Decision Analytics (Topics)

– SEMINAR SUMMER SEMESTER 2015 –

Bayesian Network Learning

– SEMINAR PAPER –

Submitted by:
Nicole Ludwig

Advisor:
Prof. Dr. Dirk Neumann

Contents

1 Introduction

Nowadays available data, which can be used to find solution to specific problems or relationships between different variables of interest, is massive. In a large dataset, with a huge number of variables which might be interacting, getting an overview over those variables and their interactions can be difficult. Although one could calculate dependences and independences for each and every combination of variables using tables or so, this does not seem to be a good approach when the data becomes big. To be able to graphically and mathematically represent dependences between a large amount of variables and afterwards ask specific queries, we can use Bayesian networks. Bayesian networks are graphical models based on a directed acyclic graph (DAG). Graphical models in general describe how variables interact locally and help us express conditional independence assumptions through the estimation of local interactions. If we chain the local interactions together we get global, indirect interactions. Instead of using Bayesian networks we could just use full joint distribution tables as probabilistic models. The problem with those, which make it a good idea to use the Bayesian networks, is the fact that they become way too big to represent explicitly if there are many variables and they are hard to learn and estimate. Thus Bayesian networks are a good technique to describe complex joint distributions (models) using simple local distributions (conditional probabilities) and hence save space and time.

An economical problem, where the use of Bayesian networks seems reasonable is the phenomenon of undepricing in initial public offerings (IPO). There exist several empirical studies showing that during a short period at the beginning of the share being trade-able, the initial public achieve above average returns, maybe indicating that the initial offering might have been too low, thus underpriced (Ruud 1993). There are many theories and models that try to explain why we have this underpricing. A more recent approach is it to include textual information, e. g. the tone in the offering (see e. g. Feuerriegel et al. (2014), into the analysis. We have thus a sizeable amount of information which could help to either explain or predict underpricing of initial public offerings.

We thus introduce to the concept of Bayesian network learning and the application of this concept to initial public offerings in this paper. The remainder is structured as follows, Section 2 gives an introduction to initial public offerings and underpricing in general, while Section 3 introduces to Bayesian networks in general. Thereafter, Section 4 applies Bayesian network learning to information about initial public offerings and discuses the results.

2 Initial Public Offerings

An *Initial Public Offering* (IPO) is generally the initial sale of shares by a corporation to the overall public. Where the main aim often is the provision of equity to the general public, initially with the purpose of obtaining funds. In this section we will first give some general information about IPOs and then address the phenomenon of *underpricing*.

2.1 General Information on IPOs

In the following we have a look at the several steps that lead to an Initial Public Offering in the U.S. First, after the company has decided to go public, they have to engage an underwriter, who generally issues securities. The company then hands in a so called prospectus to the Securities and Exchange Commission (SEC). This prospectus is supposed to help investors form a decision. It thus also includes the offer terms. The anticipated offer price is given as a price range from a minimum to a maximum price, where the midpoint of this range is expected to be the offer price (Hanley 1993). After having filed this preliminary prospectus the "waiting period" follows. This period, which lasts until the final offer date, the underwriter gathers information from regular investors if they are interested in buying a share. We define regular investors after Hanley (1993) as "*those that are actively involved on an ongoing basis in purchasing shares of firms going public*". From these interest indications the underwriter assesses the expected demand and adjusts the final offer price. Hence, increasing the offer price for higher than expected demand and decreasing the offer price for lower than expected demand.

2.2 The Concept of Underpricing

A phenomenon which is often encountered and is the main interest of the application in this paper, is the phenomenon of underpricing. In general, underpricing is the positive deviation of the stock price at the first trade from the initial offer price. Thus, the assessed value of the stock during the waiting period is lower, sometimes by a large amount, than the price of the stock at the first trading day. Interestingly, underpricing often results in the stock falling significantly the following days after the introduction of the company to the stock market. As Ruud (1993) states "*while hardly a cause for complaint from investors, such underpricing might hurt emerging firms trying to raise capital expansion. The high average initial return on new-issue shares is therefore an anomaly that invites further study*". The reasons for underpricing are still an ongoing research topic, nevertheless, many explain underpricing with information asymmetries and state that underpricing of Initial Public Offerings is done on purpose. In the following we give some important relationships between underpricing and other IPO related variables, which can be found in literature:

- Underpricing is highest (lowest) when offer prices are increased (decreased) and positive (negative) information is revealed during the waiting period (Hanley 1993).
- The percentage change in the NASDAQ index from the filing of the preliminary prospectus to the offer date is positively and significantly related to underpricing (Hanley 1993).
- There is an equilibrium relation between the expected underpricing of an initial public offering and the ex ante uncertainty about its value(Beatty and Ritter 1986)
- Managers strategically underprice IPOs to maximize personal wealth from selling shares at lock-up expiration. First-day underpricing generates information momentum by attracting attention to the stock and thereby shifting the demand for the stock outwards. This allows managers to sell shared at the lock-up expiration prices higher than they would otherwise obtain (Aggarwal et al. 2002).

3 Bayesian Networks

To be able to apply the concept of Bayesian networks to our real world problem, we first have to establish what exactly a Bayesian network is and how we can use it correctly. As Bayesian networks are based on some fundamental concepts in probability theory, we first revise the corresponding basics, before we explain Bayesian networks themselves and their learning process.

3.1 Probability Theory

As the Bayesian networks utilize the fundamental concepts of Bayesian statistics, this subsection is concerned with revising the most important ones in our context. The first, an possibly most important one is the Bayes' Theorem which describes an events probability conditional on other related events.

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x)$$
$$\rightarrow P(x|y) = \frac{P(y|x)}{P(y)}P(x).$$

This concept is useful for us, as it might be difficult to find the probability of a cause given an effect, but it is way easier in most cases to observe the probability of an effect given a cause

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}.$$

Another important concept is independence and conditional independence. Especially the latter one will be of great importance for the learning of a Bayesian network later.

Definition: Two variables are independent

$$P(x) : X \perp\!\!\!\perp Y$$

if: $\forall x,y : P(x,y) = P(x)P(y) \rightarrow P(x|y)$.

And as the absolute independence might be a too strong assumption, we also need the conditional independence

Definition: Two variables X and Y are conditionally independent

$$X \perp\!\!\!\perp Y|Z$$

iff $\forall x,y,z : P(x,y|z) = P(x|z)P(y|z)$ or iff $\forall x,y,z : P(x|z,y) = Px|z)$.

Bayesian networks, or graphical networks in general help us to express conditional independence assumptions locally. To get indirect global interactions we can thus chain the local interactions together with the help of the chain rule. In general, a Bayesian network implicitly encodes joint distributions. The probability of a full assignment then becomes

$$P(x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} P(x_i|parents(x_i)).$$

We can proof that this is a proper joint distributions with the chain rule and the assumption of conditional independence. After having repeated some of the fundamental concepts that are used in the construction of Bayesian networks, we can move on to the general structure and characteristics of such networks in the next section.

3.2 Structure of Bayesian Networks

In general, Bayesian networks are graphical models which help us to represent a probability distribution over a set of random variables $X = \{X_1,X_2,\ldots,X_p\}$. They consist of two parts (Margaritis 2003):

1. The topology of the graph, a *directed acyclic graph* (DAG) $G = (V,A)$, with each node representing a random variable, and
2. a set of *local probability distributions* (local pdfs), conditional on each value combination of the parents.

An important quantity, that we need for constructing our Bayesian network, is the concept of a Markov blanket. Figure 1 shows a Markov blanket with the following formal definition.

Definition: The *Markov blanket* of a node $A \in V$ is the minimal subset S of V such that

$$A \perp\!\!\!\perp_P V - S - A|S.$$

Or in an informal way, the Markov blanket of a node is the minimal set of nodes that must be observed to make this node independent of all other nodes. In any Bayesian network (directed model), the Markov blanket of a node A is the set of the parents of A, the children of A, and all the other nodes sharing a child with A.

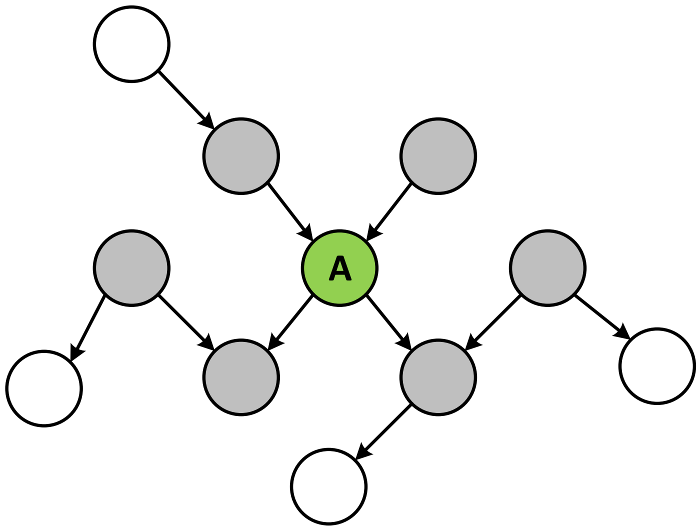


Figure 1: Illustration of the Markov blanket of node *A*. Where all grey nodes are part of the Markov blanket, while all white nodes are not.

The structure of the Bayesian network generally describes independences, thus in its mathematical sense there do not have to be causal relationships through the edges and their direction. Nevertheless, if certain assumptions are fulfilled we can use the Bayesian networks as causal models, thus have a direct causality from the parent node to the child node. Margaritis (2003) describes the two conditions as follows.

1. Whether or not there are any common unobserved causes (variables) of two or more observed variables in the domain. If there are no such variables, the property of *causal sufficiency* is said to hold true. Unobserved variables are also called latent or hidden variables.
2. Given causal sufficiency, whether or not it is possible for more than one network structure to fit the constraints that have been observed in the domain. These constraints are statistical independences that are observed in the data for the purposes of this paper. Only one of these networks can be the "“true"” underlying generative model that embodies the real cause-effect relationships that govern the data-generating mechanisms of the domain.

3.3 Structure Learning

Having established what the basic structure of the Bayesian network is, we need to find ways to learn or build those networks. Learning a Bayesian network can be divided in two steps. The first step is to learn the structure of the network. This network structure limits the possibilities of data which can be fitted to it, as "“*the structure implies a specific set of conditional independence relationships among the variables involved*"” (Margaritis 2003). The various algorithms, which have been proposed for this structure learning process, can be divided into three categories: *constraint-based*, *score-based* and *hybrid* algorithms. Additionally, the network can also be build by hand, using specific expert knowledge and prior information on the data to create the structure. In this paper we have a look at Bayesian network learning algorithms and do not try to build the model on expert knowledge.

3.3.1 Constraint-based Algorithms

Constraint-based algorithms use, as the name already indicates, constraints for the structure learning process. Typically, those constraints are conditional independence statements (Margaritis 2003). In practice, the structure is constructed with the help of statistical independence tests on the data. But in order for the resulting structure to be valid the assumptions of causal sufficiency, causal Markov and faithfulness have to hold.

An example constraint-based algorithm is the Incremental Association Markov Blanket Algorithm (IAMB) of Tsamardinos et al. (2003). The algorithm can be separated in two phases. A forward and a backward phase. The algorithm starts with an empty candidate set for the estimate of the MB(T), the CMB. the forward phase includes adding those variables to this set that maximize a heuristic function $f(X;T|CMB)$. Typically this heuristic function is some measure of association between X and T given CMB. For the set of candidates to be as small as possible after the forward phase, the heuristic function needs to be informative and effective. Tsamardinos et al. (2003) uses Mutual Information as heuristic. Where mutual information is defined as

$$\mathrm{MI}(X,Y|Z) = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{L}\frac{n_{ijk}}{n}\log\frac{n_{ijk}n_{++k}}{ni+kn_{+jk}},$$

with the conditional probability tables that are implied by the observed frequencies $\{n_{ijk}, i = 1,\dots,R, j = 1,\dots,C, k = 1\dots,L\}$ for the variables X,Y and the conditioning variable Z (Nagarajan et al. 2013). The second phase, the backward conditioning, removes those features from the CMB that do not belong to the $MB(T)$ through an independence test, testing if a feature X in the CMB is independent from $T|CMB_{-X}$. A proof of correctness and several variations of the IAMB algorithm can be found in Tsamardinos et al. (2003).

3.3.2 Score-based Algorithms

Another popular method to learn the Bayesian network structure from data is the use of score-based algorithms. The leaning process involves the calculation of a score for each possible network. This should typically be a measure for how well the structure of the network fits the given data. Assuming a structure G, the score of a network can be written as being equal to the posterior probability of G given the data set D (Margaritis 2003)

$$Score(G,D) = Pr(G|D) = \frac{Pr(D|G)Pr(G)}{Pr(D)}.$$

The aim of the score-based algorithms is it to maximize this probability.

A commonly used score is the BIC score, which was introduced by Schwarz (1978)

$$BICscore(G,D) = \log Pr(D,\hat{p},G) - \frac{d}{2}logN,$$

where $\hat{p}$ is the set of maximum-likelihood estimates, d the number of dimensions of p and N the size of the data set. The main advantage of this score is the fact that we do not need any prior information to obtain it. This makes it especially valuable in cases, where there is no prior information obtainable.

In the following we will discuss a popular score-based algorithm known as *hill-climbing* algorithm. The hill-climbing algorithm can start with an empty, full or random network. The first step is typically a maximum-likelihood estimation procedure for all the parameters of the local pdfs given a network structure. This step is followed by a loop which adds, removes or reverses single-edges, such that the current candidate is the network with the highest score. The loop stops if the score cannot be improved further through a change in a single-edge. As there is no guarantee that the algorithm finds the global optimum, one can use different methods to make the arrival at a global optimum more likely, e. g. permutation, random restarts or simulated annealing (Margaritis 2003).

In addition to using either the constraint-based or the score-based algorithms, one can also combine the two approaches to get a hybrid algorithm. After having learned the structure and the parameters of the network, we can compute probabilities, which we are interested in, from the model. This is also referred to as *probabilistic inference* (Heckerman 1995). Interestingly, exact, as well as approximate, inference in any Bayesian network are NP-hard (Cooper 1990). We still try to use inference now for our data set on initial public offering.

4 Results

In this section we discuss the results found from learning different Bayesian networks from our data. We first explain the dataset underlying our analysis more precisely. Afterwards we learn the Bayesian networks. we first consider the case of continuous Bayesian networks, before we have a look at learning the network from discrete data. We compare different algorithms in both cases and compare them. The resulting networks are then used to answer specific queries and give information about our data. All analyses are done with the help of the statistical software R.

4.1 Dataset

The dataset used for the analysis is extracted from the Kennedy-Patton IPO database and includes all initial public offerings filed with the SEC between 2003 and 2010. It also includes the corresponding final prospectus for every initial public offering. The news corpus we utilize is the Thomas Reuters News Archive for Machine Readable News. A detailed description of the corpus and the pre-processing steps for the final dataset can befound in Feuerriegel et al. (2014). For out further analysis we consider twelve different variables, which are described below (most are taken from Feuerriegel et al. (2014))

- *ShareOverhang* A variable defined as in Feuerriegel et al. (2014)

$$ShareOverhang = \frac{S_{Outstand}}{S_{Outstand} + S_{Sold}},$$

  where $S_{Outstand}$ is the number of shares left after the offer and S_{Sold} is the number of shares which have been sold to the public.

- *NewsSentiment* A measure that contains information about the sentiment in the news corpus. More precisely, it gives information about the uncertainty of the words.
- *NewsCount* The number of news announcements in the period.

- *Prior1M* Number of newspaper articles in the period.
- *ProspectusSentiment* A measure that contains information about the sentiment in the prospectus itself. More precisely, it gives information about the uncertainty of the words.
- *UpRevision* A control variable that measure how much the final offer price deviates in percentage from the expected price in the prospectus.
- $\log \Delta Days_{SEC,IPO}$ The log difference in calendar days between the filing of the prospectus to the SEC and the IPO.
- $NASDAQ_{-15days}$ To control for the general stock market development that might be unrelated to the IPO, we include the return of the last 15 days before the IPO.
- *lnSales* This variable is defined as

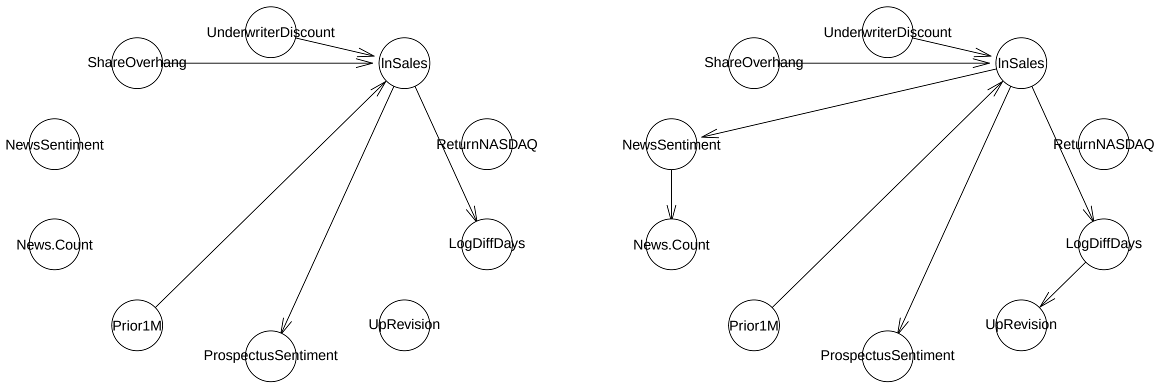
$$\log Sales = \log N * p,$$

where N is the number of sold shares in the offering and p is the intial price of the shares. The variable hence controls for the volume of the initial public offering.
- *UnderwriterDiscount* A variable including per share discounts and commissions of the underwriters.
- *YearDummy* A dummy variable for the year of the initial public offering to control for seasonal effects.
- *SectorDummy* A dummy variable for the sector in which the IPO took place, to control for sector-specific events.

4.2 Learning a Continuous Data Bayesian Network

The dataset we have, consists of mostly continuous data, only the two dummy variables are discrete factors. We therefore start learning a continuous Bayesian network, leaving the two different dummies aside for the moment. As we discussed above we first have to learn the structure of a network before we can learn the corresponding parameters and start asking specific queries. Finding the structure of our Bayesian network can be quite difficult and highly depends on the learning algorithm and the tests or scores we use. To illustrate this we will construct Bayesian networks from our data with the help of a constraint-based and a score-based algorithm. Namely, we utilize the incremental association Markov blanket algorithm as constraint-based alternative, and the hill-climbing algorithm as score-based alternative. Additionally, we have a look at how the algorithms perform with different independence tests and scores, alternatively.

The first algorithm we use is the above described incremental association Markov blanket algorithm. In Figure 2 we can compare the network constructed with this algorithm, depending on the test score, which is either the exact Student's t-test or the mutual information based Monte Carlo permutation test. The network resulting form the t-test consists of 6 connected nodes, while the one constructed with the mutual information test consists of 9 nodes. The mutual information based network adds additional relationships between *lnSales* to $NewsSentiment$ and $News.Count$, while also making $UpRevision$ a child of $LogDiffDays$.

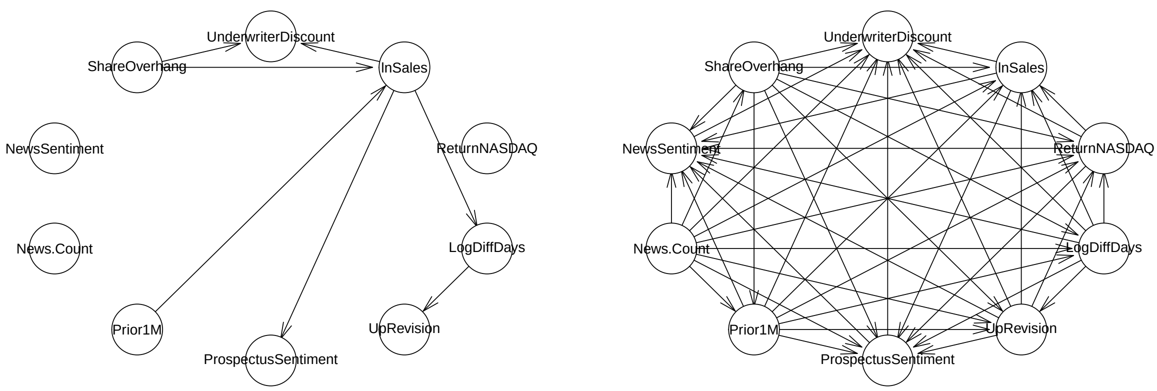


(a) Resulting Bayesian network with the exact student's t-test.

(b) Resulting Bayesian network with the mutual information based Monte Carlo permutation test.

Figure 2: Comparison of the Bayesian networks resulting from learning the structure of the network using the interleaved incremental association Markov blanket algorithm with different conditional independence tests, while the underlying variables are continuous.

The second algorithm we utilize is the hill-climbing algorithm. Here we vary the score from the Bayesian information criterion (BIC) to the log-likelihood criterion. As can be seen in Figure 3, the difference between those two network structures is quite significant. While the BIC score based network consist of 7 directed arcs, the log-likelihood based network is build upon 45 directed arcs. In comparison to the contraint-based algorithm networks, the BIC score one has all arcs, which are also included in both constraint-based networks. It also features *UpRevision* as a child of *LogDiffDays* as the mutual information based one. Additionally, the BIC network sets *UnderwriterDiscount* as a child of *ShareOverhang*. Interestingly, the directed arc between *lnSales* and *UnderwriterDiscount* is in the opposite direction for the BIC network than for the constraint-based networks.



(a) The network constructed with the hill-climbing algorithm and the BIC as score.

(b) The network constructed with the hill climbing algorithm and the log-likelihood as score.

Figure 3: Comparison of the Bayesian networks resulting from learning the structure of the network using the score-based hill-climbing algorithm with permutations, five restarts and two different scores to optimize.

Although the direction of some arcs in the graph might be contra intuitive we can still, after having learned the parameters in this structure, find answers to all interesting queries. If we e. g. consider the node *UpRevision*, we can now give a probability that there has been an above average upward adjustment of the price given e. g. an above average sales volume and above average time between the filing and the final offering day. In our case this is

$$P(UpRevision > \overline{UpRevision} \mid (lnSales > \overline{lnSales} \ \& \ \log\Delta Days_{SEC,IPO} > \overline{\log\Delta Days_{SEC,IPO}}) = 0.4371,$$

where a $\bar{x}$ indicates the average value of variable x. Thus when the log sales volume and the log time difference of the days are above average, with a probability of 43.71% the percentage of price adjustment from expected price to final price is also above average. If we consider the share overhang, we might want to know how likely it is for us to have an above average overhang given a low or high uncertainty sentiment in the prospectus.

$$P((ShareOverhang > \overline{ShareOverhang}) \mid (ProspectusSentiment > \overline{ProspectusSentiment})) = 0.5058$$
$$P((ShareOverhang > \overline{ShareOverhang}) \mid (ProspectusSentiment < \overline{ProspectusSentiment})) = 0.4796.$$

The above probabilities are calculated with the fitted values of the hill-climbing algortihm with BIC as score and maximum likelihood estimation for the parameters. Changing the the network structure obviously changes the probabilities. In our case, the differences between the structures and the different probabilites are however negligible as the probability differs by a maximum of 3%. For the Gaussian (continuous) data networks to be a good fit for the data, the variables under observation have to fulfil the Gaussian normality assumptions. In Figure 4 we plot the histograms of the variables we used for the Bayesian networks above.

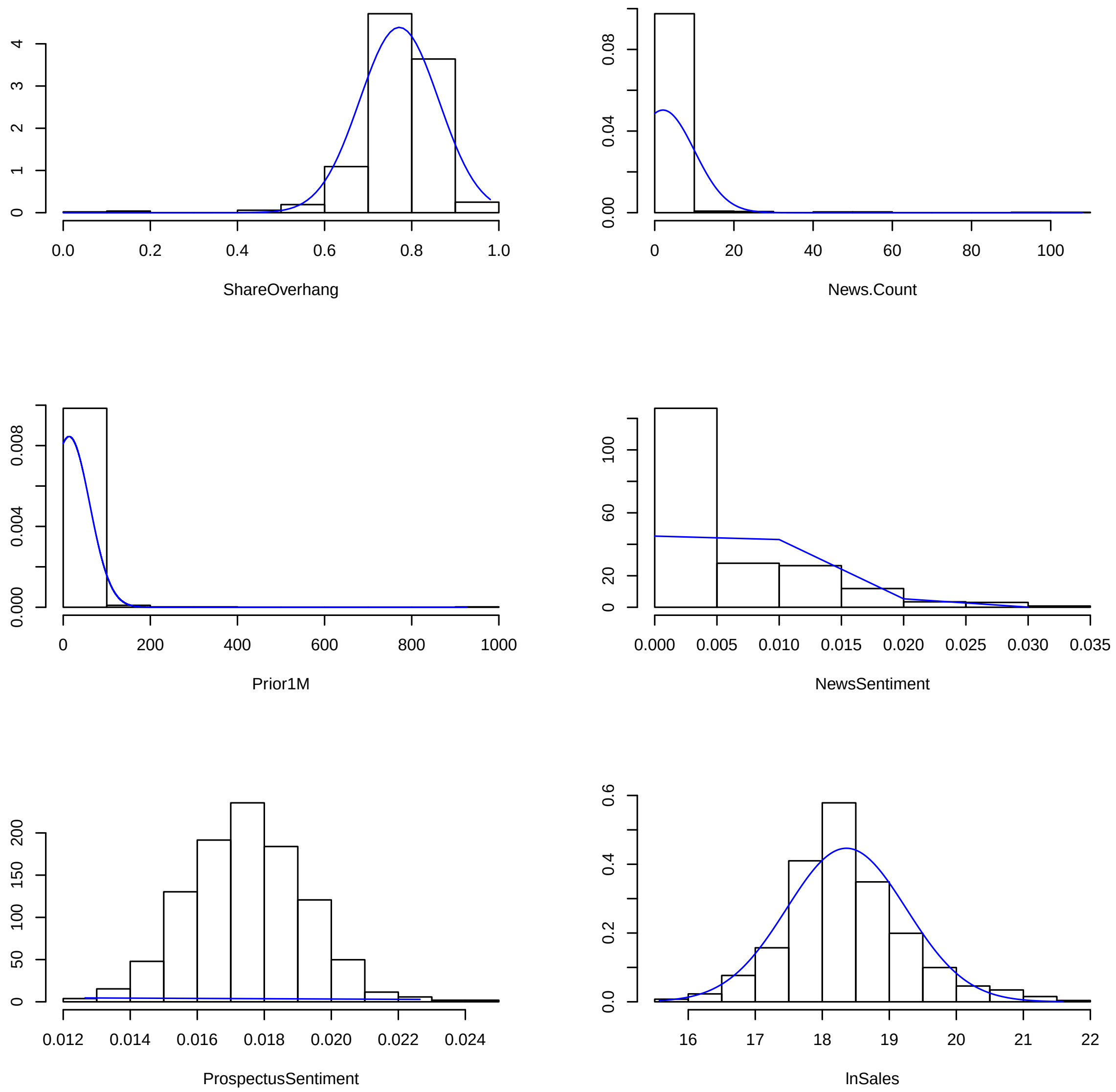


Figure 4: Histograms of the variables used for the continuous data Bayesian networks to show if they can be assumed to follow a normal distribution.

As we can see, most of the variables do not fulfil the assumption that they follow a normal distribution, most are not even symmetric. This lead us to the next section, where we have a look at discrete networks for the same data set.

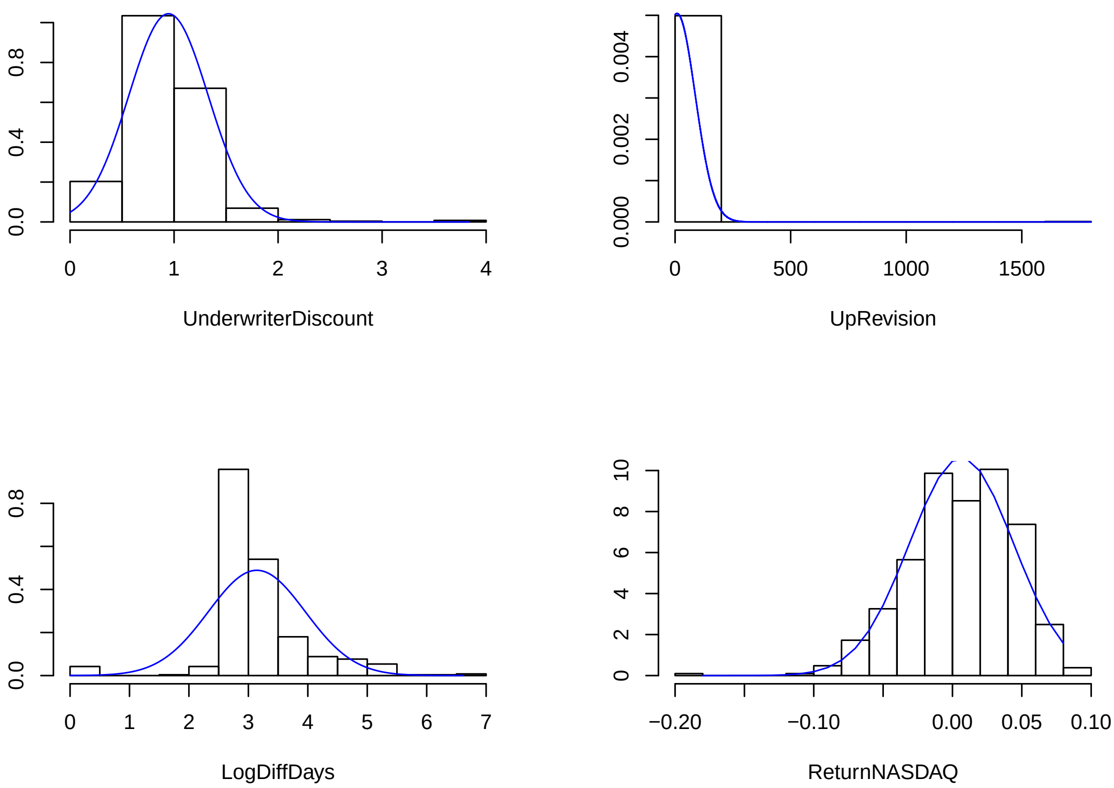
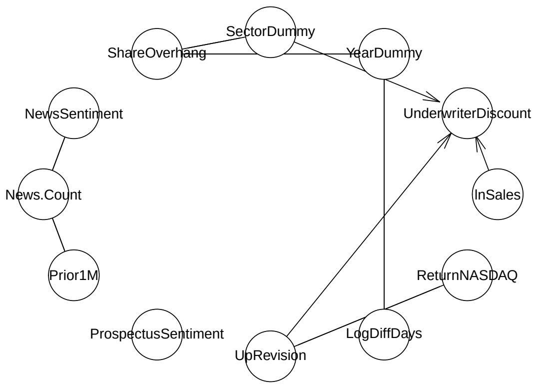


Figure 4: Histograms of the variables used for the continuous data Bayesian networks to show if they can be assumed to follow a normal distribution.

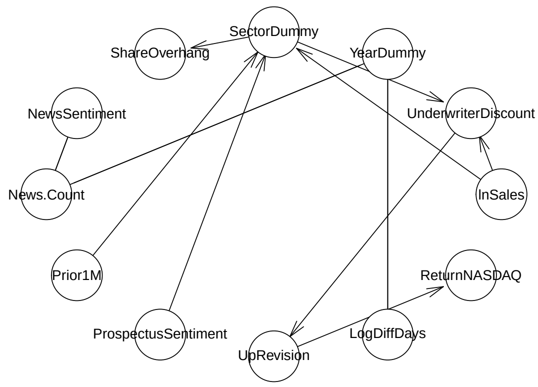
4.3 Learning a Discrete Data Bayesian Network

As we concluded in the previous section, a continuous Bayesian network might not be the best representation of our data as we do not fulfil the assumptions underlying Gaussian models. Our next step is thus to create some discrete data Bayesian networks. To do so, we have to transform the continuous variables which we have given in our data set into discrete ones. We do this by categorizing them into three different quantile groups. Corresponding to a group for low, average and high values of the variable. We then also add the dummy variables for year and sector, which we discarded beforehand. As for the case of the continuous data models, we use two different algorithms, again the incremental association Markov blanket algorithm and the hill-climbing algorithm, to learn the structure and different tests or scores, respectively. Obviously the tests and scores will now be different from those in the continuous data case.

We start again having a look at the incremental association Markov blanket algorithm. In Figure 5 we can see the resulting the structure using this algorithm and the chi-squared test statistics or mutual information Monte Carlo permutation test. Surprisingly, we now do have three out of undirected arcs in the mutual information network and only three out of nine arcs are directed for the chi-squared network. The mutual information based network also includes way more nodes into the network than the chi-squared one.



(a) Resulting Bayesian network with the chi-squared test.

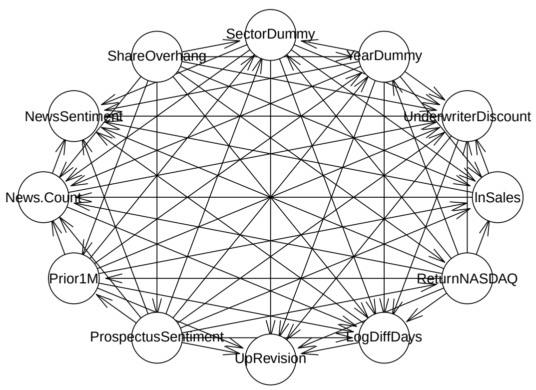


(b) Resulting Bayesian network with the mutual information based Monte Carlo permutation test.

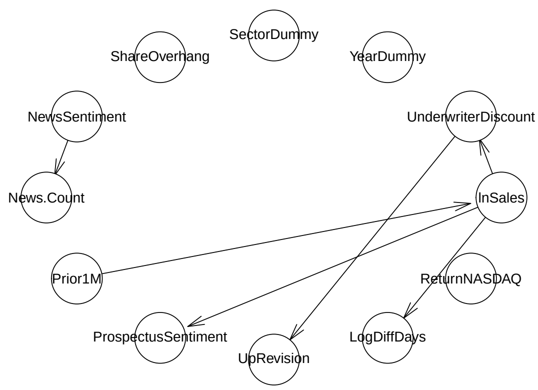
Figure 5: Comparison of the Bayesian networks resulting from learning the structure of the network using the interleaved incremental association Markov blanket algorithm with different conditional independence tests, while the underlying variables are discrete.

In comparison to the IAMB algorithm, we can have a look at the hill-climbing algorithm. The resulting networks differ quite substantially from those of the IAMB algorithm. Different to those, they do not have any undirected arcs. As in the continuous case the BIC score based network adds way less nodes into its final network than the log-likelihood one.

If we have a look at the probabilities that the mutual information based network assigns to our queries, we get some interesting results. However, before we can estimate our parameters with the maximum-likelihood estimation or Bayesian parameter estimation, we remove the undirected arcs from the network. As our relationships learned in the network are different from the cotinous data case, we cannot ask the same queries. Nevertheless, we get an interesting result, when we estimate the probability of the share overhang being high, average or low given a specific



(a) Resulting Bayesian network with the log-likelihood as score to be optimized.



(b) Resulting Bayesian network with the information theoretic Bayesian Information Criterion (BIC) as score to be optimized.

Figure 6: Comparison of the Bayesian networks resulting from learning the structure of the network using hill-climbing and different scores, while the underlying variables are discrete.

sector. Below we give some resulting probabilities for the sector Alternative Energy (2), Financial Services (10), Pharmaceuticals and Biotechnology (30) and Beverages (5):

$$P(ShareOverhang == High \mid SectorDummy == 2) = 0.775,$$
$$P(ShareOverhang == High \mid SectorDummy == 10) = 0.347,$$
$$P(ShareOverhang == High \mid SectorDummy == 30) = 0.341,$$
$$P(ShareOverhang == High \mid SectorDummy == 5) = 0.021.$$

Those results indicate a quite high dependency of the share overhang on the sector in which the initial public offering takes place. We have to be careful though with a generalization as we have only a small number of observations for some of the sectors.

5 Conclusion

In this paper, we introduce to graphical models in the form of Bayesian networks. We also tackle the economical problem of underpricing in initial public offerings (IPO), where the use of Bayesian networks seems reasonable. There are many theories and models that try to explain why we have underpricing. As more recent approaches also include e. g. textual information into the analysis, we have a sizeable amount of information which could help to either explain or predict underpricing of initial public offerings. We find that our Bayesian networks can successfully detect relationships among the variables and do in there results not differ but strengthen most empirical findings. Although the probabilities resulting from queries to the different networks are not always a very strong help when it comes to decision making, they do give information about how some variables are influencing each other. Some relationship however, such as the dependency of the share overhang on the sector have a strong foundation in the network and might be interesting to examine further.

In future research one could try to build a Bayesian network based on a theoretical approach to the phenomenon. As there is not only one but several explanations for underpricing and the relationships among variables concerning initial public offerings, this seems to be a challenging and possibly controversial approach.

A References

AGGARWAL, R. K., L. KRIGMAN, and K. L. WOMACK (2002). *Strategic IPO underpricing, information momentum, and lockup expiration selling*. In: *Journal of Financial Economics*, Vol. 66, No. 1, pp. 105–137.

BEATTY, R. P. and J. R. RITTER (1986). *Investment banking, reputation, and the underpricing of initial public offerings*. In: *Journal of Financial Economics*, Vol. 15, No. 1–2, pp. 213–232.

COOPER, G. F. (1990). *The computational complexity of probabilistic inference using bayesian belief networks*. In: *Artificial Intelligence*, Vol. 42, No. 2-3, pp. 393–405.

FEUERRIEGEL, S., J. T. SCHMITZ, and D. NEUMANN (2014). *What Matters Most? How Tone in Initial Public Offering Filings and Pre-IPO News Influences Stock Market Returns*. In: *SSRN Electronic Journal*.

HANLEY, K. W. (1993). *The underpricing of initial public offerings and the partial adjustment phenomenon*. In: *Journal of Financial Economics*, Vol. 34, No. 2, pp. 231–250.

HECKERMAN, D. (1995). *A Tutorial on Learning With Bayesian Networks*.

MARGARITIS, D. (2003). *Learning Bayesian network model structure from data*. PhD thesis. US Army.

NAGARAJAN, R., M. SCUTARI, and S. LÈBRE (2013). *Bayesian networks in R: With applications in systems biology*. Vol. 48. Use R! New York, NY: Springer.

RUUD, J. S. (1993). *Underwriter price support and the IPO underpricing puzzle*. In: *Journal of Financial Economics*, Vol. 34, No. 2, pp. 135–151.

SCHWARZ, G. (1978). *Estimating the Dimension of a Model*. In: *The Annals of Statistics*, Vol. 6, No. 2, pp. 461–464.

TSAMARDINOS, I., C. ALIFERIS, A. STATNIKOV, and E. STATNIKOV (2003). *Algorithms for Large Scale Markov Blanket Discovery*. In: *In The 16th International FLAIRS Conference, St*. AAAI Press, pp. 376–380.

B List of Figures

C List of Tables