Stock Investing Using HUGIN Software

An Easy Way to Use Quantitative Investment Techniques

Abstract

Quantitative investment methods have gained foothold in the financial world in the last ten years. This paper shows how Bayesian Networks can be used to create a computerized stock-picking model. By using historical data for 14 different economic relevant variables the model is designed to give trading recommendations (buy or sell) for the different companies included in a given dataset. The model has a hitrate of 60% and it generates an average return of 15.1% in each of the five investment periods tested. The model is found to give a significant higher return than the mean value of randomly generated portfolios and can therefore be said to posses 'skills' when it comes to 'stock picking'.

Introduction

Every day thousands of investors around the world evaluate an almost infinite set of information in the pursuit of predicting future stock prices and thereby earn as high returns as possible. Investors have through history developed a countless number of investment methods and strategies that include almost all kinds of financial and macroeconomic data one can think of. However, due to the numerous factors affecting the stock market, no methods have yet proven to be able to consistently predict future stock prices correct and because of the continuous inflow of potentially relevant information no methods probably ever will. So the quest of the investor is to make the best possible use of the available historical data and be able to combine it with the newest relevant information in order to make the most profitable investment decisions at any given time.

Bayesian Networks are ideal to use in exactly this setting, where it is of the highest importance for the investors to be able to combine quantitative and qualitative data in a systematic way. Further, Bayesian Networks have the advantage that new information easily can be taken into account, which is vital in the very dynamic and ever changing world of the stock market.

This paper provides an example of how to use Bayesian Networks as a 'stock picking' tool when investing in the Danish stock market. By using 10 different economic relevant factors, the Bayesian Network model comes up with buying/selling recommendations in a universe consisting of 14

Danish large cap stocks in a five years period. To assess the model's 'skills' when it comes to making profitable trading recommendations, the performance of the model is tested against the mean value of randomly generated portfolios.

Bayesian Networks – A brief overview

Bayesian Networks are networks of relations as they reflect the states of some part of a world that is being modeled, and describe how those states are related by probabilities. They are represented in the form of a directed acyclic graph consisting of nodes and directed arcs. Each node represents a stochastic variable that directly or indirectly affects the future stock price of a given company. Each arc denotes a causal or influential relationship between a pair of variables. Each variable in the network is assigned a conditional probability distribution based on either historical data or on subjective beliefs.

Bayesian Networks applied to stock investing

Since Graham and Dodd (1934) introduced fundamental analysis it has been the predominant method used when investing in stocks by both private and professional investors. However, since then a vast amount of different investment methods have been developed - for example the CAPM and APT models. The common thing for many of these investment methods is, that they try to predict future stock prices based on historical data for different economic variables. As the technological possibilities have evolved rapidly in the last twenty years, so have the complexity of the models used by professionals to decide which stocks to invest in. Whereas CAPM, which was developed in 1964, only uses three variables to predict the future price of a given stock, the quantitative models used today, can have over 100 different variables included. Today quant-funds, which use computer programs to find profitable trading strategies based solely on various kinds of historical economic data, make up 16% of the total value of actively managed assets in USA[1]. The quantitative investment methods used today differs from the fundamental analysis method by only using information that can be quantified. That means that more qualitative information such as the trustworthiness of a given company's management or the reliability of their financial statements are not taken into consideration as is the case with fundamental analysis.

---

[1] **Burke, Kevin (2006)**: "Not the man, but the machine", Registered Rep., September 1st 2006.

An important reason to why Bayesian Networks are useful when investing in stocks is that they combine the quantitative investment methods with the qualitative elements from fundamental analysis. Where the quantitative methods lack incorporation of qualitative measures, the fundamental analysis method does not always include all the relevant variables that might influence the future stock price and to which extend they correlate. With Bayesian Networks it is possible to use the best qualities from the two investment methods, while they at the same time are easy to use and update with new information. A Bayesian Network model reflects the investor's understanding of a particular set of relations and that makes it an ideal tool to handle decisions under uncertainty, which stock investing can be characterised as.

Modelling the Stock Picking as a Bayesian Network

Aim and method

The aim of the Bayesian Network model shown in this paper is to be able to pick out the stocks in a given investment universe, that in period t have a return that surpasses the required return they had at the beginning of period t.

The Bayesian Network model is used to provide a probability distribution for each stock's expected return in a future period. Based on this probability distribution and the required return, it is possible for the investor to make an investment decision on whether or not to buy the different stocks at the beginning of each investment period. The required return for each stock in each investment period is estimated with the CAPM model and the investment decision is taken separately for all the stocks included in the investment universe. Hence, the model is useful for stock-picking but not for optimal portfolio selection. All the qualitative relations are determined subjectively to get the most intuitive relations between the variables included in the network. The probability distributions of the model are, on the other hand, estimated from historical data.

The variables

One of the most important steps in developing a Bayesian Network model to use in stock investing is to single out which variables should be included in the model. When the variables included in this model was chosen, the limitations of the historical data for Danish stocks was taken into consideration but the most important issue was to get a model that both contains a given company's *i)* underlying business performance, *ii)* relative valuation in the market and *iii)* correlation with the general stock market. Together these three areas determine a probability distribution for the expected return of each of the stocks in the investment universe. The variables included in the three different areas are as follows (see Figure 1):

i) *Business performance:* The variables in this area are *EPS-growth*, *Sales-growth*, *Change in EBITDA-margin* and consensus estimates for the first three variables. In the model, *EPS-growth* is dependent on *Sales-growth*, *Change in EBITDA-margin* and consensus estimates for *EPS-growth*. *Sales-growth* and *Change in EBITDA-margin* are both dependent on the consensus estimates for each of the two variables.

ii) *Relative valuation:* The only variable in this area is the *Price-Earnings* (*PE*) multiple for each company at the beginning of each investment period.

iii) *Correlation with the general stock market:* The variables included in this area are the return on the S&P 500 index (*SPUS Return*), the Price-Earnings multiple for the S&P 500 index (*PE-SPUS*) and the value of the Chicago Board Options Exchange Volatility Index (*VIX*).

The data

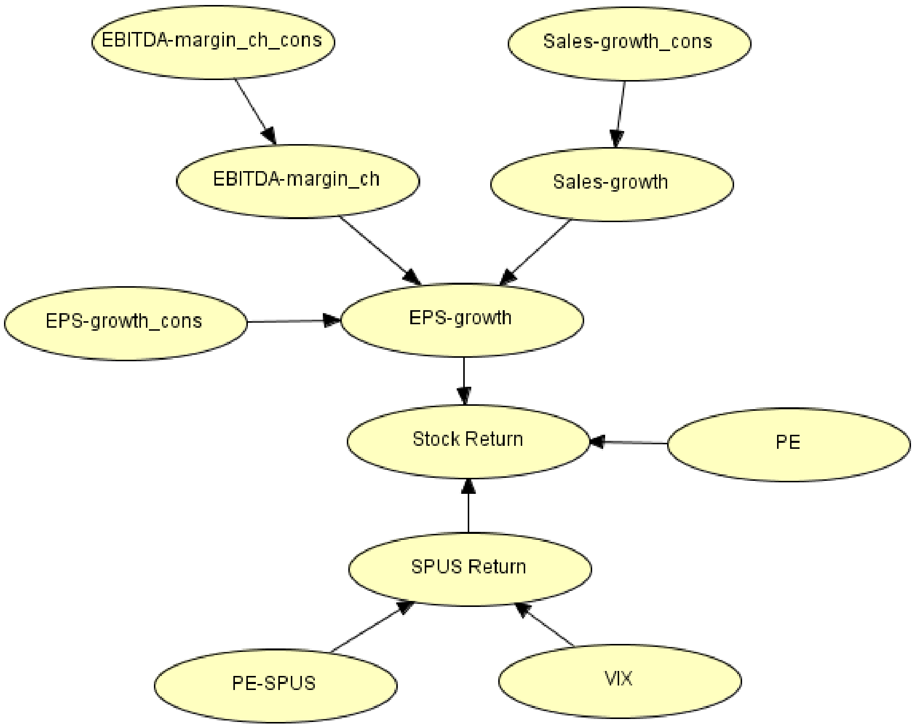
Because of the fundamental character of the variables included in the first area of the model, the historical data for these variables are on a yearly basis. The data sources used are each company's financial statements and The Thomson Corporation's financial database called Datastream while the consensus estimates are taken from Factset's JCF Estimates database. Data for the two other areas and the historical stock returns are taken from Reuters EcoWin. When the probability distributions of the variables are determined by historical data, it is important that the datasets have a certain size and quality. Given the variables included in the model, it is the availability of financial statements and consensus estimates for the Danish companies that put an upper bound on the number of companies included in the investment universe.

It is possible to apply both discrete and continuous probability distributions in a Bayesian Network model. Due to the somehow limited data material this model uses discrete probability distributions where each variable can take on only two values – *high* or *low*. All variables except *Stock Return* have been discretized based on the median. The *Stock Return* is discretized based on the estimated required return. The number of states can be varied, but it is necessary to do it in correspondence with the available data material in order to get the most precise probability distributions. Because the variables are discretized based on the median, it is not the marginal posterior probabilities that are interesting but instead the relations between the different variables expressed by the estimated conditional probability distributions.

The network

The Bayesian Network model for stock investing is shown in figure 1 below. The target variable is, as stated earlier, the *Stock Return*. By using existing information about the 10 economic variables in the past, it is possible to get a probability distribution for the future return of a given stock and from here make a qualified investment decision.

Figure 1- Bayesian Network model for stock investing



Applying the model

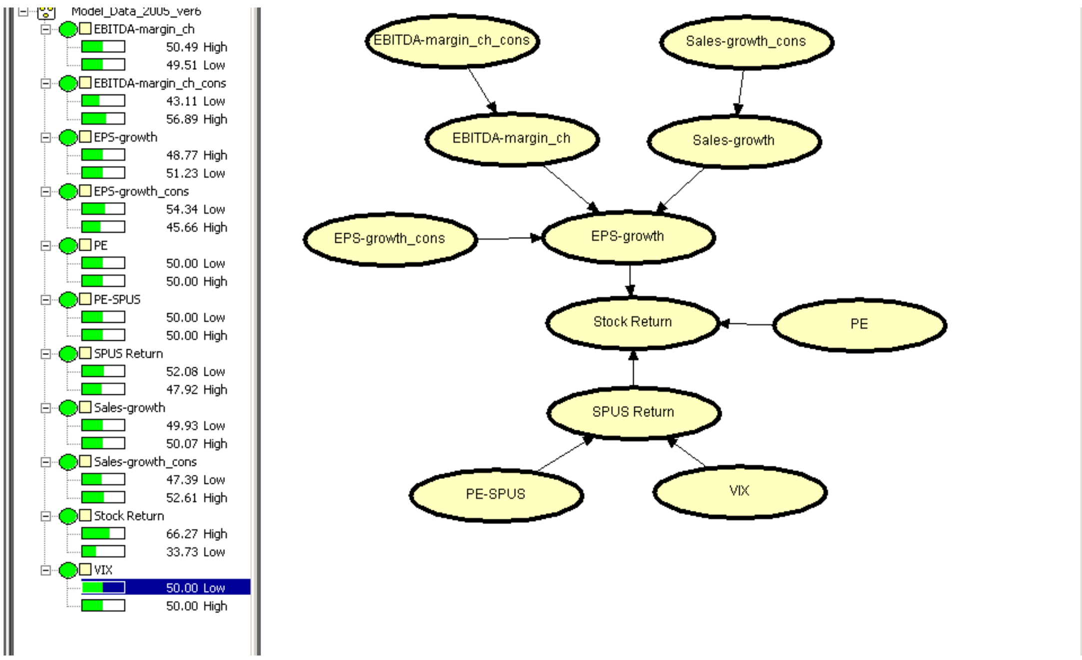
For each company in the investment universe an investment period of nine months has been determined. It starts three months after the end of each company's fiscal year in order to exclude the effect the forward looking statements, that all companies put forward in connection with the release of their annual report, have on stock prices, as it is impossible to catch these relations in the historical data.

The historical data for all the variables up to the start of the given investment period are used to estimate the probability distributions of the network. The financial data for each company such as *EPS-growth, Change in EBITDA-margin* and *Sales-growth* are measured as of fiscal year end. Stock Return and Return on S&P 500 are measured in the historical investment periods that correspond to the future investment period in which the investor wants to decide which stocks to buy. For the rest of the variables it is possible to update the model at the start of the investment period with the latest available data and thereby improving the foundation on which the investment decision is taken.

Example

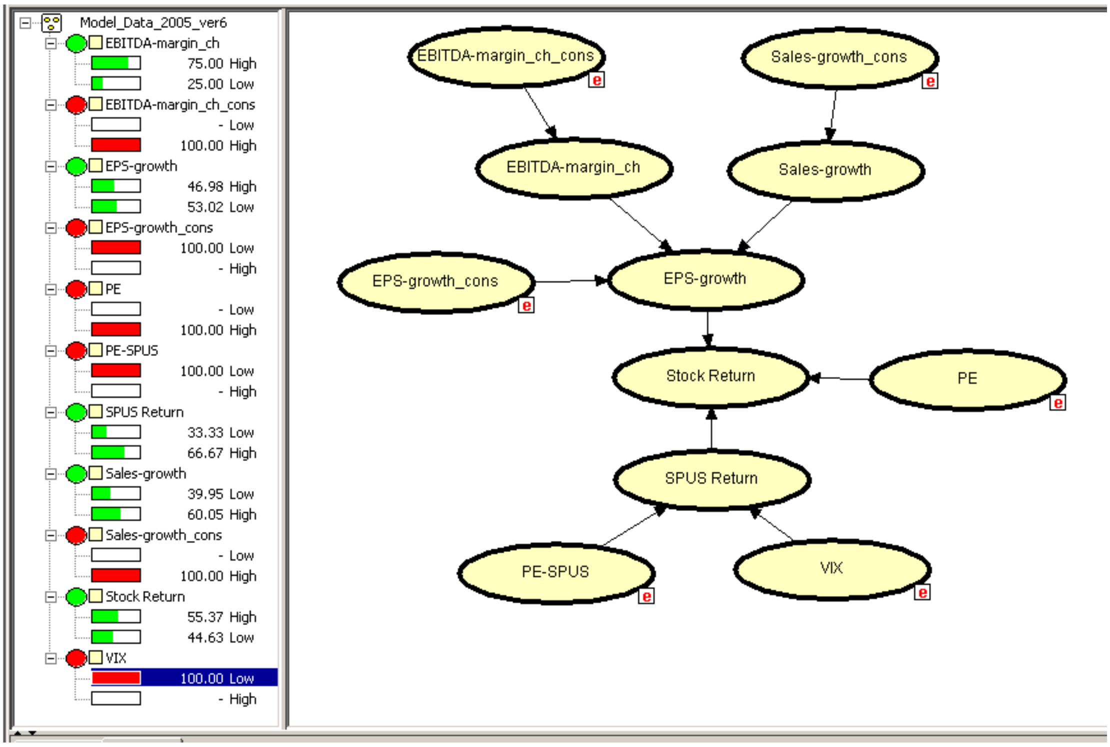
As an example, consider an investor who has to decide on whether or not to invest in the Danish company Novo Nordisk at 31 March 2006 (three months after the end of the 2005 fiscal year). On the grounds of the historical data for the different variables a marginal posterior probability distribution for all the variables can be calculated using HUGIN as shown in figure 2 below.

Figure 2 – Marginal posterior probability distributions in HUGIN



At 31 March 2006 the investor can update the model with the latest consensus estimates and the values for *PE-SPUS*, *VIX* and *PE* as of 31 March. As can be seen in Figure 3 below, the probability distributions for these variables are now updated to reflect the most recent information. Because these values are known for certain, the given states take on a 100% probability in the updated network.

Figure 3 - Marginal posterior probability distributions updated with new information in HUGIN



As can be seen the probability distributions for the remaining variables changes as well to reflect the new information. As an example it is seen that the probability of the S&P 500 to have a high return in the coming investment period is 66.7%. This reflects that the S&P 500 historically has had a high return in the investment period 66.7% of the times where *VIX* and *PE-SPUS* were low at the beginning of the investment period.

The investor makes his investment decision based on the probability distribution of the variable *Stock Return*. In the example, the investor should invest in Novo Nordisk since the probability that company will have a high return in the coming investment period is 55,4% (i.e., higher than the required return). If the probability had been below 50% the investor should on the other hand not invest in Novo Nordisk. The actual return for Novo Nordisk in the period from 1 April 2006 to 31 December 2006 was 22.85% and thereby well above the estimated required return of 5.36%.

Measuring the performance of the Bayesian Network model

There are several ways to measure the performance of a Bayesian Network model when it comes to stock-picking. One method is to test the model *out of sample* by measuring how many times it makes correct trading recommendation for each of the fourteen companies in a given number of investment periods. When testing the model in five different investment periods in the years 2002-

2006 the model is found to make the correct trading recommendations in 60% of the cases. It varies within the different companies from a hitrate of only 20% to hitrates of 100%.

A more interesting way of measuring the model's performance is to calculate what an investor who had followed all of the model's trading recommendations would have gotten in return over the five investment periods from 2002-2006. Given certain assumptions[2], the investor would have gotten a total return of 98.4% from 2002-2006 corresponding to an average return of 15.1% (arithmetic). Even though this return would seem satisfying for most investors, it is not sufficient to conclude that the Bayesian Network model has 'skills' when it comes to stock-picking. In order to do that, it is necessary to compare the model's return with a given benchmark or comparable investment strategies with the same risk profile. As there are no obvious benchmarks or comparable investment strategies to use (because of the small sample of 14 Danish companies), the method of 'Random Portfolios' or 'Monte Carlo Simulation' is used to assess the investment skills of the Bayesian Network model.

The most precise way to judge the investment skills of the Bayesian Network model would be to compare the realized return with the return of all the possible returns for portfolios containing the 14 Danish stocks and T-bills in the five investment periods. However, with 14 companies and five investment periods there are 2^{14*5} possible portfolios. This makes the task infeasible. Instead of that you can use a random selection of all the possible portfolios, which will be sufficient to assess the investment skills of the Bayesian Network model statistically.

The average return for the random portfolios in the five investment periods has a mean value of 12.81% and a standard deviation of 2.39%. With a sample size of 1,000 this gives a 99%-confidence interval of [12.61%, 13.00%]. As the random portfolio's average returns are normally distributed it can be concluded that the Bayesian Networks model's average return of 15.1% is significantly higher and the model can therefore be said to have skills when it comes to stock-picking.

[2] It is assumed that the investor can only invest 1/14 of his portfolio in each of the 14 stocks in every investment period and that he cannot short any of the stocks. So when the model gives him a sell signal on a given stock he will place 1/14 of his portfolio in Danish T-Bills. He will invest the full portfolio amount in each investment period. Taxes and transaction costs are not taken into consideration.

The above results are robust to changes in the applied equity risk premium used in the estimations of each stock's required return, the Experience Count used in HUGIN and the number of investment periods used.

Conclusion

The applied Bayesian Network model for stock investing gives very good out-of-sample test results when compared to the return of Random Portfolios. Because of the few testing periods and the small number of included companies you should of course be careful with making general conclusions based on the shown results. However, it shows that Bayesian Networks works well even in the absence of a large data material as long as you have a well-defined qualitative structure.

The good results obtained with this, to some extent, simple model underlines the potential of Bayesian Network models in the world of stock investing. With a larger set of data it will be possible for investors to build more sophisticated models that not only give trading recommendations but that also can be used for portfolio decision-making and risk analysis. Because of the great flexibility of Bayesian Network models, an investor can choose whatever investment period he wants. Bayesian Networks can handle both long-term relations and intraday trading relations. As the future stock returns are uncertain a rational investor bases his investment decisions on probabilities for the future stock returns and Bayesian Network models are an ideal tool for calculating these probabilities.

About the author:

Daniel Andersen holds a M.Sc. degree in Economics from the University of Copenhagen, and work as a Sales Analyst in the Equity Department of FIH-Kaupthing. His key focus is to generate daily trading ideas as well as developing new products to use on the sales desk or in research.