Analyzing the Effects of the Dwelling Characteristics on Space Heating Costs with the Rank Correlation Bayesian Network Model

Volkan Sevinç (✉ vsevinc@mu.edu.tr)

Mugla Sitki Kocman University Faculty of Science: Mugla Sitki Kocman Universitesi Fen Fakultesi

<https://orcid.org/0000-0003-4643-443X>

Research Article

Keywords: Rank correlation Bayesian network model, Space heating costs, house heating, household economy, Bayesian networks

Posted Date: June 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-436208/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Analyzing the Effects of the Dwelling Characteristics on Space Heating Costs with the Rank Correlation Bayesian Network Model

Volkan Sevinc

*Mugla Sitki Kocman University, Faculty of Science, Department of Statistics, Kotecli
Kampusu, 48000, Mugla / Turkey, vsevinc@mu.edu.tr, +90 252-211-1532*

ABSTRACT

Energy is one of the main concerns of humanity because energy resources are limited and costly. In order to reduce the costs and to use the energy for space heating effectively, new building materials, techniques and insulations facilities are being developed. Therefore, it is important to know which factors affect the space heating costs. This study aims to introduce the novel Rank Correlation Bayesian Network model and its application in analyzing the effects of dwelling characteristics on the space heating costs. The results show that the constructed Rank Correlation Bayesian Network model performed better than the Bayesian networks models estimated by Bayesian search, PC and Greedy Thick Thinning algorithms, which are kinds of structure learning algorithms having different kinds of estimation mechanisms to build Bayesian networks. The constructed Rank Correlation Bayesian Network model shows that the space heating costs of the dwellings are mostly affected by the heating systems used. Coal stoves, air conditioners and electric stoves appear to be the costliest heating systems. The second most important factor appears to be the existence of external wall insulation. The lack of external wall insulation almost doubles the

space heating costs. The third most important factor is the building age. Dwellings on the ground floors and the first floors appear to pay the highest space heating costs. Therefore, dwellings on these floors need to be more effectively insulated. As the size of the dwelling increases the heating cost increases too. Another result is that facing directions and floor levels of the dwellings have the least effects on their space heating.

Keywords: Rank correlation Bayesian network model; Space heating costs; house heating; household economy; Bayesian networks

1. Introduction

Dwellings use energy for various purposes such as space and water heating, cooking, lighting, running electrical appliances. According to Eurostat (2017) the main use of energy in dwellings, which is 64% of the final energy consumption, is for space heating. Thus, space heating is the major cause of energy consumption in dwellings. Energy is costly. Thus, space heating cost is one of the important elements in the family budgets. In addition, it is reported that dwellings are the fourth largest source of CO₂ emissions in the EU and they produce 9.9% of the total emissions (Martinopoulos et al., 2016). Thus, finding ways to control or cut back on these costs are necessary to reduce energy consumption, to save natural resources and to reduce carbon emissions. The dwelling characteristics and building design have significant effects on the heating energy performance of the buildings (Kazanas et al., 2014). This study aims to investigate the effects of the dwelling characteristics on the space heating cost by the use of Bayesian networks.

In literature, there are various studies concerning heating cost and employing various methods. Rehdanz (2007) investigated the effects of space heating costs, building characteristics and the socio-economic status of dwellings in Germany. As a result, it was understood that the heating costs of the home owners are less than those of tenants. Hassouneh et al. (2010) performed a study to investigate the influence of window types on the energy balance of the apartment type buildings in Amman, Jordan. Meier and Rehdanz (2010) used multiple linear regression analysis to investigate the factors determining the space heating costs in the United Kingdom. An important finding in the study is that, unlike in Germany, house owners have greater heating costs than tenants. Zoric and Hrovatin (2010) investigated the effects of heating and dwelling types on the energy efficiency of the houses in Slovenia. It was observed that heating type directly affects energy efficiency. Nair et al. (2010) conducted a survey on house owners to reduce the energy use of individual detached houses in Switzerland. As a result of the study, it was concluded that energy-efficiency investments should be supported. Laureti and Secondi (2012) investigated different family groups, heating type, heating technologies, dwelling structural features and the effects of family socio-economic status on the heating costs in Italy. They showed that the space heating concept is not only influenced by the socio-economic characteristics of the families but also the external factors which they cannot control. Kazanasmaz et al. (2014) examined the relation between architectural considerations and energy performance of the buildings in Izmir, Turkey and they made suggestions to predict the level of energy performances of the buildings in the early design phase. Ziemele et al. (2015) studied the economy of heat cost allocation in apartment buildings in Riga, Latvia and suggested a model to fairly allocate the heating costs in each apartment. Wessels (2015) stated that the energy consumption of the buildings built in the Netherlands before 1970 was very high and

these buildings constituted approximately 12.5% of the buildings in the country. Therefore, he discussed ways of minimizing the energy consumption of these buildings. Longhi (2015) investigated the effects of changes in socio-economic conditions of dwelling members and building properties on energy costs. It was observed that socio-economic changes had little effect on energy costs, but dwelling size had a significant effect on energy costs. Schmitz and Madlener (2016) investigated the effects of socio-demographic variables such as age and gender on building types and technical characteristics in explaining the heating costs of houses. One of the important findings is that the heating cost rate of low-income dwellings is higher than the dwellings with high income levels. Hill (2015) aimed to investigate the determinants of residential energy cost in Austria. It was concluded that energy costs are more affected by regional characteristics, dwelling characteristics, socio-economic factors and income level. Risch and Salmon (2017) aimed to identify the main factors affecting the energy consumption of the dwellings in France. It was concluded that residential and climatic characteristics made the greatest impact on the energy consumption in the houses. However, socio-demographic factors were observed to have a very low effect. Rose and Kragh (2017) studied the distribution of heating costs in multi-story apartment buildings in Denmark by suggesting a cost calculation method. Recently machine learning techniques like Bayesian networks are becoming popular as the performance and capacity of the computers are increasing. Compared to other statistical estimation methods that work with additive models, Bayesian networks advantageous and robust especially in handling big data because they have no restrictive distributional or probabilistic assumptions. This study aims to introduce the novel model Rank Correlation Bayesian Network (RCBN) model with its application in space heating cost estimation of dwellings.

2. Materials and Methods

2.1. Bayesian networks

Bayesian networks are a branch of probabilistic graphs, which were introduced by Pearl (1985). A Bayesian network can be looked upon a graphical representation of conditional joint probability distributions of the variables that they involve. As well as representing a probability distribution, Bayesian networks also display the causal inference among these variables. A Bayesian network consists of two main components as graphical structure and probabilistic structure. The graphical structure consists of nodes which represent the random variables and the edges which are directed arrows among the nodes. The nodes connected to each other by an edge are called adjacent and this connection indicates an association between the two nodes. The lack of any edge between two nodes, on the other hand, indicates the absence of an association between them. When each pairs of nodes are adjacent, the graph is said to be a complete graph. An example of a Bayesian network is shown in Figure 1.

Figure 1. Example of Bayesian network

Figure 1 consists of 6 nodes labeled as C1, C2, C3, C4, C5 and C6. A node from which an edge is directed is called a parent node while a node which an edge is directed to is

called a child node. In Figure 1, for example, C1 is a parent node of C2 and C5, while C5 is a child node of C1 and C3. A path can be defined as the series of edges and the nodes connected by them. Bayesian networks have the Directed Acyclic Graph (DAG) structure in which no cycles are allowed on the paths. That is, it is impossible to return to any node by following the path of any edge directed from this node.

The probabilistic structure of the Bayesian networks consists of the conditional probability tables. Conditional probability tables are determined by the joint probability distribution of the nodes. Let the set of the nodes in a Bayesian network be $V = (V_1, V_2, \dots, V_k)$. For a specific value of $V_i = v$ and a configuration u of the elements in $U \subseteq V|V_i$, the conditional probabilities are given as follows.

$$P(v_i|u) = \frac{P(v_i, u)}{P(u)} = \frac{\sum_{s \in S} P(v_i, u, s)}{\sum_{s \in S, v \in V_i} P(u, s, v)} \quad (1)$$

where s is a configuration of $S = V|U, V_i$. s and v_i denote the sample spaces of S and V_i respectively. Moreover, the joint probability density function of V is as follows.

$$P(V_1, V_2, \dots, V_k) = \prod_{i=1}^k P(V_i|v) \quad (2)$$

There are two approaches to build the structure of a Bayesian network. In the first approach the structure is constructed manually based on expert opinion. The number of nodes and edges and their directions are determined by the expert considering the causal relations among the nodes. Hence, in this kind of Bayesian networks, the edges among the nodes also imply a causal relation. The second approach is called structure learning. In the structural learning the Bayesian network is built directly from the data set by the help of various algorithms. In a Bayesian network built by any kind of structural learning algorithm, however, the edges do not need to imply a causal relationship;

instead, they just indicate the presence of a dependency. More detailed information about Bayesian networks can be found in Pearl (1988), Jensen and Nielsen (2007), Holmes and Jain (2008).

2.2. Rank correlation Bayesian network model

While constructing the edges of the Bayesian network manually, it may be useful to refer some measures of dependency as well as expert opinion. As it is known, correlation does not imply causality; on the other hand, it indicates a dependency between two variables. Thus, we introduce the Rank Correlation Bayesian Network (RCBN) model in which the dependencies among the nodes are detected by an appropriate rank correlation measure. After detecting the dependencies between node pairs, the edges between the node pairs having a significant rank correlation coefficient are determined by expert opinion also considering the directions of the dependencies. Afterwards, using the data set the posterior probabilities are calculated and the conditional probability tables are constructed by EM algorithm. In order to measure the dependency between two variables, various measures like Spearman's rank correlation coefficient, Goodman and Kruskal's gamma, Kendall's tau rank correlation coefficient exist. All of these measures are based on the rank of the observations. When the variables are categorical, Spearman's rank correlation coefficient is an appropriate measure of correlation to use in the RCBN model because it preserves the given order between the nodes. Spearman's rank correlation coefficient is calculated as follows (Sprent and Smeeton, 2001).

$$r_s = 1 - \frac{6 \sum (r_i - s_i)^2}{n(n^2 - 1)} \quad (3)$$

where r_i is the rank of the category i of the first variable, s_i is the rank of the category i of the second variable and n is the total number of observations.

2.3. Data and model

The data was derived from 500 dwellings in Mugla province in Turkey. The data involves 10 variables as the nodes of the Bayesian network: space heating costs, building age, heating system, dwelling type, number of bedrooms, dwelling size, floor level, facing direction, external wall insulation and double glazing. The list of these nodes and their definitions are given in Table1.

Table 1. Nodes with their definitions and levels

Nodes	Definitions	Levels
Space heating costs	The average monthly amount of money (Turkish Lira-TL) paid for space heating in the dwellings	0-99 100-199 200-299 300-399 400 or over
Building age	The age of the building in years	0-9 10-19 20-29 30 or over
Heating system	The heating system used for space heating in the dwelling	Air conditioner or Electric Stove Coal Stove Central Heating or Boiler
Dwelling type	The type of the dwelling as a flat or a detached house	Flat Detached House
Number of bedrooms	The number of the bedrooms aside from 1 living room	1 Living Room 1 Living Room and 1 Bedroom 1 Living Room and 2 Bedrooms 1 Living Room and 3 Bedrooms 1 Living Room and 3 Bedrooms or more
Dwelling size	Size of the dwelling in meter-squares (m^2)	0-59 60-119 120-179 180 or over

Floor level	The level of the floor on which the dwelling is located	Basement Ground Floor First Floor Second Floor Third Floor Fourth Floor Fifth Floor or higher
Facing direction	The direction which the longest side of the dwelling faces to	East West North South
External wall insulation	Existence of the external wall insulation at the dwelling	Available Not Available
Double glazing	Existence of the double glazing at the dwelling	Available Not Available

167 In order to detect the existing dependencies between the nodes, Spearman's rank
168 correlation coefficients were calculated in Knime software (Berthold et al., 2007). The
169 node pairs, the calculated correlation coefficients and their significances at $\alpha = 0.05$ are
170 presented in Table 2.

171 **Table 2.** Spearman's rank correlation coefficients and their significances ($\alpha = 0.05$)

First Node	Second Node	Correlation	p-value	Significant
Dwelling Type	Heating System	-0.259251238	4.47391E-09	*
Dwelling Type	Building Age	0.141531624	0.001560306	*
Dwelling Type	External Wall Insulation	0.118609027	0.008123311	*
Dwelling Type	Facing Direction	-0.073144651	0.103370971	
Dwelling Type	Space Heating Cost	0.237781242	8.10564E-08	*
Dwelling Type	Dwelling Size	-0.071110886	0.113347771	
Dwelling Type	Number of Bedrooms	0.218851243	8.3707E-07	*
Dwelling Type	Double Glazing	0.214617784	1.37279E-06	*
Dwelling Type	Floor Location	0.493375672	0	*
Heating System	Building Age	-0.505712851	0	*

Heating System	External Wall Insulation	-0.426623893	0	*
Heating System	Facing Direction	0.133215213	0.002923896	*
Heating System	Space Heating Cost	-0.569036641	0	*
Heating System	Dwelling Size	0.087124907	0.052243661	
Heating System	Number of Bedrooms	-0.008765185	0.845456886	
Heating System	Double Glazing	-0.368024686	0	*
Heating System	Floor Location	-0.117972134	0.008473714	*
Building Age	External Wall Insulation	0.519070819	0	*
Building Age	Facing Direction	-0.042463102	0.344815376	
Building Age	Space Heating Cost	0.499467006	0	*
Building Age	Dwelling Size	-0.086257	0.054642132	
Building Age	Number of Bedrooms	0.10829607	0.015721123	*
Building Age	Double Glazing	0.30682401	2.70384E-12	*
Building Age	Floor Location	0.05209005	0.246402249	
External Wall Insulation	Facing Direction	0.017064623	0.704316735	
External Wall Insulation	Space Heating Cost	0.565623456	0	*
External Wall Insulation	Dwelling Size	-0.119807065	0.007499046	*
External Wall Insulation	Number of Bedrooms	0.046851055	0.29721869	
External Wall Insulation	Double Glazing	0.309273733	1.77591E-12	*
External Wall Insulation	Floor Location	-0.074438547	0.097394746	
Facing Direction	Space Heating Cost	-0.0894886	0.046152876	*
Facing Direction	Dwelling Size	0.155379023	0.000508279	*
Facing Direction	Number of Bedrooms	-0.179230033	5.8648E-05	*
Facing Direction	Double Glazing	-0.022037722	0.624048756	
Facing Direction	Floor Location	-0.035003599	0.436199876	
Space Heating Cost	Dwelling Size	-0.233134159	1.46506E-07	*
Space Heating Cost	Number of Bedrooms	0.414932942	0	*
Space Heating Cost	Double Glazing	0.21728928	1.00584E-06	*
Space Heating Cost	Floor Location	0.019383787	0.666406719	
Dwelling Size	Number of Bedrooms	-0.427150608	0	

Dwelling Size	Double Glazing	-0.082698176	0.065453841	
Dwelling Size	Floor Location	0.02827045	0.529489223	
Number of Bedrooms	Double Glazing	0.006183758	0.890626291	
Number of Bedrooms	Floor Location	0.097168815	0.030318401	*
Double Glazing	Floor Location	0.082823335	0.065045791	

In Table 2, the node pairs having a significant correlation and hence, a dependency between them are marked with a (*) sign. At this stage the signs of the correlation coefficients can be ignored because we are only interested in detecting the dependencies. The increase or decrease in the levels of the nodes with respect to each other will be determined later in the RCBN model to be constructed more accurately also considering the directions of the dependencies. Hence, only the node pairs having a significant correlation between them were connected with edges and the edges are directed from the effecting nodes to the affected nodes by expert opinion. The RCBN model, for analysis of the effects of the dwelling characteristics on the space heating costs, was built manually using Netica (2020) software. The constructed RCBN model is presented in Figure 2. The scores behind the bars represent the probability of occurrence belonging to the corresponding levels.

Figure 2. Rank Correlation Bayesian Network model for analysis of the effects of the dwelling characteristics on the space heating costs

2.4. Model validation

The confusion matrix that demonstrates the estimation performance of the RCBN is given in Table 3. The values on the diagonal of the matrix indicate the number of correctly classified cases. The validation was performed by using simulated data of 100,000 cases. Instead of a portion of 500 data used in the study, simulated data was preferred to measure the performance of the model with as many cases as possible.

Table 3. Confusion matrix for the Rank Correlation Bayesian Network model

Space heating costs		Predicted				
		0 - 99	100-199	200-299	300-399	400 or over
Actual	0 - 99	7901	550	0	0	1
	100 - 199	7994	13463	745	0	1

200 - 299	7821	2498	11257	791	178
300 - 399	7866	737	1747	13345	764
400 or over	7963	308	368	2797	10905

When Table 3 is examined, it is seen that the correct estimation rate of the RCBN model for the space heating costs is 0.56871 (56871/100000).

In order to make a comparison between the RCBN model and other structure learning models, three different Bayesian networks were built based on Bayesian Search (BS) (Heckerman et al., 2000), PC (Spirtes et al., 1991) and Greedy Thick Thinning (GTT) (Cheng et al., 1997) algorithms, which are different kinds of structural learning algorithms having different mechanisms to build Bayesian networks. The Bayesian networks based on the algorithms were generated in GeNIe (2020) software and rearranged in Netica (2020) software. The Bayesian network model for the analysis of the effects of the dwelling characteristics on space heating costs that is based on BS algorithm is shown in Figure 3.

Figure 3. Bayesian network estimated by Bayesian Search algorithm

The confusion matrix that demonstrates the estimation performance of the Bayesian network that was built using BS algorithm is presented in Table 4. The values on the diagonal of the matrix indicate the number of correctly classified cases. The confusion matrix was obtained by using simulated data of 100,000 cases.

Table 4. Confusion matrix of the Bayesian network estimated by Bayesian Search Algorithm

Space heating costs		Predicted				
		0 - 99	100-199	200-299	300-399	400 or over
Actual	0 - 99	0	714	171	0	82
	100 - 199	0	17530	2823	0	425
	200 - 299	0	11664	12147	0	2937
	300 - 399	0	4421	9221	0	12393

	400 or over	0	2087	5117	0	18268
--	-------------	---	------	------	---	-------

When Table 4 is examined, it is seen that the correct estimation rate of the Bayesian network built by BS algorithm for the space heating costs is 0.47945 (47945/100000). However, BS algorithm failed to correctly estimate the space heating costs that are between 0-99 TL and 300-399 TL.

The Bayesian network model for the analysis of the effects of the dwelling characteristics on space heating costs that is based on PC algorithm is presented in Figure 4.

Figure 4. Bayesian network estimated by PC algorithm

223 The confusion matrix that demonstrates the estimation performance of the Bayesian
 224 network that was constructed by PC algorithm is presented in Table 5. The values on
 225 the diagonal of the matrix indicate the number of correctly classified cases. The
 226 confusion matrix was built by using simulated data of 100,000 cases.

227 **Table 5.** Confusion matrix of the Bayesian network estimated by PC Algorithm

Space heating costs		Predicted				
		0 - 99	100-199	200-299	300-399	400 or over
Actual	0 - 99	5463	1196	1033	1168	827
	100 - 199	5494	12876	1129	1148	767
	200 - 299	5495	4406	8459	2113	1244
	300 - 399	5518	2258	2425	11787	2299
	400 or over	5449	896	1650	3427	11473

228 When Table 5 is examined, it is seen that the correct estimation rate of the Bayesian
 229 network built by PC algorithm for the space heating costs is 0.50058 (50058/100000).

230 The Bayesian network model for the effects of the dwelling characteristics on space
 231 heating costs that is based on GTT algorithm is presented in Figure 5.

Figure 5. Bayesian network estimated by Greedy Thick Thinning algorithm

Table 6 shows the confusion matrix demonstrating the estimation performance of the Bayesian network that was constructed by GTT algorithm. The values on the diagonal of the matrix indicate the number of correctly classified cases. The confusion matrix was built by using simulated data of 100,000 cases.

Table 6. Confusion matrix of the Bayesian network estimated by Greedy Thick Thinning Algorithm

Space heating costs		Predicted				
		0 - 99	100-199	200-299	300-399	400 or over
Actual	0 - 99	0	655	136	35	117
	100 - 199	0	17027	1853	851	361
	200 - 299	0	11081	7630	5247	2551

	300 - 399	0	3958	3898	11735	6956
	400 or over	0	928	1547	4397	19037

When Table 6 is examined, the correct estimation rate of the Bayesian network built by GTT algorithm for the space heating costs appeared to be 0.55429 (55429/100000). However, the Bayesian network built by GTT algorithm could not estimate the space heating costs between 0-99 at all. Table 7 presents the accuracy rates of the RCBN and the other four Bayesian network models built by BS, PC and GTT algorithms.

Table 7. Accuracy rates for the space heating costs provided by the four different models

Accuracy Rates			
Rank Correlation Bayesian Network	Bayesian Search	PC	Greedy Thick Thinning
0.56871	0.47945	0.50058	0.55429

It can be seen in Table 7 that the RCBN model performed significantly better than the Bayesian network models built by BS and PC algorithms. However, although the accuracy rates of the RCBN model and the Bayesian network model given by GTT look close to each other, GTT algorithm could not estimate the space heating costs that are between 0-99 at all. Thus, the most successful model appears to be the RCBN.

2.5. Sensitivity analysis

In order to see how much space heating cost is affected by the changes in the levels of other nodes, it is possible to perform a sensitivity analysis on the Bayesian network estimated. Entropy Reduction (ER) can be used as a measure of sensitivity in Bayesian networks. The calculation of ER is as follows (Marcot et al., 2006).

$$I = H(Q) - H(Q|F) = \sum_q \sum_f \frac{P(q,f) \log_2[P(q,f)]}{P(q)P(f)} \quad (4)$$

where $H(Q)$ is the entropy of the node Q without any evidence, $H(Q|F)$ is the entropy of the node Q conditional to an evidence belonging to the node F and q, f are the states of Q and F respectively. Table 8 shows the ER scores of the nodes on the output node space heating cost obtained in Netica (2020) software.

Table 8. Results of sensitivity analysis for space heating cost node

Node	Entropy Reduction (ER)
Heating system	0.12479
External wall insulation	0.11879
Building age	0.10709
Number of bedrooms	0.06072
Dwelling size	0.04493
Double glazing	0.01937
Dwelling type	0.01844
Facing direction	0.01656
Floor level	0.01531

According to the results of the sensitivity analysis given in Table 8, the space heating costs of the dwellings are affected mostly by the heating systems with a level of 0.12479 entropy value. The second most important factor appears to be the existence of external wall insulation which has an entropy value of 0.11879. The third is the age of the building whose entropy value is 0.10709. It is also seen that the floor levels of the dwellings seem to have the least effect on the space heating costs with an entropy value of 0.01531

3. Results and Discussions

In Bayesian networks, it is possible to perform analyses and make estimations for any node by propagating the evidences through the network. For instance, for a flat type dwelling, on the third floor, heated by air conditioner or electric stove, having no double

glazing and external wall insulation, Figure 6 demonstrates the state of the RCBN model when these evidences are propagated through the network.

Figure 6. State of the RCBN model after propagation of the evidences

Figure 6 shows that the space heating cost of this dwelling is estimated to be 400 TL or over with the highest probability of 32.7% and between 300-399 TL with a probability of 20.4%. Continuing the propagation process for other combinations, various results provided by the estimated RCBN model are provided in the following lines.

As far as floor level is considered, the dwellings in basements and on ground floors seem to pay the highest space heating cost compared to the ones on the other floors. 29% of the dwellings in the basement pay 400 TL or more and 25.9% of the dwellings on the ground floor pay 300-399 TL per month. On the other hand, third and fourth floors seem to be paying less than the others. The third floors pay 100-199 TL with a probability of 25.9% the fourth floors pay 200-299 TL with a probability of 23.5%.

288 These results are consistent with the results given by Ziemele et al. (2015) who report
289 that the apartments located on the top and bottom floors of a building have more heat
290 loss than the other floors. When floor level factor is considered along with building age,
291 it is seen that the ground floors in new building pay 100-199 TL (38.3%) whereas in
292 buildings that are aged 30 or over, the ground floors pay 300-3999 TL (31.5%).
293 However, using coal stove in old buildings increases the space heating costs of the
294 ground floors to 400 TL or more level (44.4%). The Least space heating costs are paid
295 by residents living on the third floors in new buildings as 100-199 TL (49.8%). If
296 residents living on the third floors are living in old buildings and using air conditioners
297 or electric stoves for heating, their space heating costs increase to 400 TL or over class
298 with a probability of 29.2%

299 The size of the dwelling, as expected, increases the heating cost as it gets larger. The
300 dwellings between 0-59 m² pay 100-199 TL with the highest probability of 24.1% and
301 the ones between 60-119 m² pay 200-299 TL with the highest probability of 27.7%.
302 However, when the size becomes 120-179 m² the most popular cost becomes 300-399
303 TL with a probability of 28.5%. The residents, on the other hand, living in dwellings
304 having an area of 180 m² or over, the most likely space heating cost is 400 TL or over.
305 These results are also supported by the findings of Rehdanz (2007) stating that heating
306 costs decrease with the size of the property in Germany.

307 When the number of bedrooms is considered, dwellings having 3 bedrooms seem to pay
308 the highest amount of cost 300-399 TL with a probability of 29.3%. The dwellings
309 having only 1 living room, 1 bedroom and 2 bedrooms pay the least amount of space
310 heating costs 100-199 TL with probabilities 22.8%, 37.5% and 29.7% respectively.

The average monthly space heating costs of the dwellings having external wall insulation appears to be 100-199 TL with the highest probability of 37.1%, while the ones that do not have external wall insulation pay 400 TL or more on the average with the highest probability of 31.6%. Thus, it can be concluded that the lack of external wall insulation almost doubles the space heating costs. Indeed, in the sensitivity analysis, external wall insulation appeared to be the second most important factor affecting the space heating costs. This result is supported by many other studies such as Alvarez et al. (2016) who showed that insulation reduces energy costs in low income housings in Mexico, Cheung et al. (2015) who suggest that as well as the presence of the external insulation, the thickness of the insulation is also significant to reduce the energy consumption of the high rise-apartments in Hong Kong.

It was found out that building age is another important factor affecting the space heating costs. In fact, older buildings seem to pay higher space heating costs compared to the newer ones. This result is consistent with the findings in United Nations (2018) stating that although new buildings are more expensive to build, they are more energy-efficient compared to the old ones. Kazanasmaz et al. (2014) also support this result by reporting that older buildings are less energy-efficient due to the lack of insulation in Izmir, Turkey. In addition, Rehdanz (2007) also suggests that old buildings are more expensive to heat in Germany. The average amount space heating costs are 100-199 TL for the buildings 0-9 years old with the highest rate of 38.2%. However, for the buildings that are 30 years old or more, the average rate of cost is 400 TL or over with the highest rate of 34.4%.

When the heating system used in the dwelling is considered, it appeared to be the most important factor affecting the space heating costs. Coal stoves appear to be the costliest

heating systems. Dwellings heated by coal stoves spend 400 TL or more with the highest probability of 42.8%. Air conditioner and electric stoves also seem to be costly, since, 28.4% of the dwellings heated by these systems also pay 400 TL or more on the average. Central heating or boiler systems, however, are comparatively cheaper systems which lead to a cost 100-199 TL with the highest probability of 30.1%. These results are similar to the findings of Meier and Rehdanz (2010) in Great Britain. They state that space heating costs are highest if electricity is used while costs for gas are lowest. Rehdanz (2007) also found out that there is a strong effect of the kind of heating system on space heating costs in Germany.

Existence of the double glazing at dwellings slightly changes the space heating costs. When it is available, the proportion of the costs is distributed almost equally among 100-199 TL, 200-299 TL and 300-399 TL categories. However, when it is not available space heating costs sharply increase to the category of 400 TL or more with a probability of 37%. Cheung et al. (2005) also supports this result by stating that the thermal insulation performances of external walls are more effective than those for windows, in their study performed in Hong Kong.

The study suggests that there is a significant difference between the space heating costs of the detached houses and the flat type dwellings. Detached houses pay 400 TL or more on the average with the highest probability of 28.7%, while flats pay 100-199 TL with the highest rate 25.7%. This result is also consistent with the findings of Meier and Rehdanz (2010) in Great Britain. They found out that tenants, who mainly live in flats, pay less heating costs than owners, who mostly live in detached houses. This difference can be explained by the energy efficiency of the flats compared to the houses. On the other hand, Rehdanz (2007) suggests that, in Germany, owners, who mostly live in

houses, are likely to pay less space heating costs than tenants, who mostly live in flats. This is because the owners are more careful about the energy efficiency and insulation conditions of their houses.

As far as facing directions of the dwellings are considered dwellings facing to east and west and north pay space heating costs at similar rates. The most likely costs for east and west directions are 300-399 TL with the probabilities 24.9%, 28.3% and 28.4 respectively. However, the costs slightly change when the direction changes to south. Dwellings facing to south pay 100-199 TL with the highest probability of 27.6. Thus, facing directions of the dwellings do not appear to have a considerable amount of effect on the space heating costs. A study by Hassouneh et al. (2010) in Amman, Jordan, however, suggest that facing directions of the buildings are important when combined with the suitable kind of window glasses at the appropriate direction to reduce the costs of heating and cooling of the apartment type buildings.

4. Conclusions

In this study the construction and an application of the novel Rank Correlation Bayesian Network (RCBN) model is presented. Bayesian networks are built using two approaches in general. The first approach is structure learning. In structure learning the structure of the network is completely learned from the data through some algorithms. In the second approach, however, the structure of the network, that is, the number and the directions of the edges among the nodes are determined by an expert manually.

In the Bayesian networks that are constructed by learning algorithms, the directions among the edges do not have to imply a causal relation still, they represent a dependency relation. In the manually constructed networks, however, the directions of

the edges can imply causality, when appropriately directed. Causality is a complex and somehow a philosophical matter. However, dependency is a measure that can be mathematically calculated. Moreover, as it is known, correlation does not imply causality, on the other hand, it indicates a dependency between two variables. While developing the RCBN model, the point was the idea of building a novel model that benefits a dependency measure as well as the causality sense of a human being.

In general, mechanisms of the Bayesian networks depend on dependency relations among the nodes. Thus, when building a Bayesian network manually, an expert may have difficulty to decide about or determine the degrees of mathematical dependency among the nodes in the model. Thus, in the RCBN model the dependencies among the nodes are determined by using a rank correlation measure (Spearman's rank correlation coefficient) which can also be calculated for categorical variables. Then, an expert locates the edges between the node couples that have a dependency relation with each other, considering the direction of the cause-effect relations. For example, in the study, no arrow was directed from the node space heating cost as it is the affected variable. However, an arrow was directed from the node dwelling size to the node space heating cost, as the size of the dwelling is a factor affecting the space heating costs. At the stage of building the structural learning models, no restrictions were implied on the algorithms. In order to be able to make a fair comparison, the structural learning models were built as unrestricted models to avoid any manual intervention. Thus, the structural learning algorithms were just let build their models using their own mechanisms. Our research shows that the RCBN model performed better than the three different structural learning Bayesian network models, which were estimated by BS, PC and GTT algorithms. During the comparison of the RCBN model with the other structural

learning Bayesian network models, a simulated dataset of 100,000 cases were preferred for the testing data to include as many cases as possible.

The better performance of the RCBN model can be explained by the procedure applied while constructing it. In the RCBN model, the dependencies among the nodes are determined by rank correlation measures. Afterwards, between the node pairs that have significant correlations, directions of the edges are manually determined by expert opinion considering the causality relations and their directions. However, structural learning depends on various kinds of learning algorithms detecting the dependencies among the nodes directly from data with different calculation approaches. Thus, the directions of the edges do not necessarily imply a causality relation. Therefore, while structural learning algorithms can also detect the dependencies among the nodes, they miss the causal relations and the directions of the causalities among them. Hence, as the RCBM model include both dependency and causality relations as well as the directions of causalities within its network structure, it is more advantageous than the conventional structure learning algorithm-based Bayesian networks.

As far as the findings of the research related to the space heating costs are considered, it is seen that space heating costs are affected by various factors. In this study, the effects of various dwelling characteristics such as building age, heating system, dwelling type, number of bedrooms, dwelling size, floor level, facing direction, external wall insulation and double glazing on space heating costs were analyzed. In this analysis, both the amounts of costs that these factors cause and their levels of effect on the space heating costs were provided. The estimated RCBM model shows that the space heating costs of the dwellings are mostly affected by the heating systems used. When the heating systems are considered, coal stoves, air conditioners and electric stoves appear

to be the costliest heating systems. Hence, these systems should be replaced by central heating or boiler systems which are comparatively less costly. The second most important factor appears to be the existence of external wall insulation. In fact, the lack of external wall insulation almost doubles the space heating costs. Therefore, it is strongly recommended to use external wall insulation in this area. The third most important factor is the building age. The older buildings seem to pay more space heating costs compared to the new ones. Another finding is that existence of the double glazing at the dwelling slightly decreases the space heating costs. Additionally, the dwellings on the ground floors and the first floors appeared to pay the highest space heating cost. Therefore, dwellings on these floors need to be more effectively insulated. As the size of the dwelling increases the heating cost increases too. Another result is that facing directions and floor levels of the dwellings have the least effects on their space heating costs.

Compliance with Ethical Standards:

Funding: This study was not funded by anyone or any establishment.

Conflict of Interest: The sole author of this manuscript declares that he has no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by.

References

450 Alvarez, L. J., Munoz, N. R. & Domínguez, I. M. (2016). The Effects of roof and wall
 451 insulation on the energy costs of low-income housing in Mexico. *Sustainability*, 8(7),
 452 590.

453 Berthold, M.R., Cebron, N., Dill, F., Gabriel, T.R., Kötter, T., Meinl, T., Ohl, P., Sieb,
 454 C., Thiel, K. & Wiswedel, B. (2007). KNIME: The Konstanz Information Miner. In
 455 *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*.
 456 Springer.

457 Cheng, J., David A. B. & Weiru L. (1997). An Algorithm for Bayesian Belief Network
 458 Construction from Data. *Proceedings of AI & Statistics*, pages 83-90.

459 Cheung, C., Fuller, R., & Luther, M. (2005). Energy-efficient envelope design for high-
 460 rise apartments. *Energy and Buildings*, 37(1), 37-48

461 Eurostat, (2017).
 462 https://ec.europa.eu/eurostat/statisticsexplained/index.php/Energy_consumption_in_hou
 463 [seholds](https://ec.europa.eu/eurostat/statisticsexplained/index.php/Energy_consumption_in_hou)

464 GeNIe, (2020). BayesFusion, LLC. Retrieved from <http://www.bayesfusion.com/>

465 Hassouneh, K., Alshboul, A., & Al-Salaymeh, A. (2010). Influence of windows on the
 466 energy balance of apartment buildings in Amman. *Energy Conversion and*
 467 *Management*, 51(8), 1583-1591.

468 Heckerman, D., Chickering, D. M., Meek, C., Rounthwaite, R., & Kadie, C. (2000).
 469 Dependency networks for inference, collaborative filtering, and data visualization.
 470 *Journal of Machine Learning Research*, 1(Oct), 49-75.

471 Hill, D. R. (2015). Regional determinants of residential energy costs and the principal-
 472 agent problem in Austria. *Region*, 2(1), 1.

473 Holmes, D. E., & Jain, L. C. (2008). *Innovations in Bayesian networks: Theory and*
 474 *applications*. Berlin: Springer.

475 Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs*
 476 *February 8, 2007*. New York, NY: Springer New York.

477 Kazanasmaz, T., Uygun, I. E., Akkurt, G. G., Turhan, C., & Ekmen, K. E. (2014). On
 478 the relation between architectural considerations and heating energy performance of
 479 Turkish residential buildings in Izmir. *Energy and Buildings*, 72, 38-50.

480 Laureti, T., & Secondi, L. (2012). Determinants of Dwellings' Space Heating type and
 481 Costs in Italy. *Int. J. Environ. Res.*, 6(4), 1025-1038.

482 Longhi, S. (2015). Residential energy costs and the relevance of changes in dwelling
 483 circumstances. *Energy Economics*, 49, 440-450.

484 Marcot, B. G., Steventon, J. D., Sutherland, G. D., & Mccann, R. K. (2006). Guidelines
 485 for developing and updating Bayesian belief networks applied to ecological modeling
 486 and conservation. *Canadian Journal of Forest Research*, 36(12), 3063-3074.

487 Martinopoulos, G., Papakostas, K. T., & Papadopoulos, A. M. (2016). Comparative
 488 analysis of various heating systems for residential buildings in Mediterranean
 489 climate. *Energy and Buildings*, 124, 79-87.

490 Meier, H., & Rehdanz, K. (2010). Determinants of space heating costs in Great Britain.
 491 *Energy Economics*, 32(5), 949-959.

492 Nair, G., Gustavsson, L., & Mahapatra, K. (2010). Factors influencing energy efficiency
 493 investments in existing Swedish residential buildings. *Energy Policy*, 38(6), 2956-2963.

494 Netica, (2020). Norsys Software Corp. Retrieved from
 495 <https://www.norsys.com/netica.html>

496 Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential
 497 reasoning, *7th Conference of the Cognitive Science Society*, 15-17 August 1985,
 498 California: 329-334.

499 Pearl, J. (1998). *Bayesian networks*. Los Angeles, CA: Computer Science Dept.,
 500 University of California.

501 Rehdanz, K. (2007). Determinants of space heating costs in Germany. *Energy*
 502 *Economics*, 29(2), 167-182.

503 Rose, J., & Kragh, J. (2017). Distribution of heating costs in multi-story apartment
 504 buildings. *Energy Procedia*, 132, 1012-1017.

505 Risch, A., & Salmon, C. (2017). What matters in residential energy consumption:
 506 Evidence from France. *International Journal of Global Energy Issues*, 40(1/2), 79.

507 Schmitz, H., & Madlener, R. (2016). Heterogeneity in Price Responsiveness for
 508 Residential Space Heating in Germany. *SSRN Electronic Journal*.

509 Spirtes, P., Glymour, C., An Algorithm for Fast Recovery of Sparse Causal Graphs,
 510 *Social Science Computer Review*, 9(1), 62-72, 1991.

511 Sprent, P., & Smeeton, N. C. (2007). *Applied Nonparametric Statistical Methods*,
 512 *Fourth Edition*. Philadelphia, PA: CRC Press.

- 513 United Nations. (2008). *State of the World's Cities 2008/2009 Harmonious Cities*. New
514 York: United Nations, 2008.
- 515 Wessels, R. (2015). Improving the energetic building performance, renovation measures
516 and strategies for achieving zero-energy 1960's portico-flats. Delft University of
517 Technology, Department of Architecture 14th Graduation Studio Architectural
518 Engineering, June 16th, 2015.
- 519 Ziemele, J., Pakere, I., Blumberga, D., & Zogla, G. (2015). Economy of Heat Cost
520 Allocation in Apartment Buildings. *Energy Procedia*, 72, 87-94.
- 521 Zoric, J., & Hrovatin, N. (2012). Determinants of residential heating preferences in
522 Slovenia. *International Journal of Sustainable Economy*, 4(2), 181.

Figures

Figure 1

Example of Bayesian network

Figure 2

Rank Correlation Bayesian Network model for analysis of the effects of the dwelling characteristics on the space heating costs

Figure 3

Bayesian network estimated by Bayesian Search algorithm

Figure 4

Bayesian network estimated by PC algorithm

Figure 5

Bayesian network estimated by Greedy Thick Thinning algorithm

Figure 6

State of the RCBN model after propagation of the evidences