PAK Publishing Group
Growing Knowledge for Future

Proceedings Book of 2nd ICEFMO, 2014, Malaysia
Handbook on Economics, Finance and Management Outlooks
ISBN: 978-969-9952-06-7

A Study on Factors that Influence Monthly Household Expendıture Amongst Educators in Serı Iskandar, Perak

Nurul Aidha Mohd Tarimizi[1] --- Ong Hong Choon[2]

[1]*Quantitative Science Department, Kolej Profesional MARA Seri Iskandar, Perak and Malaysia*
[2]*School of Mathematical Sciences, Universiti Sains Malaysia, Penang and Malaysia*

ABSTRACT

In Malaysia, Household Expenditure Survey (HES) was conducted every five years to collect information on the level and pattern of consumption expenditure. Basically, monthly household expenditure represents the total outlay that a household has to make in order to satisfy needs and commitments incurred in a family. In 2011, Seri Iskandar had been chosen as the hub of education and commercial centre for Perak state. In conjunction with that, it is crucial to identify the factors which influence the household expenditure amongst the educators due to the rapid economics in Seri Iskandar. Bayesian Network approach is conducted in this study to analyse the causal relationship in monthly household expenditure and educators. Eight different structural learning algorithm are used in this study, which are Grow-Shrink, Incremental Association Markov Blanket, Fast Incremental Association, Interleaved Incremental Association, Hill-Climbing, Tabu Search, Max-Min Hill-Climbing and General 2-Phase Restricted Maximization. In this study, bnlearn package from R programming language is utilized in order to run all the eight structural learning algorithms. The network scores are used to identify which algorithm gives the best fitted network. Moreover, the arc strength is applied in the final network to determine the most influential relationship in this study.As a result, the network from Tabu Search algorithm is identified as the best final network in this study. Furthermore, the outcome shows that household size is identified as the main factor which influence the monthly expenditure amongst the educators in Seri Iskandar according to their gender.

Keywords: Household expenditure, Educators, Bayesian network

Contribution of Study

We highlight the issues of expenditure and economics literacy amongst the educators as a catalyst of creating awareness in satisfying needs and commitments. Educators are highly recommended to learn the basic knowledge of economics in order to manage their financial wisely. Educators should realise that aggressive developments in Seri Iskandar is going to influence their monthly household expenditure. The demands and liability of the household expenditure will rise up for sure. In conclusion, a proper financial planning and financial behaviour should be adopted for the educators in preventing them from financial problems in advance.

1. Introduction

1.1. Background on Monthly Household Expenditure

According to Lofquist *et al.* (2012), a "*household*" includes all members who occupy in a housing unit and one person in each household is designated as the "*householder*". Basically in business, expenditure refers to payment of cash for goods or services. Moreover, expenditure also refers as a charge against available funds in settlement of an obligation as evidence by using invoice, receipt, voucher or other documents. It has been reported (Income and Expenditure, 2012) that Household Expenditure Survey (HES) was first conducted in year 1957/58. However, beginning 1993/94, five years interval expenditure survey was carried out consistently to represent the current expenditure pattern of household in Malaysia. Besides, the information from the survey also contributes in determining the rate of change in prices of goods and services included in the basket of Consumer Price Index (CPI). From the expenditure trend, it was recorded that the household expenditure among Malaysian rose up to RM2190 from RM1953. It is also stated that Malaysia's population is 28.3 million and 2,258 428 are residents of Perak. Furthermore, Teacher Statistics (2013) revealed that there are 418,146 teachers around Malaysia with 69.27% being females and 30.73% being males. We decided to focus on Seri Iskandar area because it is known as the hub of education that provides carrier opportunities, family planning and salary. In addition, a rapid development in infrastructure and facility also plays a major influence on the household expenditure growth because of demographic factors, investment, savings and spending habits. We use Bayesian network to analyze the causal relationship of the monthly household expenditure amongst educators in Seri Iskandar, Perak.

1.2. Bayesian Network

A Bayesian network which also known as belief network or directed acyclic graphical (DAG) model is a probabilistic graphical representation of a multivariate joint probability distribution that exploits the dependency structure of distributions to describe them in a compact and natural manner (Pearl, 1988). Ge *et al*. (2010) used Bayesian network for determining the probabilistic relationship among set of variables. In graphical models, there are two types of structures namely as undirected and directed acyclic graph (DAG) to represent the relationship about an unknown domain. An undirected edge refers to direct probabilistic dependencies among the random variables where as DAG refers to the nodes that correspond to the variables in the domain. A Bayesian network which considers a finite set of *n* random variables of **X** = $\{X_1, X_2, ..., X_n\}$ is a pair of B = $\{G, \theta\}$ where G encodes each variable X_i is independent of its non-descendants given its parents in G while θ represents the set of parameters that quantifies in the network. Therefore, set B contains a parameter $\theta_{x_i|\pi_i} = P_B(x_i|\pi_i)$ for each realization x_i of X_i with conditioned of π_i.Thus, B defines as unique joint probability distribution over **X**, namely $P_B(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P_B(X_i|\pi_i) = \prod_{i=1}^{n} \theta_{X_i|\pi_i}$ where π_i represents the causes (parents) of variable X_i. Figure 1.1 shows an example of DAG consisting random variables X_1, X_2, X_3 and *Q*.

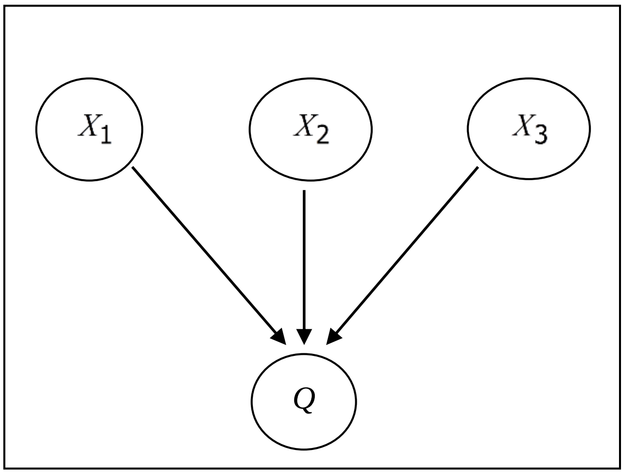


Figure-1. An example of DAG

Based on Figure 1.0, the random variables X_1, X_2, X_3 are said to be parents of Q which means outcome of X_1, X_2, X_3 influence the outcome of event Q based on the arrow from the nodes respectively. The direction of the arrow is useful information for decision maker in order to define the relationship of two random variables as conditional probability, $P(Q|X)$ where the probability for event Q to occur depends on the outcome of X.

1.3. Objectives of the Study

The aims of this study are

i) To investigate the factors that influence monthly household expenditure amongst the educators.
ii) To use Bayesian network to study the relationships between the factors.
iii) To create awareness about economics literacy amongst educators.

2. Literature Review

In this section, there are three major elements to cover which are household expenditure, educators' economics literacy and Bayesian network. Therefore, contents from several sources related to this study will be discussed in order to gather information about the three main components.

2.1. Review of Household Expenditure

According to Income and Expenditure (2012), graph of household expenditure trend 1993/94 to 2009/10 shows that housing, water, electricity, gas and other fuels were identified as the main contributors which surged by 15.1% to RM495 in 2009/10 as compared RM430 in 2004/05. This is then followed by food and non-alcoholic beverages which increased by 13.0% to RM444 from RM393, restaurants and hotel soared by 12.2% to RM239 from RM213 and transport was up by 4.1% to RM327 from RM314. Overall, average monthly households in Malaysia rose 12.1% to RM2190 in 2009/10 from RM1953 in 2004/05.

From Population Distribution and Basic Demographic Characteristics (2010), census in 2010 shows that the total of Malaysia's population is 28.3 million compared than 23.3 million in year 2000. It means that the average annual population growth rate for Malaysia is 17.67% from 2000 to 2010. It is also stated that 91.8% are citizens of Malaysia where it covers 67.4% for Bumiputera, 24.6% for Chinese, 7.3% for Indian and 0.7% for others. Besides, the Malaysia's population density rose up to 86 people per square kilometre in 2010 compared to 71 people in 2000. Specifically in Peninsular Malaysia, the Malays are the main ethic group with 63.1% which is equal to 17,857,300 people. Moreover, proportion of working age population from 15 to 64 years old is increased by 4.7% from 62.8% to 67.3%. This trend indicator shows the age structure of population aging in Malaysia. Population in Perak is stated as 2,258,428 people where 1,138,018 are male and 1,120,410 are female. In 2010, census shows that 35.1% are single where as 59.6% are married. Mean age of marriage for male is 28 years old and for female is 25.7 years old. Furthermore, the average household size is 4.31 in 2010 compared 4.62 in 2000.

Household Expenditure (2013) stated that household expenditure as the total consumption and non-consumption expenditure incurred for a family. It is said that household expenditure is supposed to satisfy the family needs and their legal commitments. Food expenditure is identified as the main contributor which covers 51% of the total, followed by transportation (11%), housing and utilities (10%) etc.

Jalleh (2011) reported in The Star that S.M Mohamed Idris from Consumers Association of Penang (CAP) revealed Malaysian households use almost half of their income to pay debts. The biggest portion of the Malaysian household expenditure goes to pay off housing loan, cars, personal use, securities purchase and credit cards. He is worried that families with high household debts would suffer from stress, depression, mental problems, suicides and family break-ups. Besides, Suhaila (2011) reported in Perak Today that Datuk Seri Dr Zambry Abdul Kadir confirm that Seri Iskandar is known as Perak's hub of education. In addition, Director of Dikir maju Sdn Bhd, Marcus Loh said that the location of Seri Iskandar is expected to change local economic landscape and offers employment opportunities to Perak citizens.

2.2. Reviews on Educators' Economics Literacy

In this era, economic stability contributes in affecting the educators' stress level based on their awareness about savings, spending habits, investment and economics literacy. Yunus *et al.* (2010) stated that economic literacy is vital because teachers as consumers also face problems of making choices in the market. Fontana & Abouserie (1993) did mention that educators of different grades, different counties and over different time period have all reported moderate to high level job stress. According to Wood and Doyle (2002), there is a significant relationship in between educated individual and economics literacy. It shows that, awareness of economics literacy amongst the educators is crucial in order to help them out in managing their household expenditure that can satisfy needs and commitments. Besides, it provides long term stability in terms of family planning, career opportunities and financial aspects. According to Walsh and Mitchell (2005), consumers are sometimes confused in making decision to buy and shop according to gender, age and educational level. Graham et al. (2002) stated that strategies in investment are influenced by gender factor. Yunus *et al*. (2010) refers that a number of teachers save for various purposes without even knowing the economics aspect as a whole. Furthermore, he manages to prove that expenditure is significantly correlated to economics literacy among the secondary school teachers in Perak. The result is consistent with a study done by Wood and Doyle (2002) which stated that teachers who have taken economics as subject in high school have more economics literacy compared than who didn't. He also recommended other researchers to stratify random sampling methods and expand the study.

According to Gorham *et al.* (1998), good financial behaviour is described by having effective behaviour in preparing financial record, documented cash flow, planning expenses, paying utility bills, controlling usage of money well in savings plan. Zaimah *et al.* (2013) identified statistics from the Ministry of Education (MOE) in 2012 shows that the total number of female teacher exceeds the total number of male teacher in Malaysia. In their study, they are inspired to investigate on financial behaviour among the female teachers in terms of age, education level, monthly income and level of financial knowledge.

Therefore, in this study, educators in Perak Tengah district were chosen as the respondents. Based on Perak Tengah District Council portal, the educational institution has been categorized based on their academic levels which are primary school, secondary school and higher level institution. Basically, there are fourteen primary schools, fifteen secondary schools and nine higher level institutions in Perak Tengah District. Since 2011, Seri Iskandar is rapidly developing economically and is one of education centres in Perak. Thus, this meets the demands of our study.

2.3. Reviews on Bayesian Network

Refer to Ben-Gal I (2007), Bayesian network is determined as *belief networks* (Bayes) where it belong to probabilistic graphical models (GMs) that represent the knowledge about an uncertain domain. Generally, GMs with undirected edges are known as *Markov random fields* or *Markov networks*. According to Heckerman (1995), Bayesian network is important because it can readily handle incomplete data sets, facilitate the combination of domain knowledge with data, provide knowledge about causal relationship and has efficient approach for avoiding over fitting of data. Cooper and Herskovits (1992) mentioned that they used Bayesian network in their study and calls it as *cases* where it can provide insight into probabilistic dependencies that exist among the variables in the data sets. This statement was supported by Friedman *et al.* (2000) who says that Bayesian network represents the dependence structure between multiple interacting quantities by using structural learning algorithm.

A study on factors of floating women's income in Jiangsu province was conducted by Ge *et al.* (2010) where researchers apply Bayesian network in socio economics field. They used 1757 samples aged in between 15 to 49 who migrated at least three months in Jianye. They considered 8 possible variables namely province, age, education, city, training experience, job, time and income. Furthermore, they adopted Bayesian network to identify the influence factors for floating women's in Jiangsu using six different structure learning algorithm which are Grow-Shrink (GS), Hill-Climbing (HC), Incremental Association Markov Blanket (IAMB), Fast Incremental Association (FAST.IAMB), Interleaved Incremental Association (INTER.IAMB) and Max-Min Parents and Children (MMPC). Based on the network scores, they identified Hill-Climbing as giving the best result and hence considered the final network for the study. In conjunction with the study, it is found that income of the floating women in Jiangsu province is influenced by the type of job which is different between cities, but not the education and training experience.

3. Methodology

In this chapter, we are going to explain about the data collection such as questionnaires, sampling design, Bayesian network and structural learning algorithms. Furthermore, we will also discuss briefly on network scores as a performance measuring for the algorithm.

3.1. Data Set

The sample for this study is randomly selected among the educators Perak Tengah District through questionnaires. There are total of 525 samples. In this study, we measure the characteristics of interest by using 15 variables and all the details are presented in Table 3.1.

Table-3.1. 15 variables of the data

Variable	Possible values	Description
Gender	2	2 types: Male and Female
Marital Status	3	3 types: Single, Married and Widowed
Household Size	4	3 groups: 1-4, 5-8 and 9-12
Race	4	4 types: Malays, Chinese, Indian and Others
Age	3	3 groups: 16-30, 31-45, 45-60
Sector	3	3 types: Government, Private and Statutory
Experience	3	3 groups: <3 years, 3-7 years, >7 years
Students' Academic level	2	2 levels: Post - SPM, School
List	2	2 types: Yes, No
Frequency	3	3 groups: <4 times, 4-6 times, >6 times
Buy	4	4 types: Grocery Store, Supermarket, Hypermarket and Others
Focus	6	6 types: Food & Beverage Housing, Water, Electricity, Gas & Other Fuels Jewellery Communication and Technology Restaurant & Hotels Others
Choose	6	6 categories: Price/cost, Brand, Fashion/Style, Quantity, Promotion, Others
Income	3	3 groups: <RM3000, RM3000-RM5000 and >RM5000
Expenditure	3	3 groups: <RM2000, RM2000-RM4000 and >RM4000

3.1.1. Questionnaires Design

In this study, we construct self-administered questionnaires because it requires low budget and high response rate. Furthermore, we categorized it to three types of measurement namely nominal scale, interval scale and ratio scale. Nominal scale data describe variables in term of its category and differs in term of quality such as marital status, gender and race. Interval scale data measures the variables such as household size and age. However, ratio scale measures variables such as income and expenditure. The questionnaire was designed to focus on demographic characteristics and expenditure style among the educators in Seri Iskandar.

Based on Table 3.1, out of 525 respondents, there are 28% are male educators and 72% are female educators. 94.1% are Malays, 4% are Chinese and 1.7% are Indians. The ages of the respondents from this study consist of 37.90% from 16 to 30 years old, 60.4% from 31 to 45 years old and 1.7% around 45 to 60 years old. Percentage of single educators in Seri Iskandar is 25.9% compared than 73.5% who are married and 0.6% widowed. Household size is 84.2% for one to four, 15.6% for five to eight and 0.02% for nine to twelve. There are 73.7% working with the government, 14.3% in private sector and 12.0% in statutory sector. Besides that, 41.1% teaching post-SPM student and 58.9% teaching secondary, primary and pre-school. 20.6% of the educators claim that they do have the list of items they want to buy while 79.4% don't. From the survey, we found out that 60.2% of educators choose their major expenditure as food and beverage while 32.4% for housing, water, electricity gas and other fuels. 66.3% of the educators choose to buy their needs in hypermarket while 19.6% go to grocery store, 9.1% go to supermarket and 5% go to others. 53.5% educators choose their items based on price/cost where else 22.1% quantity, 9% brand, 7.2% promotion, 2.9% fashion/style and 5.3% others. An open ended question is provided for the respondent to answer for average household income and expenditure.

3.1.2. Sampling Design

In order to fulfil the requirement of this study, a total 525 respondents were selected from Seri Iskandar educators. Stratified sampling is used as the process of selecting samples that represents each stratum in a population. According to Income and Expenditure (2012), population in Perak is stated as 2,258,428 people where 1,138,018 are male and 1,120,410 are female. However, according to Teacher Statistics (2013), the total numbers of female teachers is 289,631 which are approximately 70% of the total number of all teachers in Malaysia and given the total of educators in Perak is 413, 759. Therefore, the percentage of educators in Perak is set to be 18% where the number of educators divided by the population in Perak. We identify 38 educational institutions from different levels in Perak Tengah District but we focus on educators in Seri Iskandar area only by including those who are teaching in primary school, secondary school, higher level education, tuition centre, Islamic school and pre-school. Figure 1.1 shows the stratified sampling has apply in this research.

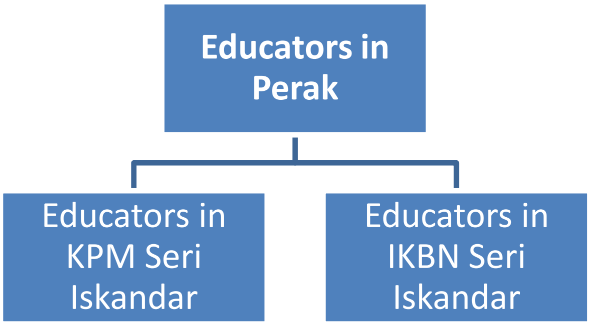


Figure 1.1. Example of stratified sampling of study

3.2. Structure Learning Algorithms

In order to reduce the complexity of data while running the analysis of this study, several related learning algorithms for Bayesian network were applied. Structure learning algorithm can be categorized into three main components which are constrained –based algorithm (Cooper, 1997; Margaritis, 2003), score-based structure algorithms (Singh & Valtorta, 1995; Margaritis, 2003) and hybrid structure algorithm (Acid and De Compos, 1996).

Scutari & Brogini (2010) mentioned a Bayesian network B = $\{G, \theta\}$ as a graphical model represented by a Directed Acyclic Graph (DAG) as G = $\{A, E\}$ where each node represents the random variable for $X \in A$ and the arcs in E specify the conditional independence structure of A. Learning Bayesian network is also performed in two steps with structure learning and parameter learning. On structure learning we need to find a graph structure that encodes the conditional independence (CI) relationship in the data. On the other hand, on parameter learning is where we need to fit the parameters of each local distribution given the graph structure in the first step.

3.2.1. Constrained-Based Structure Algorithms

According to Dash & Druzdzel (2003), constraint-based searches a database for independence relationships and constructs graphical structures called "*pattern*" which represent a class of statistically indistinguishable DAG. Cooper (1997), suggested that a Bayesian independence test as part of an

approximate constraint-based learning algorithm. However, constraint based is sensitive to errors in the individual tests.
The following structure learning algorithms are classified under constraint-based structure algorithm:

A. Grow-Shrink (GS)

According to Margaritis (2003), GS is considered as the simplest Markov blanket detection in structure learning algorithm. It consists of two phases which are a grow phase and shrink phase. In GS, growing phase of a variable X will continue by trying to add each variable Y to the current set of hypothesized neighbours of X, PX. Then, PX grow during every iteration at the same time variable X is added if and only if Y is dependent on X given the current set of hypothesized neighbours of PX. However, false positive situation occurs when some of the variables in PX are not true neighbours of X at the end of grow phase due to unspecified ordering on the variables. Hence, each false positive of PX is determined by testing the independence with X conditioned on PX - {Y}.

B. Incremental Association Markov Blanket (IAMB)

Tsamardinos *et al.* (2006) said that IAMB is based on a two phase selection scheme and follows the same two phase structure with GS algorithm. Moreover, IAMB adopts one dynamic heuristic in the growing phase in order to enhance the static and inefficiencies heuristic of GS structure. IAMB iteratively reorders the variables when a new variable enters the Markov blanket at the same time reordering operation is applied using mutual information heuristic.

C. Fast Incremental Association Markov Blanket (FAST.IAMB)

FAST.IAMB is considered similar to GS and IAMB since it is two phase structure: grow and shrink phase. It reduces the number of conditional independence test by using speculative stepwise forward selection (Yaramakala and Margiritis, 2005).

D. Interleaved Incremental Association (INTER.IAMB)

INTER.IAMB structural algorithm interleaves the growing phase and shrinking phase attempting to keep the size of Markov blanket as small as possible during all steps of the algorithm's execution. Aliferis *et al.* (2003) said that this algorithm used a forward stepwise selection which avoids the false positive in Markov Blanket.

3.2.2. Score-Based Structure Algorithms

Score-based structure algorithms addresses learning as a model of selection problem, reflects how well a structure matches the data and at the same time search for the best network by looking at the highest score (Na & Yang, 2010). Chickering (1995) also mention that score function is usually score equivalent since the same score is assigned for the same probability distributions that the network defined.
There are two categories classified under score-based structural learning algorithms which are:

A. Hill-Climbing (HC)

HC is identified as the common score-based algorithm on the space of a directed graph due to its ease in implementation and quality of obtained output that generates locally optimal solution (Gamez *et al*., 2010).

B. Tabu Search (TABU)

Scutari (2010) stated that TABU algorithm is a modified HC which able to escape local optima by choosing a network that minimally decreases the score function.

3.2.3. Hybrid Structure Algorithms

Hybrid algorithm is a combination of constraint-based and score-based algorithms where it refers to conditional independence test and network scores at the same time (Scutari, 2013).

A. Max-Min Hill-Climbing (MMHC)

Tsamadinos *et al.* (2006) briefly explained that MMHC first learn the skeleton of a Bayesian network using Max-Min Parent Child (MMPC) to restrict the search space. Then, it orients the skeleton using greedy hill-climbing search to figure out the optimal network structure in the restricted space.

B. General 2-Phase Restricted Maximization (RSMAX2)

Scutari (2010) said that general implementation of MMHC algorithm can be found in RSMAX2 algorithm.

3.3. Network Scores

Ge *et al.* (2010) proposed that a scoring function Score (G, D) for Bayesian network structure is decomposable and formally can be expressed as:

$$Score\ (G, D) = \sum_{i=1}^{m} S(D_i, D_{G_i})$$

where G is a directed acyclic graph(DAG) and D is considered as certain data set for $i = 1, 2, .., m$. Scutari (2013) mentioned that available network scores can be categorized into two categories which are discrete case (multinomial distribution) and continuous case (multivariate normal distributions). However, in this study we focus on discrete case where we use five different methods to find scores which are a multinomial log-likelihood (loglik), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Bayesian Dirichlet Equivalent (BDE) and logarithm of K2 score (K2).

A. Log-likelihood (loglik)

Log-likelihood (loglik) score is equivalent to the entropy measure used in Weka (Witten & Frank, 2005). According to Witten *et al.* (1999), Weka refer as Waikato Environment for Knowledge Analysis is a comprehensive group of Java class libraries that implements many state of the art machine learning and data mining algorithms.

B. Akaike Information Criterion (AIC)

Akaike (1973) tells that AIC is used to choose the model that able to minimize the negative likelihood penalized by the number of parameters as shown in equation (1).

$$AIC = -2 \log p(L) + 2p \quad \cdots\cdots\cdots (1)$$

Where *L* represents the likelihood under the fitted model and *p* is the number of parameters in the model.

C. Bayesian Information Criterion (BIC)

Generally, BIC is derived within a Bayesian framework as an estimate of the Bayes factor for two competing models (Schwarz, 1978; Jensen, 2009). The score of the BIC can be defined as equation (2)

$$BIC = -2 \log p(L) + \log n \quad \cdots\cdots\cdots (2)$$

where *n* is a sample size.

D. Bayesian Dirichlet Equivalent (BDE)

BDE was developed by Heckermen *et al.* (1995). BDE score uses Bayesian analysis to evaluate and estimate data set in a network.

E. Logarithm of K2 score (K2)

K2 score is proposed by Cooper and Herskovits (1992). It is considered another Dirichlet posterior density. According to Chen *et al*. (2006), K2 attempts to choose the network structure that maximizes the network's posterior probability given the experimental data. The K2- like greedy search method will incrementally adds a node to a parent set and finds the best parent set to maximize the joint probability of the structure and the database (Yang *et al*., 2006). K2 is different from BDE because it does not compare an equivalent score.

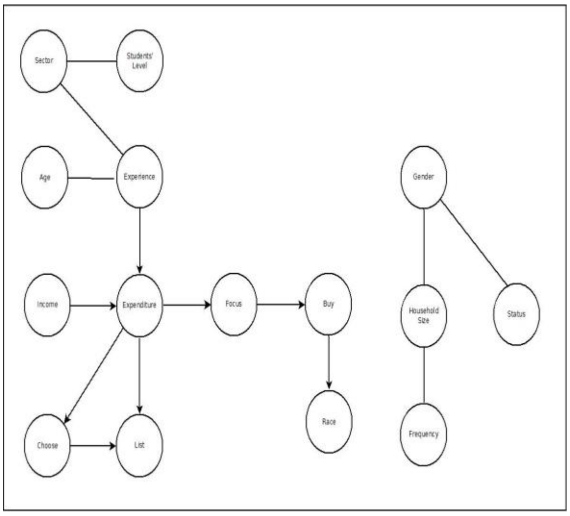
3.4. An Overview of Performance Measure

In the previous components of this study, the network scores are briefly explained in order to have a better understanding of estimating the fitting of all algorithms used. By using the five different score methods, we can measures the network performances and examine the possible final network. The structural learning algorithm which possessed the best network scores will be selected as the final network for this study. Furthermore, the arcs strength is used to determine the strength of relationship among two variables in the final network. Hence, we can determine the strength of causal relationship in between two linked nodes. Finally, we will be able to identify the factors that influence the study.

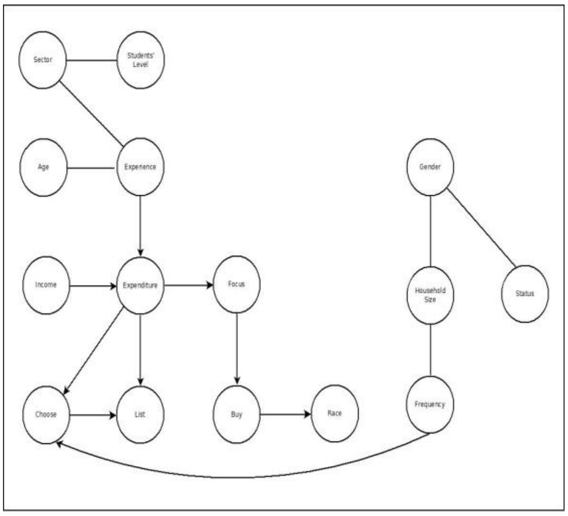
4. Results and Discussion

4.1 Result of Bayesian Networks

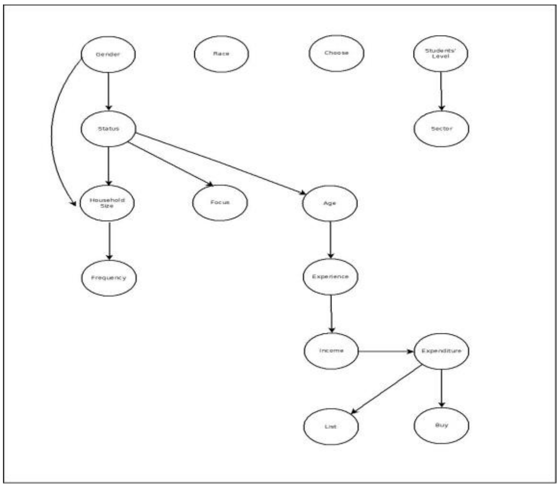
Similar to Ge *et al*. (2010), the *bnlearn* package in R language is used to run the structural learning algorithms. From the structural learning algorithms, there are eight different networks outcomes are analyzed. These eight networks are from Hill-Climbing (HC), Grow- Shrink (GS), Incremental Association Markov Blanket (IAMB), Fast Incremental Association Markov Blanket (FAST.IAMB), Interleaved Incremental Association Markov Blanket (INTER.IAMB), Max - Min Hill Climbing (MMHC), Tabu Search (TABU) and Restricted Maximization (RSMAX2) are shown in Figure 4.1. Basically as we discussed in the previous section, the arcs represent direct dependent relationships between the connection variables but the existence of conditional independence relationships of the absence of arcs gave meaning to the network (Ge *et al*. 2010). Besides, these diagrams also provide the logical causal effect between the variables. Based on the chosen structural learning algorithm, we can evaluate the pattern of the network that leads us to next process of analyzing the data.



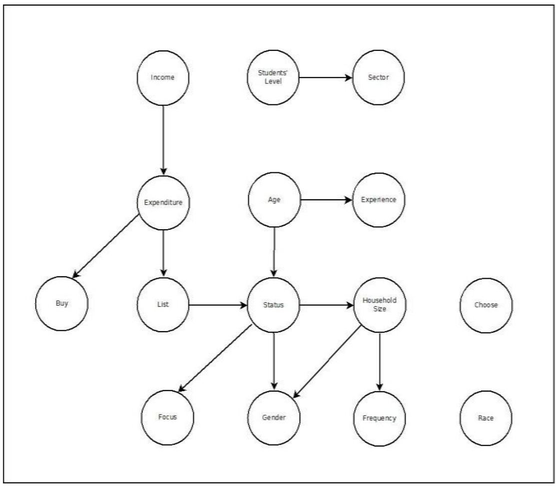
(a) IAMB



(b) INTER.IAMB



(c) HC



(d) TABU

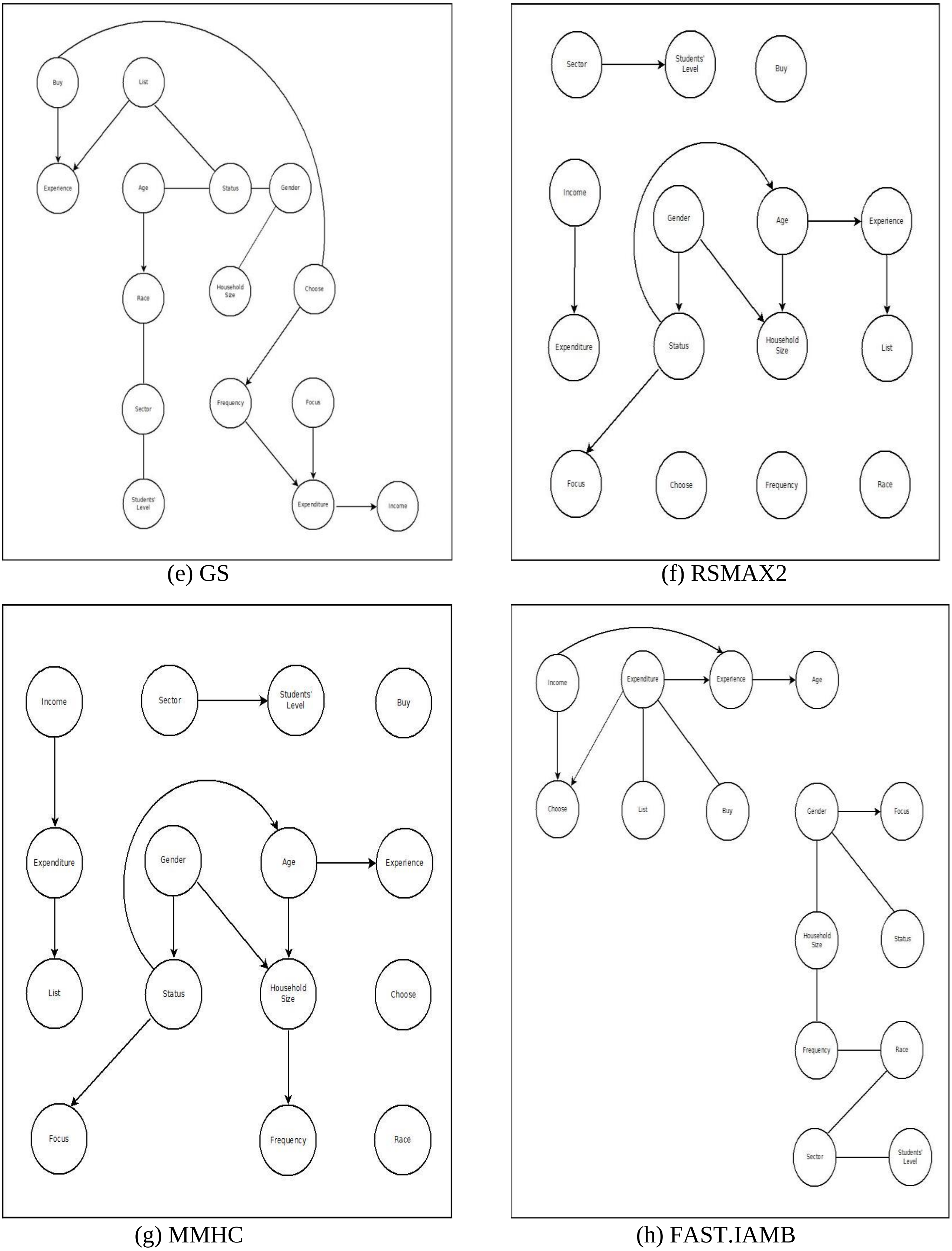


Figure-4.1.: Network structures learned by selected algorithms. (a) Incremental Association Markov Blanket (b) Interleaved Incremental Association Markov Blanket; (c) Hill – Climbing; (d) Tabu Search; (e) Grow – Shrink; (f) General 2-Phase Restricted Maximization; (g) Max – Min Hill Climbing; (h) Fast-Incremental Association Markov Blanket

4.1.1. Common Links and Arcs of the Learning Algorithm

Generally, the number of links and arcs of Figure 4.1 are displays in Table 4.1. Here, we indicate the edge as the number of common links that exists by omitting its direction. Moreover, we refer the arcs as the common number for links with similar direction exists. Table 4.1 shows the common edges/arcs that produces mutual relationship in the network

Table-4.1. Number of common links, arcs and edges between each pairs of the learned networks

	TABU	GS	HC	IAMB	FAST. IAMB	INTER. IAMB	MMHC	RSMA X2
TABU	13/13	5/0	13/10	7/2	8/0	8/4	9/5	5/2
GS	-	14/7	5/0	6/0	3/0	4/0	5/1	5/0
HC	-	-	12/12	7/3	5/2	6/3	9/9	5/5
IAMB	-	-	-	14/8	5/0	14/8	7/3	4/0
FAST.IAMB	-	-	-	-	14/6	8/1	4/2	3/0
INTER.IAMB	-	-	-	-	-	15/9	7/2	4/1
MMHC	-	-	-	-	-	-	10/10	5/5
RSMAX2	-	-	-	-	-	-	-	6/6

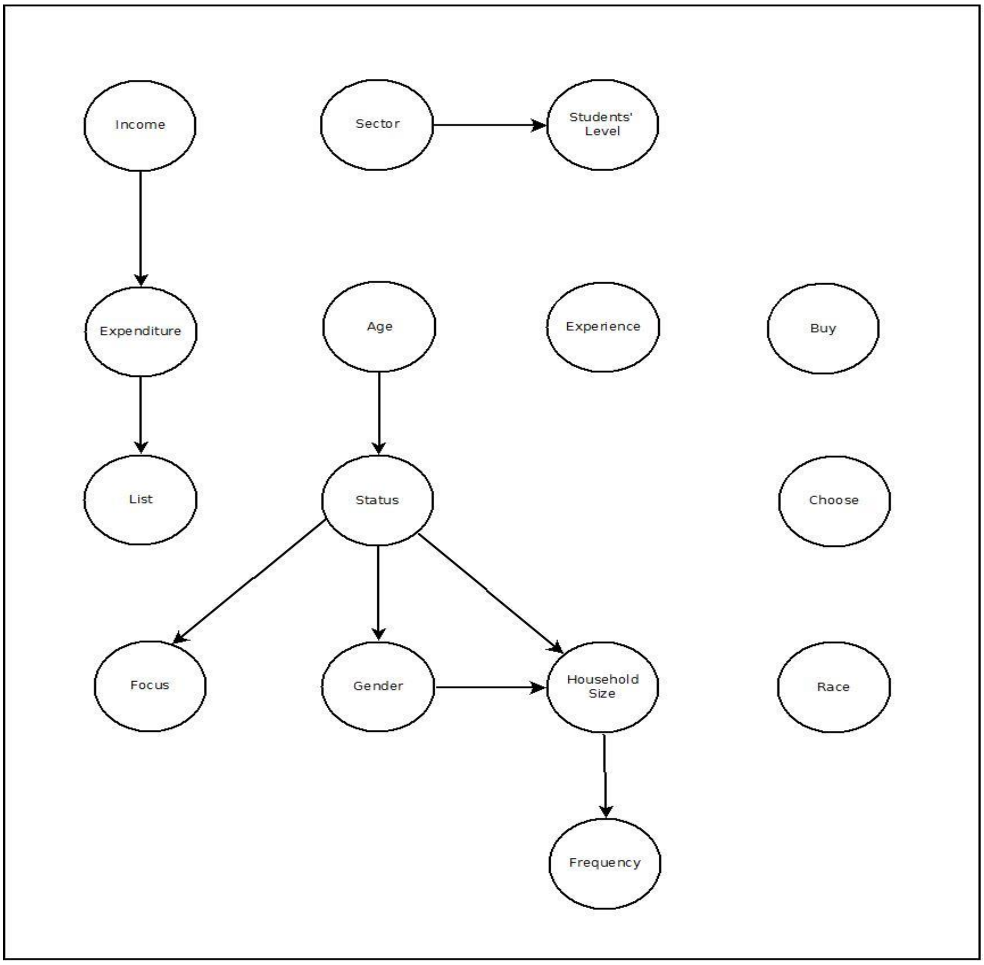
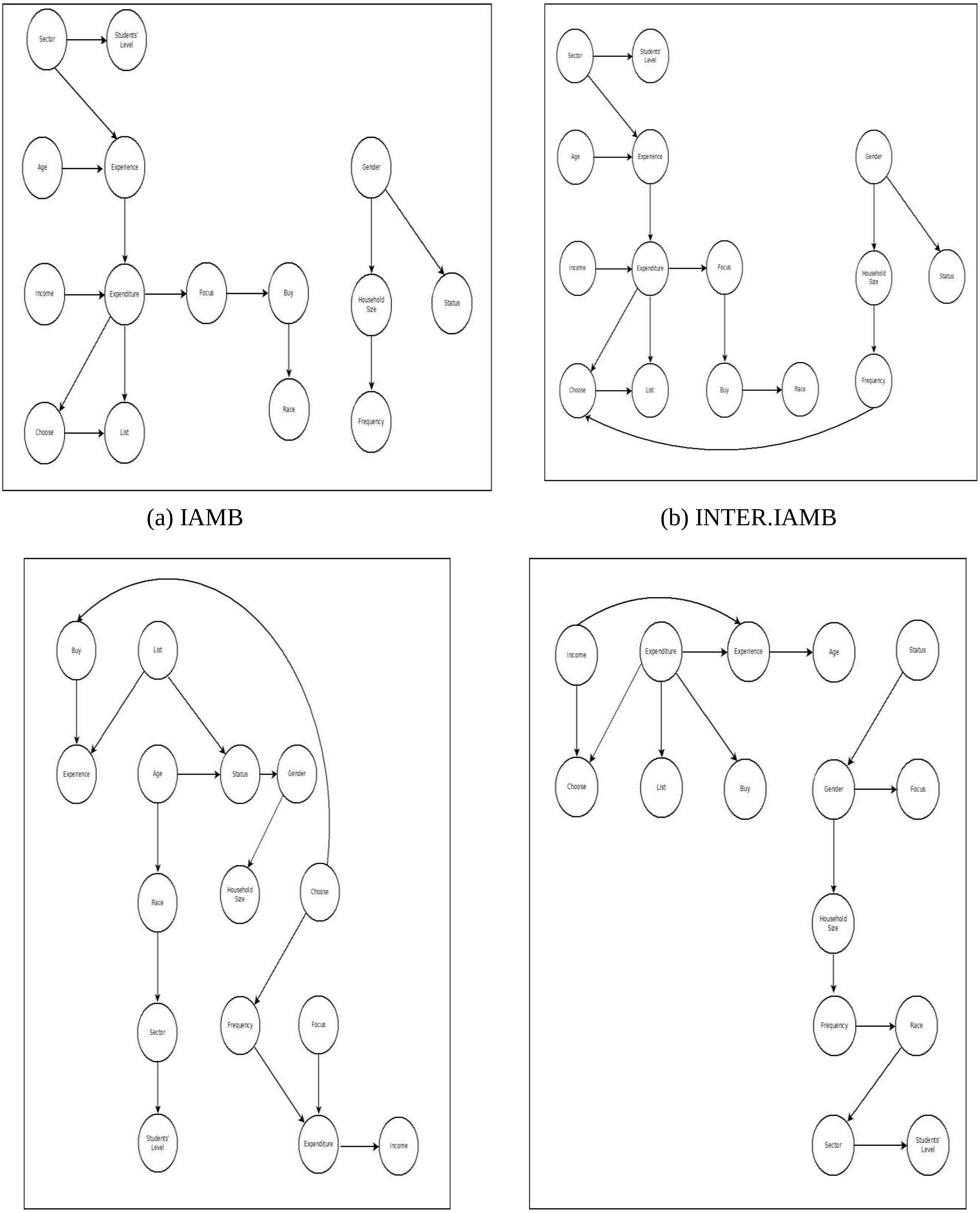


Figure-4.2. Common edges of all the learned networks

As shown in Figure 4.2, there are nine common edges which are with directions and it can be identified from Figure 4.1. Next, we are going to run the structural learning algorithms again and whitelist the common edge.

4.1.2. Whitelisted Structure Learning Algorithm



(a) IAMB

(b) INTER.IAMB

(c) GS

(d) FAST.IAM

Figure 4.3. Network structures learned by selected algorithms after whitelist (choose direction). (a) Incremental Association Markov Blanket; (b) Interleaved Incremental Association Markov Blanket; (c) Grow – Shrink; (d) Fast Incremental Association Markov Blanket

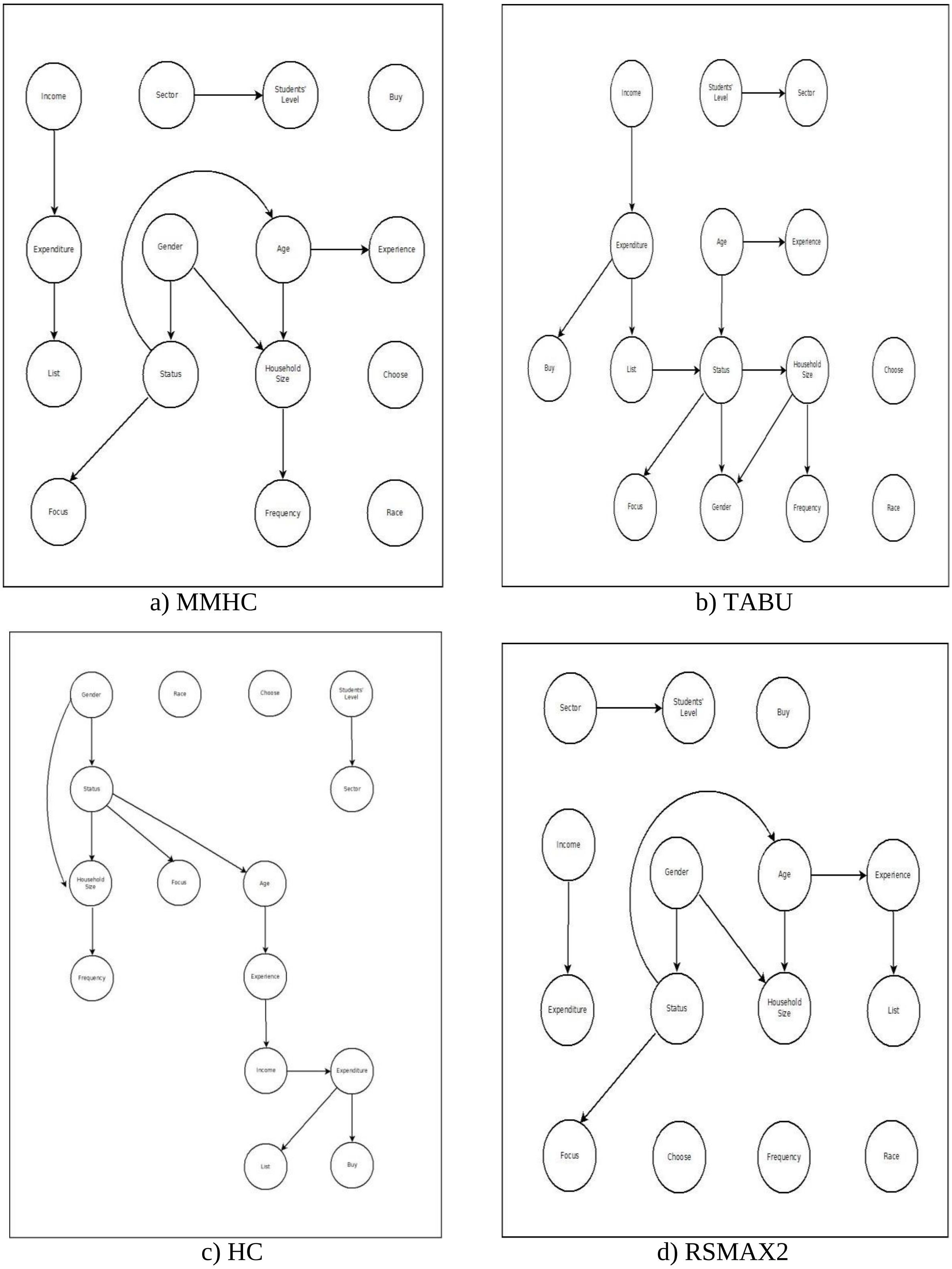


Figure 4.4. Network structures learned by selected algorithms after whitelist (without choosing direction). (a) Max – Min Hill Climbing; (b) Tabu Search; (c) Hill – Climbing; (d) General 2-Phase Restricted Maximization

Figure 4.3 shows four network structures that we have to choose the direction of the arrows by depending on the p-value from the R language. The p-value is used to determine which direction of two variables is better than another. Figure 4.4 shows another four network structures that remain the same after whitelist since we do have direction for all two related variables.

Table-4.2. shows the network scores of each structural learning algorithm using five different scoring functions. The result is vital for the study since it provides the knowledge for choosing the best fitted network.

Table-4.2. The result of scores of all learned networks for each algorithm

	bde	k2	loglik	aic	bic
GS	-5718.205	-5716.777	-5379.971	-5554.971	-5928.018
HC	-5606.452	-5591.215	-5389.450	-5630.591	-5630.591
IAMB	-5660.635	-5682.311	**-5328.051**	**-5492.051**	-5841.650
FAST.IAMB	-5651.561	-5659.292	-5353.982	-5499.982	-5811.210
INTER.IAMB	-5660.635	-5682.311	**-5328.051**	**-5492.051**	-5841.650
MMHC	-5651.835	-5629.92	-5456.675	-5660.235	-5660.235
TABU	**-5591.273**	**-5578.834**	-5367.508	-5627.439	**-5627.439**
RSMAX2	-5801.779	-5776.422	-5629.636	-5783.089	-5783.089

Table 4.2 shows that Tabu Search is the best to suite this study and have been chosen as the final network. Based on the network scores, Tabu Search produce the best fitted network. Hence, we proceed to find the arcs strength for the Tabu search network as shown in Table 4.3.

Table-4.3. The scores of arc strength of Tabu Search Network

RANK	FROM	TO	ARC STRENGTH
1	Household Size	Gender	-0.1519647
2	List	Status	-3.1522503
3	Expenditure	Buy	-4.6954296
4	Household Size	Frequency	-11.201775
5	Status	Focus	-13.5796364
6	Status	Gender	-18.5381907
7	Status	Household Size	-18.6126895
8	Age	Status	-26.724894
9	Experience	Income	-27.9844693
10	Expenditure	List	-29.1225328
11	Age	Experience	-71.101278
12	Income	Expenditure	-113.455402
13	Student's Level	Sector	-139.8591115

In Figure 4.5, the solid thicker lines implies the stronger relationships among the edges which have the higher scores compared with the others. The solid thin lines represent the supplementary edges which still related to the network according to their rank. The dotted lines represent the weakest relationship among the edges.

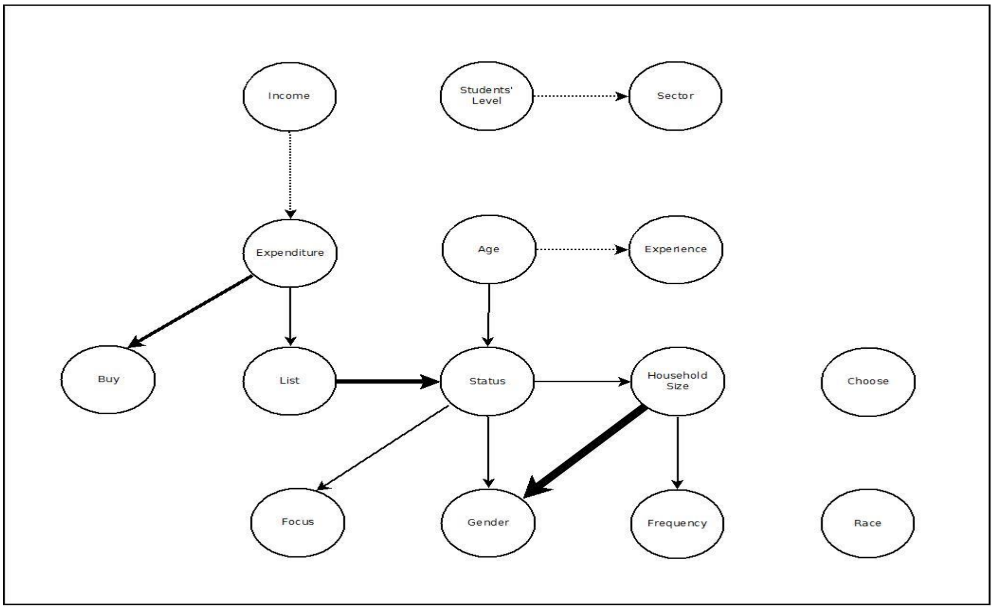


Figure 4.5. The final result of the score learned network.

4.2. Discussion

From the final network, we identified that educators' household size strongly linked to their gender. Mok *et al.* (2011) confirms that the average household size in Malaysia is 4.3 persons and about 56% of them were small size households (less than 5 persons), 37% were medium size (5-7 persons) and 8% were large (8 or more persons). Moreover, Teacher Statistics (2013) stated approximately 70% of the teachers in Malaysia are female and the rest are male. It is aligned with the result for this study where we identify that the educators' household size is one to four (small size household) and female educators outnumbered the male educators by 44%. We also found that 73.5% of the educators are married and they share the household expenditure with their spouse in order to fulfill their family needs and commitments. Considering that the educator and spouse as working parents, they have to plan the size of the family depend on their financial capability. Due to rapid economic development in Seri Iskandar, female educators are necessary to be a contributor or head in the household expenditure since they generate a consistent income every month. In conjunction with that, gender is influenced directly by the household size. This is because the liabilities incurred in the family become the priority of concern amongst the educators in order to provide a better accommodation, promote higher education, technology literate and paying utility bills on time. Joo & Grable (2004) stated that a financial well-being is closely related to their financial behaviour, income financial knowledge, levels of financial stress, liquidation of cash, financial tolerance and education.

According to Population Distribution and Basic Demographic Characteristics Report (2010), census shows that Perak state has a slow population growth rate which equal to 1.4% of Malaysia's population. Family planning and economic upturn are the factors that slow down the population growth rate in Perak. The aggressive development in Seri Iskandar, Perak in past two years also contributes to catapult the educators' household expenditure in a month. Loh and Hew (2012) mentioned that Tony Ng, manager of Hua Yang Bhd said that the price of a single-storey house costs around RM120,000 – RM180,000 and double-storey house costs about RM180,000 – RM220,000. Mathematically, on average, the educators have to pay around RM600 – RM1300 for housing loan to their monthly household expenditure.

List of expenditure is also strongly associated to marital status of Seri Iskandar educators. Marital status plays a role in influencing the educators' expenditure in terms of spending habits, savings planning, buying power and future investment. The need of listing expenditure demands is crucial in order to avoid financial problems such as over spending, debt and stress which is not good for the family institutions. According to Zaimah *et al.* (2013), surprisingly from 90% of married respondents in this study only 40% of them had carefully drawn their financial budget, reviewed and assessed their expenditure and consistently carried out financial planning.

Ironically, list of expenditure related to the expenditure estimation where it is depends on customers' choices. We found an association in between expenditure estimation and a favourite shopping venue is quite strong. In this study, 66.3% of the respondents which is equal to 348 people choose to shop in hypermarket such as Tesco, Giant, Jusco etc. The educators agreed that hypermarket offers comfortable facilities, good quality product, appropriate price and promotions compared than the other shopping venue. Hamzah (2011) reported that Hua Wen Yan (Hua Yang Bhd Chief Officer) officially announced that Tesco will be opening in Seri Iskandar, Perak. The superstore can cater to 10,000 populations in that area. Ngo (2013) reported that Datuk seri Dr Zambry Abdul Kadir, Menteri Besar Perak said the state government is set to develop a new education hub in Seri Iskandar, covering an area of more than 404.6ha focusing on tertiary education, which includes government, semi-government or private universities and colleges. This is the paradigm shift for Seri Iskandar economy that is going to inflict monthly household expenditure amongst the educators in Seri Iskandar, Perak. Therefore, it is very important to create awareness on economics literacy among the educators to provide wise economics decisions and effective financial planning.

In this study, the average household income is minimally related to expenditure estimation due to the salary factors. However, if there are no awareness pervaded it might cause the educators end up in the financial problem. Besides, it can lead to inflation in the future. Students' academic level taught by the educators is weakly correlated with sector educators' work. Monthly household expenditure is slightly affected according to students' academic level either higher level educational institutions or school under government, private and statutory.

5. Conclusion

5.1 Summary

A study on factors that influence the monthly household expenditure amongst the educators in Seri Iskandar is conducted due to the concern of rapid economics in the area. Nowadays, Seri Iskandar is formally known as the education hub and commercial centre of Perak. Therefore, it is very crucial for the educators to identify and realise about factors that influence their monthly household expenditure.

Ge *et al.*(2010) mentioned that Bayesian network is a powerful tool to explore the potential relationship between the variables of complex social problems. Bayesian network allows us to find the relationships among variables using structural learning algorithms compared the network scores and determine the strongest arcs strength between the edges. As a result, we manage to run the analysis and gather all information about the factors that influence the monthly expenditure amongst the educators.

In this study, we highlight the issues of expenditure and economics literacy amongst the educators as a catalyst for creating awareness in satisfying needs and commitments. Educators are highly recommended to learn the basic knowledge of economics in order to manage their financial wisely. Educators should realise that aggressive developments in Seri Iskandar is going to influence their monthly household expenditure. The demands and liability of the household expenditure will rise up for sure. In conclusion, a proper financial planning and financial behaviour should be adopted for the educators in preventing them from financial problems in advance.

5.2. Future Works

The study will continued by expanding the range of sampling area and choosing different groups of respondents instead of focusing on educators only. Besides, methods of collecting data using questionnaires will be not considered due to time constraint and flexibility to the researcher. Moreover, updating more structural learning algorithms and network score methods in the future should be taken into consideration in order so that it increases the data reliability and validity.

References

Acid, S. and De Campos, L.M. (1996). An Algorithm for Finding Minimum d-Separating Sets in Belief Networks. In: Proceeding of 12th Conference on Uncertainty in Artificial Intelligence. DECSAI Technical Report , 3 – 10.

Akaike, H. (1973). Information Theory and an Extension of the MaximumLikelihood Principle. In: B.N. Petrov and F. Csaki (eds.) 2ndInternational Symposium on Information Theory:267-281.Budapest:AkademiaiKiado.

Aliferis, C., Tsamardinos, I., Statnikov A. (2003) HITON, A Novel Markov

Blanket Algorithm for Optimal Variable Selection, Vanderbilt University, DSL 03-08

Ben-Gal, I. (2007). Bayesian Networks, in Ruggeri F., Faltin F &Kenett R., Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons.

Chen, X.W., Anantha, G. and Wang, X. (2006). An effective structure learning method for constructing gene networks.Bioinformatics. 22(11), 1367 – 1374.

Chickering, D. M. (1995). A transformational characterization of equivalent Bayesian network structure: In UAI '95: Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence, 18-20 August 1995, Canada. San Francisco: Morgan Kaufmann.

Cooper, G.F. (1997). A Simple Constraint-Based Algorithm for Efficiently Mining Observational Databases for Causal Relationships.DataMining and Knowledge Discovery, 1, 2203– 2224.

Cooper, G. F. and Herskovits, E. (1992). A bayesian method for the induction of probabilistic networks from data. MachineLearning, 9, 309 – 347.

Dash, D. and Drudzel, M. J.,(2003). Robust independence testing for constraint-based learning of causal structure.UAI 2003, 167-174.

Friedman, N., Linial, M. & Nachman, I. (2000). Using Bayesian network to analyze expression data. Journal of Computational Biology. 7, 601-620.

Fontana D and Abouserie R. (1993). Stress level, genderand personality factors in teachers.The British Journal of Educational Psychology.Jun;63(pt2),261-70

Gamez, J. A., Mateo, H. L. and Puerta, J. M (2010). Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood, Data Min Knowl Disc (2011) 22,106–148.

Ge, Y., Li, C. and Yin, Q. (2010). Study on Factors of Floating Women's Income inJiangsu Province Based on Bayesian Networks.Advances in Neural Network Research and Applications, Lecture Notes in Electrical Engineering 2010. 67(9), 819–827.

Gorham,E. E., DeVaney, S. A. & Bechman, J. C. (1998). Adoption of financial management practices: a program assessment. Journal of Extension. Vol.36, No, 2.

Graham J.,Stendardi E., Myers J. & Graham M.(2002). Gender differences in investment strategies: an information processing perspective. International Journal of Bank Marketing, 20(1), 17-26.

Hamzah, S. R. (2011). Tesco to open 80000 sq ft in Hua Yang 's Bandar Universiti township. [Online]. [Accessed 24th October 2012]. Available from Malaysian Reserve Online: http://themalaysianreserve.com

Heckerman, D.(1995). A tutorial on learning with Bayesian networks.Technical Report MSR-TR-95-06, Microsoft Research, Redmond, Washington.

Heckerman, D., Geiger, D. and Chickering, D.M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learning, 20 (3), 197–243.

Household Expenditure (2013) [Online],[Accessed 12th May 2013]. Available from Department of Statistics Tonga Portal: http://www.spc.int/prism/tonga

Income and Expenditure (2012) [Online],[Accessed 24th October 2012]. Available from Department of Statistics Malaysia Portal: http://www.statistics.gov.my/portal

Jalleh, J. (2011). Malaysian households use almost half their income to pay off debts [Online], [Accessed 30th May 2013]. Available from The Star Online: http://thestar.com.my

Jensen, F.V. (2009). Bayesian Networks. Wiley Interdisciplinary Reviews: Computational Statistics. 1(3), 307 – 315.

Joo, S., & Grable, J. E. (2004). An exploratory framework of the determinants of financial satisfaction. Journal of Family and Issues, 25(1), 25-50.

Lofquist, D.,Lugaila, T.,O'Cornell, M. & Feliz, S. (2012).Household and families 2010. U.S Census Bureau.U.S Department of Commerce. Economics and Statistics Administration..Suitland MD.

Loh, I. and Hew, C. (2012). Fast expanding Seri Iskandar continue to attract investors. [Online], [Accessed: 5th May 2013]. Available from The Star Online: http://thestar.com.my

Margaritis, D. (2003). Learning Bayesian Network Model Structure from Data.Ph.D.thesis, School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA.

Mok, T. P., Maclean, G. and Dalziel, P.(2011).Household size economies: Malaysian evidence .Lincoln University, Lincoln, New Zealand.

Na, Y. and Yang, J. (2010). Distributed Bayesian Network Structure Learning. International Symposium On Industrial Electronics(ISIE), 2010 IEEE: 1607-1611.

Ngo, E. (2013). New education hub in Seri Iskandar [Online], [Accessed 30th May 2013]. Available from The Star Online: http://thestar.com.my

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, SanMateo, CA.

Population Distribution and Basic Demographic Characteristics Report (2010) [Online], [Accessed on 24th October 2012].Available from Department of Statistics Malaysia Portal: http://www.statistics.gov.my/portal

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of statistics. 6(2), 461-464.

Scutari, M. (2010). Learning Bayesian network with bnlearn R package. Journal of Statistical Software, 35(3), 1-22.

Scutari, M. and Brogini, A.(2010). Constraint-based Bayesian Network Learning with Permutation Tests. Department of Statistical Sciences, University of Padova.

Scutari, M. (2013). Bayesian network structure learning, parameter learning and inference Learning Bayesian network with bnlearn R package. [Online], [Accessed 3rd April2013]. Available from bnlearn package: http://www.bnlearn.com

Singh, M. and Valtorta, M. (1995).Construction of Bayesian network structures from data: a brief survey and an efficient algorithm. International Journal of Approximate Reasoning.111-131

Suhaila (2011). Sri Iskandar Bussiness Centre Bakal Jadi Mercu Tanda Terbaru Daerah Perak Tengah [Online], [Accessed 30th May 2013]. Available from Perak Today Online: http://www.peraktoday.com

Teachers Statistics (2013) [Online],[Accessed 5th May 2013]. Available from Ministry of Education Malaysia Portal: http://www.moe.gov.my

Tsamardinos, I., Brown, L.E., and Aliferis, C.F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. Machine Learning. 65, 31–78.

Walsh G. & Mitchell, V.(2005). Demographic characteristics of consumers who find it difficult to decide.Marketing Intelligence & Planning, 23(3), 281-295

Wood, W., & Doyle, J.(2002). Economic literacy among corporate employees. Journal of Economics Education, 195-205

Witten, I. H. and Frank E.(2005). Data mining: practical machine learning tools and techniques, 2nd edn. Morgan Kauffman, San Francisco.

Witten, I. H.,Frank, E.,Trigg, L.,Hall, M., Holmes, G. and Cunningham, S.J (1999). Weka: Practical machine learning tools and techniques with Java Implementations. In: Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information System, 192-196.

Yang, Y. L.,Wu, Y. and Liu,S. Y.(2006).The constructing of Bayesian networks based on intelligent optimization. In: Proceedings of the Sixth International Conference on Intelligent Systems design and Applications (ISDA'06).1,371-376.

Yaramakala, S. and Margiritis, D.,(2005). Speculative Markov Blanket discovery for optimal features selection. In Proceeding of the Fifth IEEE International Conference on data Mining(ICDM), 2005.809-812.

Yunus, K.Y., Ishak, S., Jalil, N.A. (2010). Economic literacy amongst the secondary school teachers in Perak Malaysia. Information Management and Business Review, 1(2): 69-78.

Zaimah, R., Sarmila, M. S., N. Lyndon,., Azima A. M., S. Selvadurai, Suhana, S., Er A. C.(2013). Financial behaviors of female teachers in Malaysia, Asian Social Science, 9(8).