than the BIC optimal model, although the BIC optimal model provides the best likelihood measured by BIC and log-likelihood of all models given the data we have.

[30] The matrix compares the actual and predicted states. In our layout, the rows represent the instances in the predicted four classes and the columns the instances in the actual four classes of SWB.

- It can be reasonably expected that the prediction accuracy would be improved if other than economic factors were introduced into the models.
- The prediction accuracy around 40% seems to be low at first sight[31].Statistically, it is still significantly more than the No Information Rate (NIR)[32] on a 1% significance level. NIR is 27.80 % for all our models[33].

As the machine learned models fit the training data, they are often better than the expert model given the data. Still, the data may reflect the conditions under which it has been collected. We can reasonably expect that the expert model has more general applicability than the other models, because it reflects the state of the art in social research and depends to a lesser extent on the conditions under which the data were collected.

We mentioned earlier that non-monotone behavior may be a problem in models where the EM parameters are used. Intuitively, the two ordered sets are monotonic (monotone) if their given order is preserved or reversed. Ben David et al. (2009: 6,627) illustrated this property with the following example: "A model that guarantees monotonic classifications will never classify a young and healthy applicant in a higher life insurance premium category than an old, unhealthy one". Based on the analysis, we can reasonably expect the monotonic patterns of all parameters in our model except for the country (which is not ordinal). On the other hand, we can hardly assume our underlying dataset to be monotonic because the data is largely based on human judgment. Therefore, the EM parameters cannot be monotonic because the algorithm is not designed to satisfy the monotonicity conditions unless the dataset is monotonically consistent. The non-monotonicity leads to counterintuitive conclusions drawn from the model. We overcome this problem by using another local structure of conditional probabilities based on OLR.

Income is one of the central variables in the expert BN. It is conditionally independent of SWB given the set of four variables: relative income (two types); material deprivation; and economic stress. Income hence has no impact on SWB if the evidence for these four variables is available. In other words, the influence of the income on SWB is important, but it is not direct. For an illustration, a low income typically causes unsatisfactory relative income, high material deprivation and high financial problems. The high financial problems influence the economic stress thereafter. The four variables of relative income (two types), material

[31] Four states of SWB are predicted.

[32] P-value for all models is below the significance level. P-value is shown in Table 5.

[33] NIR is the largest class percentage in the data. If the most common class is always predicted, this resulting accuracy of such prediction will be equal to this rate. The most frequent class is SWB = 1 (906 times) in our case. When it is divided by the total number of data vectors (3,259), NIR is obtained.

deprivation and economic stress directly affect SWB. As another example, a low income does not necessarily mean either a low relative income (if it is not perceived as lower compared to others and to the past), high material deprivation (if a respondent is low in materialism), or high financial problems (if the expenses are low) or high economic stress (if there are enough savings)., SWB will be negatively affected only if these factors become burdensome. The low income itself is not necessarily the problem. We conclude that the objective material conditions themselves influence SWB rather indirectly through the subjective situational assessment of various aspects related to the individual material conditions. These subjective variables are crucial in terms of SWB. This conclusion is not far from the one reached by another paper on well-being where BNs were in use (Ceriani and Gigliarano 2016), who confirmed that *"subjective variables are strongly interlinked, as well as that objective dimensions influence subjective variables"* (Ceriani and Gigliarano 2016: 13).

Still, the material conditions constitute only a small part of the whole picture. Factors related to demography; aspirations; expectations; personality; social relations; the wider environment as well as situational factors play a role. Therefore, the results must be interpreted, taking into consideration that the material situation is only one piece in the large puzzle.

7 Conclusions

The commonly assumed association between a good material situation and SWB might be one of the reasons people want to have high material standards. Still, the relation between SWB and a particular variable of the material situation may not necessarily be direct, although it is statistically significant. The direct and the mediating factors are well visible in the corresponding BN. One conclusion from our analysis is that the objective conditions, such as income and financial problems, influence SWB indirectly through the subjective perception of relative income, material deprivation and economic stress.

The general applicability of our research results may be limited due to the research design and data. Only the data for four Central European countries are used due to their similarity. It is the purpose of this paper to study SWB only in the Central European context and any applications of our results in a different context must be undertaken with caution.

Also, the analysis has revealed certain knowledge about the "mainstream" income groups, but we know little about the very wealthy people. Furthermore, very poor people having problems with satisfying their basic needs are not specifically addressed in the study. We can hardly come to a conclusion about the SWB of any of these two groups. Furthermore, we know nothing about SWB of the people under the age of 18 as they are not included in the sample.

No causality is ascertained by the networks, the structure of the DAG implies the relationships of conditional independence only. These variables are inter-related, but no proof has been given in terms of the causality.

The accuracy of the SWB prediction is used as one of the metrics for BN assessment. Obviously, the models can be used for predicting any variables, but prediction is not the main purpose of the network. The main purpose is to model the complex relationships between the variables of the material situation and SWB. If the model is designed with the intention to maximize the prediction accuracy of a particular variable, for example SWB, different learning methods and approaches should be adopted.

So far, BNs have been used very scarcely in happiness research. As such, this paper should be understood as one of the very early steps on a longer journey. Several issues should be addressed in further research. For example, the strength values of the edges are not displayed in BNs, although it can be reasonably assumed that the links are unequally strong. Using the suggested models, causal and diagnostic inference may be employed to explore the relationship between SWB and individual material situation in Central Europe more in depth. Generally, BNs are still waiting for a greater extent of use by social scientists in the role of an analysis instrument. It is the modest wish of the author of this paper to help them along a little.

References

AHRENDT, D. - DUBOIS, H. - MEZENGER, E., 2015: An overview of quality of life in Europe. In: Glatzer, W. et al. (eds.), Global Handbook of Quality of Life, International Handbooks of Quality-of-Life. Dordrecht. Springer, pp. 625-661.

BELLANI, L. - D'AMBROSIO, C., 2011: Deprivation, social exclusion and subjective well-being. Social indicators research 104, Nr. 1, pp. 67-86.

BEN-DAVID, A. - STERLING. L. - TRAN, T., 2009: Adding monotonicity to learning algorithms may impair their accuracy. Expert Systems with Applications 36, Nr. 3, pp. 6627-6634.

BERTHOUD, R. - BRYAN, M., 2011: Income, deprivation and poverty: a longitudinal analysis. Journal of Social Policy 40, Nr. 1, pp. 135-156.

CERIANI, L. - GIGLIARANO, C., 2016: Multidimensional well-being: A Bayesian Networks approach. ECINEQ, Society for the Study of Economic Inequality 399.

CHICKERING, D. M., 2002: Optimal structure identification with greedy search. Journal of Machine Learning Research, Nr. 3, pp. 507–554.

CHRISTOPH, B., 2010: The relation between life satisfaction and the material situation: A re-evaluation using alternative measures. Social Indicators Research 98, Nr. 3, pp. 475-499.

CLARK, A. E. - OSWALD, A. J., 1996: Satisfaction and comparison income. Journal of public economics 61, Nr. 3, pp. 359-381.

CLARK, A. E. - FRIJTERS, P. - SHIELDS, M. A., 2008: Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. Journal of Economic Literature 46, Nr. 1, pp. 95-144.

CRONBACH, L. J., 1951: Coefficient alpha and the internal structure of tests. Psychometrika 16, Nr. 3, pp. 297–334.

CUSSENS, J. - BARTLETT, M., 2015: GOBNILP software.

DEMPSTER, A. P. - LAIRD, N. M. - RUBIN, D. B., 1977: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B 39, Nr. 1, pp. 1–38.

DIENER, E. - BISWAS-DIENER, R., 2002: Will money increase subjective well-being? Social indicators research 57, Nr. 2, pp. 119-169.

DITTMANN, J. - GOEBEL, J., 2010: Your house, your car, your education: The socioeconomic situation of the neighborhood and its impact on life satisfaction in Germany. Social Indicators Research 96, Nr. 3, pp. 497-513.

EASTERLIN, R. A., 1974: Does economic growth improve the human lot? Some empirical evidence. Nations and households in economic growth 89, pp. 89-125.

EASTERLIN, R. A., 1995: Will raising the incomes of all increase the happiness of all? Journal of Economic Behavior & Organization 27, Nr. 1, pp. 35-47.

EASTERLIN, R. A., 2001: Income and happiness: Towards a unified theory. Economic journal 111, Nr. 473, pp. 465-484.

EASTERLIN, R. A., 2005: Feeding the illusion of growth and happiness: A reply to Hagerty and Veenhoven. Social indicators research 74, Nr. 3, pp. 429-443.

EASTERLIN, R. A., 2015: Happiness and economic growth: The evidence. In: Glatzer, W. et al. (eds.), Global Handbook of Quality of Life, International Handbooks of Quality-of-Life. Dordrecht, Springer, pp. 283-299.

ERVASTI, H. - VENETOKLIS, T., 2010: Unemployment and subjective well-being: An empirical test of deprivation theory, incentive paradigm and financial strain approach. Acta Sociologica 53, Nr. 2, pp. 119-139.

European Foundation for the Improvement of Living and Working Conditions (Euorfound), 2012a: Third European Quality of Life Survey. Sampling report.

European Foundation for the Improvement of Living and Working Conditions (Euorfound), 2012b: Third European Quality of Life Survey. Technical report.

European Foundation for the Improvement of Living and Working Conditions (Euorfound), 2014: European Quality of Life Survey [computer file], 2011-2012: 2nd Edition. Colchester, Essex: UK Data Archive [distributor]. SN: 7316. http://dx.doi.org/10.5255/UKDA-SN-7316-2

FAHEY, T., 2007: The Case for an EU-wide Measure of Poverty. European Sociological Review 23, Nr. 1, pp. 35-47.

FERRER-I-CARBONELL, A., 2005: Income and well-being: an empirical analysis of the comparison income effect. Journal of Public Economics 89, Nr. 5, pp. 997-1019.

FRIEDMAN, N. - GEIGER, D. - GOLDSZMIDT, M., 1997: Bayesian network classifiers. Machine learning 29, Nr. 2-3, pp. 131-163.

HAGERTY, M. R. - VEENHOVEN, R., 2003: Wealth and Happiness Revisited: Growing National Income Does Go with Greater Happiness. Social Indicators Research 64, Nr. 1, pp. 1-27.

HUGIN EXPERT, 2014: Hugin Researcher API 8.0 (www. hugin. com).

JENSEN, F. V., 2001: Bayesian Networks and Decision Graphs. Department of Computer Science, Aalborg University.

LAYTE, R. - WHELAN, C. T. - MAITRE, B. - NOLAN, B., 2001: Explaining levels of deprivation in the European Union. Acta Sociologica 44, Nr. 2, pp. 105-121.

LUTTMER, E. F. P., 2005: Neighbors as negatives: Relative earnings and well-being. The Quarterly Journal of Economics 120, Nr. 3, pp. 963-1002.

MILLS, R. J. - GRASMICK, H. G. - MORGAN, C. S. - WENK, D., 1992: The effects of gender, family satisfaction, and economic strain on psychological well-being. Family Relations 41, Nr. 4, pp. 440-445.

NUNNALLY, J. C., 1978: Psychometric Theory, 2nd ed. New York, McGraw-Hill.

PEARL, J., 1982: Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach. AAAI - 82 Proceedings, pp. 133–136.

R CORE TEAM, 2014: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, (http://www.R-project.org).

RINGEN, S., 1988: Direct and indirect measures of poverty. Journal of social policy 17, Nr. 3, pp. 351-365.

SCHMITT, N., 1996: Uses and abuses of coefficient alpha. Psychological Assessment 8, Nr. 4, pp. 350-353.

SCHWARTZ, G. E., 1978: Estimating the dimension of a model. Annals of Statistics 6, Nr. 2, pp. 461–464.

SPIRTES, P. - GLYMOUR, C. N. - SCHEINES, R. - HECKERMAN, D. - MEEK, C. - COOPER, G. et al., 2000: Causation, Prediction, and Search. MIT Press, second edition.

STEVENSON, B. - WOLFERS, J., 2008: Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox. Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution 39, Nr. 1, pp. 1-102.

ŠVORC, J., 2018: Subjective well-being and individual material situation in four countries of Central Europe. Sociológia 50, Nr. 6, pp. 727-759.

ŠVORC, J., 2019: Subjective Well-Being and Individual Material Situation in Four Countries of Central Europe. PhD. Dissertation. Faculty of Management, University of Economics, Prague.

ŠVORC, J. – VOMLEL, J., 2019: Bayesian networks for the analysis of subjective well-being. In: Kratochvíl, V. (ed.), Proceedings of the 22nd Czech-Japan Seminar on Data Analysis and Decision Making, Matfyzpress, pp. 175–188.

VEENHOVEN, R., 1991: Is happiness relative? Social indicators research 24, Nr. 1, pp. 1-34.

VEENHOVEN, R. - HAGERTY, M., 2006: Rising happiness in nations 1946–2004: A reply to Easterlin. Social Indicators Research 79, Nr. 3, pp. 421-436.

WHELAN, C. T., 1992: The role of income, life-style deprivation and financial strain in mediating the impact of unemployment on psychological distress: Evidence from the Republic of Ireland. Journal of Occupational and Organizational Psychology 65, Nr. 4, pp. 331-344.

WHELAN, C. T. - LAYTE, R. - MAITRE, B. - NOLAN, B., 2001: Income, deprivation, and economic strain. An analysis of the European community household panel. European sociological review 17, Nr. 4, pp. 357-372.

WHELAN, C. T. - MAITRE, B., 2007: Income, deprivation and economic stress in the enlarged European Union. Social Indicators Research 83, Nr. 2, pp. 309-329.

Appendix 1 – EQLS questions and scales

Abbrev.	Variable	EQLS questions	EQLS scale / items
SWB	Subjective well-being	Taking all things together on a scale of 1 to 10, how happy would you say you are?	1=very unhappy; 10=very happy
		All things considered, how satisfied would you say you are with your life these days?	1=very dissatisfied; 10=very satisfied
PAST	Relative income compared to the past	When you compare the financial situation of your household 12 months ago and now, would you say it has become better, worse or remained the same?	1=better; 2=the same; 3=worse
OTHR	Relative income compared to others	Could you please evaluate the financial situation of your household? In comparison to most people in your country would you say it is much worse, somewhat worse, neither worse nor better, somewhat better or much better?	1=much worse; 2=somewhat worse; 3= neither worse nor better; 4=somewhat better; 5=much better
STRS	Perceived economic stress	Thinking of your household's total monthly income: is your household able to make ends meet very easily, easily, fairly easily, with some difficulty, with difficulty or with great difficulty?	1=very easily; 2=easily; 3=fairly easily; 4=with some difficulty; 5=with difficulty; 6=with great difficulty
DEPR	Material deprivation	There are some things that many people cannot afford, even if they would like them. For each of the following things on this list, can I just check whether your household can afford it if you want it?	keeping your home adequately warm; paying for a week's annual holiday away from home (not staying with relatives); replacing any worn-out furniture; having a meal with meat, chicken, fish every second day (if wanted); buying new, rather than second-hand, clothes; having friends or family for a drink or meal at least once a month
FPRO	Financial problems	Has your household been in arrears at any time during the past 12 months, that is, unable to pay as scheduled any of the following?	rent or mortgage payments for accommodation; utility bills such as electricity, water and gas; payments related to consumer loans; payments related to informal loans (from friends and relatives not living in the household)
HOUS	Housing problems	Do you have any of the following problems with your accommodation?	shortage of space; rot in windows, doors or floors; damp or leaks in walls or roof; lack of indoor flushing toilet; lack of bath or shower; lack of place to sit outside

Appendix 2 – Notations of conditional independencies

Step 2:

- CRY $\perp$ HOUS, FPRO, DEPR, STRS, OTHR, PAST, SWB | INC (country is conditionally independent of housing defects; financial problems; material deprivation; economic stress; relative income compared to others; and compared to the past and SWB given income).
- OTHR, PAST $\perp$ HOUS, DEPR, FPRO, STRS | INC (relative income is conditionally independent of housing problems; material deprivation; financial problems; and economic stress given the income).
- INC, DEPR, CRY, OTHR, PAST, STRS, SWB $\perp$ HOUS | FPRO (housing defects are conditionally independent of income; material deprivation; country; relative income; economic stress; and SWB given the financial problems).

Step 3:

- STRS $\perp$ INC | FPRO, DEPR (economic stress is conditionally independent of income given financial problems and material deprivation).

Step 4:

- SWB $\perp$ FPRO | DEPR, INC, STRS (SWB is conditionally independent of financial problems given material deprivation; income; and economic stress).

Step 5:

- SWB $\perp$ INC | DEPR, OTHER, PAST, STRS (SWB is conditionally independent of income given material deprivation; relative income compared to others as well as to one's own past; and economic stress).

Appendix 3 – Confusion matrices

Model	Prediction	Reference				
		Actual=1	Actual=2	Actual=3	Actual=4	Total
Expert/EM	Predicted=1	597	301	174	128	1,200
	Predicted=2	196	328	250	180	954
	Predicted=3	64	142	173	145	524
	Predicted=4	49	122	153	257	581
Expert/OLR	Predicted=1	579	286	148	100	1,113
	Predicted=2	226	344	261	189	1,020
	Predicted=3	14	40	33	39	126
	Predicted=4	87	223	308	382	1,000
BIC optimal	Predicted=1	566	289	159	114	1,128
	Predicted=2	204	267	193	151	815
	Predicted=3	97	241	269	258	865
	Predicted=4	39	96	129	187	451
TAN/EM	Predicted=1	573	278	136	96	1,083
	Predicted=2	195	307	223	172	897
	Predicted=3	71	144	189	148	552
	Predicted=4	67	164	202	294	727
OLR	Predicted=1	576	264	130	82	1,052
	Predicted=2	224	337	259	195	1,015
	Predicted=3	33	98	101	74	306
	Predicted=4	73	194	260	359	886
	Total	906	893	750	710	3,259