UNIVERSITY OF TURKU

# ABSTRACT

| | |
|---|---|
| | Bachelor's thesis |
| x | Master's thesis |
| | Licentiate's thesis |
| | Doctor's thesis |

| | | | |
|---|---|---|---|
| Subject | Accounting and finance | Date | 6.8.2019 |
| Author | Marjut Soininen | Student number | 509850 |
| | | Number of pages | 92 |
| Title | Factor selection for multifactor models: Bayesian model averaging approach | | |
| Supervisors | Prof. Mika Vaihekoski and M.Sc. Valtteri Peltonen | | |

The optimal set of factors for multifactor model has been under discussion in the previous decades. Research has focused mostly on the Fama-French models and the inclusion of some additional factors. Moreover, the research has been mostly limited to fundamental factors and traditional hypothesis testing based on *p*-values. In this thesis, factor selection is done with the Bayesian approach, where factors are evaluated based on their posterior probabilities of inclusion rather than their *p*-values. Twelve factors are included in the algorithm as potential factors and all possible combinations of these factors are evaluated simultaneously. The model with the highest posterior probability is then chosen. The specific method used is Bayesian model averaging (BMA), which has received good results in previous studies of stock market factor selection. The significance of macroeconomic factors besides fundamental factors is examined as well. Finally, the predictability ability of the constructed multifactor model is compared to the Fama-French models and CAPM.

The time period of research is from 2007 to 2018, where 2007−2016 is the in-sample and 2017−2018 the out-of-sample period. All NYSE, AMEX and NASDAQ stocks, which have the relevant data available, are included in the research. The following six factors have the highest posterior probability: market premium, size, value, momentum, the changes in long-term interest rate, and oil price change. These factors form the BMA multifactor model. The out-of-sample results show that the BMA multifactor model constructed here has the lowest Mean Squared Error (MSE) value, so it can predict stock returns out-of-sample better than Fama-French models or CAPM. The results are robust for the test portfolios used.

The results are in line with previous findings that the Bayesian approach is a useful tool for factor selection. Widely examined fundamental factors (market, size and value premium) as well as momentum factor still seem to have some predictability ability in the US, and since inclusion of macroeconomic factors improves the predictability of model, they should be examined more alongside fundamental factors in the literature.

| | |
|---|---|
| Key words | Bayesian model averaging, factor selection, multifactor model, US stock market |
| Further information | |

TURUN YLIOPISTO

# TIIVISTELMÄ

| | |
|---|---|
| | Kandidaatintutkielma |
| x | Pro gradu -tutkielma |
| | Lisensiaatintutkielma |
| | Väitöskirja |

| | | | |
|---|---|---|---|
| Oppiaine | Laskentatoimi ja rahoitus | Päivämäärä | 6.8.2019 |
| Tekijä | Marjut Soininen | Matrikkelinumero | 509850 |
| | | Sivumäärä | 92 |
| Otsikko | Monifaktorimallin faktoreiden valinta bayesilaisella keskiarvoistamisen menetelmällä | | |
| Ohjaajat | Prof. Mika Vaihekoski ja KTM Valtteri Peltonen | | |


Monifaktorimallit ja faktoreiden optimaalinen valinta ovat kiinnostaneet tutkijoita viime vuosikymmeninä. Tutkimus on keskittynyt lähinnä Fama-French-malleihin ja niihin yhdisteltäviin faktoreihin. Lisäksi tutkimus on rajoittunut suurimmaksi osaksi fundamentaalisiin faktoreihin ja perinteiseen *p*-arvoihin perustuvaan hypoteesien testaamiseen. Tässä tutkielmassa faktoreiden valinta suoritetaan bayesilaisella menetelmällä, jossa faktoreita arvioidaan *p*-arvojen sijaan posterioritodennäköisyyksillä. Kahdentoista potentiaalisen faktorin kaikki mahdolliset yhdistelmät arvioidaan perustuen näihin todennäköisyyksiin ja malli, jolla on korkein posterioritodennäköisyys, valitaan. Käytetty menetelmä on bayesilainen mallikeskiarvoistaminen (Bayesian model averaging, BMA), jolla on saavutettu hyviä tuloksia aiemmissa faktoreiden valintaan liittyvissä tutkimuksissa. Lisäksi makroekonomisten faktoreiden merkitystä tutkitaan. Muodostetun multifaktorimallin ennustuskykyä verrataan Fama-French-malleihin ja CAPM:iin keskineliövirheellä (MSE) mitattuna.

Tutkimus suoritetaan USA:n markkinoilla 2007−2018, jossa otosikkuna on 2007−2016 ja otoksen ulkopuolinen ikkuna 2017−2018. Kaikki NYSE:n, AMEX:in ja NASDAQ:in osakkeet, joista on tarvittava data saatavilla, sisällytetään tarkasteluun. Seuraavilla kuudella faktorilla on korkein posterioritodennäköisyys ja jotka siten sisällytetään BMA monifaktorimalliin: markkinapreemio, koko- ja arvofaktorit, momentum, pitkien korkojen muutos ja öljyn hinnan muutos. Otoksen ulkopuolinen testaus osoittaa, että BMA monifaktorimallilla on pienin keskineliövirhe, joten se pystyy ennustamaan osaketuottoja paremmin kuin Fama-French-mallit tai CAPM. Tulokset ovat robusteja testiportfolioiden valinnalle.

Tulokset ovat linjassa aiempien tutkimusten kanssa ja osoittavat, että BMA menetelmä sopii hyvin faktoreiden valintaan. Lisäksi tulosten mukaan paljon tutkitut fundamentaaliset faktorit näyttävät edelleen selittävän osaketuottoja USA:ssa. Makroekonomisten muuttujien sisällyttäminen parantaa myös mallin ennustekykyä ja niitä tulisi siten tutkia enemmän fundamentaalisten faktoreiden ohella.



| | |
|---|---|
| Asiasanat | bayesilainen tilastotiede, faktorivalinta, monifaktorimalli, USA:n osakemarkkinat |
| Muita tietoja | |

UNIVERSITY
OF TURKU
Turku School of
Economics

# FACTOR SELECTION FOR MULTIFACTOR MODELS

## Bayesian model averaging approach

Master´s Thesis
in Accounting and Finance

Author:
Marjut Soininen

Supervisors:
Prof. Mika Vaihekoski
M.Sc. Valtteri Peltonen

6.8.2019
Turku

# TABLE OF CONTENTS

## LIST OF TABLES



## LIST OF FIGURES

# 1 INTRODUCTION

## 1.1 Background and motivation

The characteristics of stock returns and portfolio selection are some of the core interests in finance and have always been of interest to both practitioners and academics. One of the most crucial problems is finding a specific model, which adequately expresses the dynamics of asset returns. (Ando 2009, 556; Tsai et al. 2010, 110.) While there are numerous papers documenting predictability, there is little consensus across these articles on what the important variables are (Cremers 2002, 1223).

Modern portfolio theory, developed by Markowitz (1952), was one of the first and most important steps in understanding stock behavior (Rubinstein 2002, 1041). Since then various factors have been found and different models tested over the years. In the 60s many researchers focused on the Capital Asset Pricing model (CAPM), which is based on Markowitz's theory and was developed independently by Sharpe (1964), Lintner (1965) and Mossin (1966). The main point in the CAPM was that stock returns can be explained with only one factor, the market premium, to which different assets have different loadings (beta coefficients).

The CAPM equilibrium has been criticized since the model was introduced. Many researchers have recognized the problems in testing the model that are caused by the strict assumptions behind the model. One of the most known critics is the so-called Roll's critic, which states that only the efficiency of the market portfolio can be tested with the CAPM and it has no implications on the return-risk-relationship (Roll, 1977). Later Fama and French (2004) received similar results.

Besides the testing errors, the CAPM's assumption of a one risk factor has been widely questioned in past decades. Many researchers have found pricing anomalies that generate new risk factors. Some of the most known pricing anomalies are size effect, value effect and momentum effect. (Bender et al. 2013, 4.) Because of these findings, finance researchers have focused lately on the multifactor models that consist of various risk characteristics and solve the portfolio choice problem with models far more sophisticated than the mean-variance framework of Markowitz or the CAPM one factor model (Ando 2009, 551). Fama and French (1992; 1993; 2004) have made pioneering research among risk factors and developed perhaps the best-known multifactor models.

Researchers and financial professionals mostly agree nowadays that there is more than one risk factor that affects stock returns. A common view however, has not been reached on how many factors and which ones should be included. Some researchers, for example, have suggested that the size effect has disappeared in the last few years although it is still widely used in many funds by practitioners (Ang 2014, 229).

Many different factors have been considered as a component of multifactor models. These factors can be divided into fundamental, macroeconomic and statistical factors. (Connor 1995, 42.) The question now is how to combine these potential factors to achieve the most optimal multifactor model, that is, the combination of factors that explains stock returns most accurately and can thus be used in predicting the stock returns.

The common way to evaluate different factors and risk premiums is to add one or more new factors to the Fama-French models. For example, Carhart (1997) adds a momentum premium to the FF3-factor model and reaches better results compared to the original 3-factor model. However, this approach assumes that the combination of factors is fixed and evaluates only one model per time. This method can lead to the better model being compared to universal multifactor models but does not technically focus on factor selection as the set of factors is picked in advance.

To get the most optimal multifactor model all the existing factors and models should be included in testing. Obviously, it is not possible to include all the factors since there are thousands of them in existing literature alone and many of them are strongly correlated with each other. Some methods have been developed to test multiple combinations of factors, which leads us a little bit closer to factor selection analysis and the most optimal multifactor model. (Fernandez 2001, 381.)

One method is to utilize the Bayesian approach, where the main idea is to first select a set of possible factors and then evaluate all the possible combinations of these factors based on their prior and posterior probabilities (Ando 2009). The prior probability refers to the probability, in which the factor is included in the optimal model based on previous knowledge, views, or specific distribution. These probabilities are then updated with the new data, which results in the posterior probabilities. (Puga et al. 2015, 277.)

With these methods it is possible to examine the common problem of finding the most relevant factors and construct a model, where the number of factors is restricted. Bayesian methodologies have several advantages compared to traditional hypothesis testing (for example, the Fama-MacBeth approach). For example, they take into account parameter uncertainty, as they allow a wide range of prior distributions and evaluate all possible combinations of factors to calculate posterior probabilities. (Bianchi et al. 2017, 111.)

Common Bayesian methods however do not take into account model uncertainty, which is another possible problem in classical hypothesis testing. Barnard (1963) was the first one to mention the idea of combining models in statistical literature. Later some other approaches were developed, like Akaike's information criterion (AIC). (Hoeting et al. 1999, 384, 398.) Lately, more attention has started to be paid to applying model combining to Bayesian methodology. This has resulted in the Bayesian model averaging approach (BMA), which takes into account both parameter and model uncertainty.

Although factor selection is a very important part of constructing multifactor models, previous literature provides very little guidance on it. There is no consensus among researchers about what the best factors in predicting stock returns would be. In this thesis, the problem of selecting relevant factors is examined and therefore a set of factors is not decided beforehand. Instead, many fundamental and macroeconomic factors are recognized as potential factors and the final set of optimal factors is formed from among them with the Bayesian model averaging method. Macroeconomic factors are included in the study because the existing empirical research has lately mostly focused only on fundamental factors.

Bayesian model averaging can be used in two ways. First, it can be used to find the most optimal multifactor model among all the competing models based on the posterior probabilities of models. The model that has the highest posterior probability is selected. Second, the number of competing models can be reduced with the use of Occam's window and the remaining models can then be used as portfolio weights based on the posterior probabilities of models. (Steel 2011, 33.) This second approach uses all the competing models and greatly reduces model uncertainty.

Bayesian methodology is widely known among statisticians, and has been applied to many different purposes, but it's implications on the factor selection problem in multifactor models are relatively unknown and a limited number of previous researches are available on the subject. Especially few research papers are available on the Bayesian model averaging approach, although the results have been promising. Moreover, the BMA method has many advantages compared to other, more classical, factor selection approaches. (Young & Lenk 1998; Berger et al. 2001; Fernandez et al. 2001; Ando 2009.) It is interesting to examine how the Bayesian model averaging method works in factor and model selection at the present time and if it is possible to construct better and more accurate multifactor models than the universal FF3 and FF4 factor models.

## 1.2 Research questions and limitations

The purpose of this thesis is to consider the different factors that have an impact on stock returns and examine the problem of selecting the most optimal factors. The possible factors included in testing are based on the existing literature. The factor selection is done with the Bayesian model averaging approach. The research questions are formed as follows:

> *Which factors are important for explaining stock returns in the Bayesian model averaging approach and should therefore be included in the optimal multifactor model?*

and

> *Does this Bayesian multifactor model predict stock returns out-of-sample better compared to benchmark models?*

The first question is examined with the Bayesian model averaging technique and the method is constructed mostly based on Ando's (2009) article about Bayesian portfolio selection. In Ando's research two different Bayesian approaches are utilized and compared to the mean-variance portfolio selection style. However, in this thesis only one Bayesian construction method is used, the Bayesian model averaging method.

Factor selection is done in-sample and as a result, the optimal multifactor model according to the Bayesian method is achieved. Other competing models are also briefly analyzed, but only the best model is selected to the out-of-sample testing. In other words, the second method of using Bayesian model averaging, where all the individual models are included in the optimal model based on the weights determined by posterior probabilities, is not used in this thesis. This choice can be made because results show that the difference in the posterior probabilities between the best and second-best models is relatively large. Therefore, focusing solely on the best model does not significantly reduce the reliability of the study (Steel et al. 2011, 33).

The second research question relates to the usefulness of constructed multifactor models and their ability to predict stock returns compared to benchmark models. The out-of-sample performance of the formed Bayesian multifactor model is compared to the one factor model (CAPM), and two multifactor models, the FF3-factor model and FF4-factor model. The performance is examined using the mean squared error (MSE), which calculates the difference between the actual returns and the expected returns of the factor model.

Theoretically, in the Bayesian model averaging method, an unlimited number of factors and models can be considered. In this thesis the number of potential factors is restricted to four fundamental and eight macroeconomic factors that have strong academic relevance and logical, but not always rational, reasons for their existence in the markets. All other factors are left to future research.

The testing of the competing factor models is done for the portfolios of stocks rather than for individual assets. The reason for this is that returns need to be stationary in the sense that they have approximately the same mean and covariance and individual stocks are usually very volatile. (Ericsson & Karlsson 2003, 6.) Twelve industry portfolios are therefore used for testing the hypothesis, which is a common method in the literature.

The research is done using U.S. data and all the NYSE, NASDAQ and AMEX stocks, that have the relevant historical data available, are included. The time period for the in-sample factor construction is ten years, 2007−2016. Monthly data is used, so 120 observations are included in each industry portfolio in the BMA factor selection. The time period for out-of-sample testing is two years, 2017−2018.

## 1.3 Structure of the thesis

Section 2 focuses on the theory behind the multifactor models. First, the efficient market hypothesis is introduced and after that the most relevant stock market anomalies and factors. Finally, multifactor models, including the APT and Fama-French models, are discussed briefly based on these factors.

The next section, Section 3, concentrates on the Bayesian approach. First, the basic idea behind the model is introduced and its implications on factor selection are discussed, as well as the advantages and disadvantages of the model and prior information. Second, the Bayesian model averaging method is introduced, and the formulas derived. In addition, the existing relevant literature from the topic and previous findings are discussed in the end of Section 3.

Section 4 focuses on representing the data and methodology that is used for the empirical research of this thesis. The results are finally shown in Section 5, first from the Bayesian model selection and then the performance of the model compared to the benchmarks. A robustness check is made in the end of Section 5. The conclusions are summarized in Section 6 after which references are shown.

# 2 MARKET PREMIUM AND OTHER RISK FACTORS IN MULTIFACTOR MODELS

## 2.1 Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) is a theory whereby the securities market is efficient, which means that share prices always fully reflect all available information. The theory assumes that when new information arises, it spreads very quickly and affects share prices without delay. (Malkiel 2003, 59.) Therefore, in efficient markets, fluctuation in share prices should not appear if there is no new information available and stocks should always trade at their fair value on stock exchanges.

EMH was presented by Fama (1970) and it is based on previous findings of investors' rationality and the idea of a "random walk". Random walk describes the scenario where all subsequent price changes are unrelated to previous prices. The logic of the theory is that if all the information is immediately reflected in share prices, then tomorrow's news will affect only tomorrow's price change and be independent of the price changes of today. In addition, news is by definition unpredictable, and thus, price changes are also unpredictable and random. (Malkiel 2003, 59.)

According to EMH, the generation of consistent risk-adjusted excess returns, or alpha, is impossible in efficient markets. This means that it should be impossible to "win the markets" constantly, that is, outperform the market through stock selection or market timing. In an efficient market, market prices of shares can differ a lot from their real values, but the important thing is that these deviations must be entirely unpredictable and completely random. Because of this, it is still possible to "win the market" coincidentally even though a systematic excess return should not be possible in the equilibrium.

According to Fama's (1970) research, there are three different levels of efficiency in the markets. These levels are weak efficiency, semi-strong efficiency and strong efficiency, and they are determined as follows (Fama 1970, 388; Malkiel 2003, 59):

- The weak efficient market hypothesis suggests that today's stock prices reflect all historical data, e.g. past stock prices and volume, and therefore technical analysis cannot be effectively utilized to support investors in making trading decisions or gaining excess returns. However, if fundamental analysis is used, undervalued and overvalued stocks can be recognized. This means that investors can utilize companies' financial statements to increase their chances of gaining higher-than-market-average profits.

- The semi-strong efficient market hypothesis is based on the belief that all historical and public information is reflected in stock prices, and hence, investors cannot utilize technical or fundamental analysis to gain consistently excess returns, or alpha, in the market. This version of the market efficiency theory believes that only information that is not available to the public can aid investors in increasing their returns to a level above the market.
- The strong efficient market hypothesis states that all information, both the information available to the public and insider information, is completely accounted for in current stock prices. In other words, there is no type of information that could give an investor a market advantage.

There is no consensus among financial economists whether the EMH is true or not. Even after four decades of research and thousands of journal papers, researchers have not yet agreed on whether financial markets are efficient or not. (Findlay & Williams 2000, 196.)

The efficient market theory reached the height of its dominance in the academic world around the 1970s. At that time, the rational expectations revolution in economic theory was a fresh new idea that occupied the center of attention among researchers and it was widely believed that stock prices always incorporated the best information about fundamental values and that prices changed only because of good, new and sensible information. (Shiller 2003, 83.)

A simple indicator of market efficiency is the ability of professional fund managers to outperform the market. This is obvious, because if irrational investors and rational estimates of the present value of corporations mostly determine market prices, and if it is easy to spot predictable patterns in security returns, then professionals should be able to outperform the market. However, a significantly large body of evidence suggests that professional investment managers are not able to outperform buy-and-hold index funds, which indicates that these markets are at least semi-strong efficient (Malkiel 2003, 76−77). One assumption behind the efficient market hypothesis however, is that there are no costs or taxes in the market and underperformance of professionals relative to the market index can be explained by these extra costs. A typical active mutual fund has an expense ratio of just less than 150 basis points whereas index funds can be run with lower expense ratios, less than 20 basis points, even for non-professional individual investors. Furthermore, active managers turn over their portfolio, often as much as 100% each year, which requires additional costs from brokerage costs, bid-asked spreads and market impact. (Malkiel 2005, 3.) This leads to the conclusion that the performance of professionals cannot be counted as a reliable indicator of market efficiency.

In past decades, the academic dominance of the efficient market hypothesis has become far less universal. Many researchers have begun to believe that stock prices are at least partially predictable, which is against the random walk theory behind the efficient market hypothesis. (Malkiel 2003, 60.) There seems to be several examples, at least ex-post, where market prices failed to fully reflect available information. Periods of large-scale irrationality, such as “bubbles” in history, have convinced many researchers that the efficient market hypothesis should be rejected. (Malkiel 2005, 2; Kartašova et al. 2014, 332.)

Many researchers have recently suggested that stock prices are, to a significant extent, predictable on the basis either of past returns or some fundamental valuation metrics, or both, and have therefore rejected the efficient market hypothesis (Malkiel 2005, 2.) This would mean that new information is not the only factor that influences price changes (Kartašova et al. 2014, 332). Actually, the efficient markets model for the aggregate stock market has never been supported by any study very effectively, linking stock market fluctuations with the fundamentals (Shiller 2003, 90). These findings have resulted in the behavioral elements of stock-price determination. Many researchers have even made the far more controversial claim that it is possible to earn excess risk adjusted returns with these predictable patterns. (Malkiel 2003, 60.)

Although many researchers have questioned the EMH and proved that some predictability exists in the markets, some research has stated that even though the aggregate stock market appears to be inefficient, individual stock prices do show some correspondence to EMH. That is, the stock market is micro efficient but macro inefficient, since among the investors there is considerable predictable variation across individual stocks in their predictable future paths of dividends but only a little predictable variation in aggregate dividends in the whole stock market. Thus, it is easier to predict the aggregate stock market than the individual stocks. (Shiller 2003, 89.) It is also clear, that market efficiency should not be expected to be so wrong that immediate profits would continuously be available for investors (Shiller 2003, 101). To conclude, there is surely at least some predictability in stock markets, but also some support to EMH. Today the most common opinion is probably that the markets could be weak form efficient or semi-strong form efficient, but never strong form efficient. (Kartašova et al. 2014, 332.)

## 2.2 Pricing anomalies

The Capital Asset Pricing Model, which is based on modern portfolio theory and the efficient market hypothesis, was one of the earliest widely accepted models for stock pricing. In the CAPM, stocks have only two main drivers: systematic risk and idiosyncratic risk. Systematic risk refers to the risk that arises from exposure to the market and is captured by beta, the sensitivity of a stock’s return to the market. (Bender et al. 2013, 4.)

In past decades, many researchers have found different anomalies that explain the variation in stock returns and have criticized the CAPM equilibrium (Avramov & Chordia 2006, 1001). The reason these anomalies have gained a wide-spread interest among researchers is that their behavioral explanations challenge semi-strong-form market efficiency. Another reason for the increased interest in these additional factors is their persistence. In theory, all pricing anomalies tend to disappear in efficient markets soon after they have been found, but many researches state that the most known anomalies still exist in the market. (Pätäri & Leivo 2017, 80.)

These anomalies are often behind the additional factors in multifactor models. Researchers usually look for factors that have been persistent over time and have strong explanatory power over a broad range of stocks. There are three main types of factors: macroeconomic, statistical, and fundamental. Macroeconomic factors include measures related to, for example inflation, GNP and the yield curve. Statistical factor models identify factors using statistical methods, where the factors are not pre-specified in advance. The mostly widely used factors today are fundamental factors. Fundamental factors capture stock characteristics such as industry membership, valuation ratios, and technical indicators. (Bender et al. 2013, 4.)

The stocks that are related to these stock market anomalies, tend to gain better returns than other stocks in the market on average (for example, small stocks vs. large stocks). This better return can be measured as an absolute return or risk-adjusted return, depending on the definition. (Pätäri & Leivo 2017, 80.) Some researchers believe that only the anomalies, which capture risk-adjusted extra returns, are real anomalies in the market as this is an indication that the anomaly cannot be entirely explained by the higher riskiness of stocks, and there must be some irrational explanation behind it as well.

The most popular factors are value, size and momentum, which were originally represented by Fama and French (1992; 1993; 2004) and Carhart (1997). In the following subsections, these fundamental factors as well as some macro-economic factors, that have a strong academic importance, are introduced.

### *2.2.1 Size Effect*

The size effect is perhaps the most investigated anomaly in finance. According to the risk-return relationship, small firms should, on average, provide higher returns than large firms since they are riskier. However, some researchers have found that small firm stocks actually generate excess returns even after adjusting for risk. This finding indicates that the size effect is inconsistent with efficient market theory. (Patel 2012, 653.) Many researchers have questioned the existence of the size effect recently. It has been stated that whereas the effect has mostly disappeared from developed markets, it still remains rather strong in developing markets. (Rutledge et al. 2008, 117.)

The size effect was originally reported by Banz (1981). He investigated the effect with a time period from 1926 to 1980. His evidence showed that small capitalization firms earned higher stock returns on average than large firms. Reinganum (1981) also discovered the size effect in his study which showed that small firms have higher returns even after adjusting for risk via the CAPM. This indicated that the null hypothesis, that the CAPM was correct and the market efficient, should be rejected. In addition, Reinganum (1982) found that during the years 1964−1978, the average return for small capitalization firms exceeded the return for large firms by more than 0.1 percent per day and over 30 percent per year.

Since these pioneering articles, size effect has been investigated by various researchers. For example, Fama and French (1992) investigate U.S. stocks from 1963 to 1990 and support the existence of the size effect and later include the size factor in their three-factor model. Marquering et al. (2006) examine CRSP data from 1960 through 2003 and find that small firms generated higher returns than large firms in the later years of their study from 1999 to 2003. Bauman et al. (1998) examine the size effect in international markets from 1986 to 1996 and find support for the size effect in Europe, Australasia and Canada. Mills and Jordanov (2003) discover the size effect in the London Stock Exchange in 1982−1995. Their results showed that small firms had significantly greater excess returns than large firms. Hwang et al. (2014) also find strong evidence that the size effect exists in UK markets when using the time period of 1985−2012.

Various different explanations for the size effect have been suggested during the years. Traditional explanations are risk-based and usually related to, for example, operational, financial, liquidity and default risks. Higher operational risk is explained by the less diversified product base, less sophisticated technology and lower customer loyalty. These operational risks lead to greater financial risk exposure, which leads to, for example, higher borrowing costs. (Pandey & Sehgal 2016, 46.) Another risk-based explanation is the higher liquidity risk related to small firms. Amihud (2002) find that small firm returns are sensitive to variation in market liquidity. Amihud and Mendelson (1986) and Liu (2006) also find out that the illiquidity of small stocks is related to the size effect. Small firms are also more sensitive to changes in economic conditions, which can be one explanation for higher premiums (Merton 1973; Chen et al. 1986). Vassalou and Xing (2004, 832) among many others argue that the size effect is the result of the higher distress and default risk faced by small capitalization firms.

If these risk-based explanations are the whole truth behind the size effect, one question that arises is if the size effect is an anomaly in the market at all as the higher returns are only compensation for the higher risk. For example, Chen (1983) discover that the size effect is captured by factor loadings of the APT and firms with different sizes do not have significantly different average returns after adjusting for factor risks. This means that the higher risk of small companies is the explanation for size effect and the market is therefore efficient. Chan et al. (1985) confirm Chen's findings. They examine the size effect in NYSE stock data from 1953 to 1977 and find that the average return for the smallest size portfolio was 1.513% whereas for the largest size portfolio it was only 0.558%, due to the small firms' higher covariations with changing business conditions. This indicates that the size effect arise from the higher risk and is not an anomaly in the market. (Chan et al. 1985, 456−457, 463.)

Another differing view to the risk-based explanations argues, based on behavioral finance literature, that the size effect is caused by the absence of rational investors who usually drive prices to equilibrium. Markets are dominated by naïve investors who just follow market trends or irrationally use past information for the future. Usually these investors would rather overreact than underreact to past information. (Lakonishok et al. 1994.) Dissanaike (2002) shows that the size effect simply indicates an investor's overreaction. Daniel et al. (1998) provide an explanation for the size effect based on the investors' overconfidence and self-attribution bias driving stock prices from their fundamental values. These behavioral explanations challenge the efficient market hypothesis, but there is no clear consensus in the research on whether these explanations are more likely to be true than the rational explanations described earlier.

One common explanation for the size effect is the role of the January effect. Keim (1983) is one of the first researchers who showcases this relationship. He investigates all the NYSE and AMEX firms from 1963 to 1979 and reports that almost half of the annual difference between returns on small and large capitalization firms occurred in January. Blume and Stambaugh (1983, 388) use the time period of 1963−1980 and find an even stronger significance on the January effect than Keim: the size effect averaged about 0.60 percent per day in January and almost zero in other months. Kim and Burnie (2002) also find the January small firm effect for the period of 1976−1995 as well as Patel (2012, 658) for both developed and emerging markets in 1996−2010.

Recently, the size effect has been questioned among researchers. For example, Chan et al. (2000), Amihud (2002) and more recently van Dijk (2011) in his literature review find that the size effect has disappeared after the early 1980s. Moreover, Cederburg and O'Doherty (2015) use U.S. data from 1963 to 2011 to find that the size effect relates only to differences between the returns of micro and small size firms.

Some studies show, that the premium did not disappear but has actually changed sign. For example, Dimson and Marsh (1999) examine the size effect in the UK from 1955 to 1997 and do not find small firm premiums after the year 1987. In contrast, they find negative size premiums in the later years of their study. Al-Rjoub et al. (2005) supports Dimson and March and state that there is a size reversal in the US stock market: their results show that large firms had higher returns than small firms in average from 1970 through 1999. However, it has also been argued that a reversed size effect during certain periods does not necessarily imply that the positive size premium has disappeared. Indeed, over long periods of time, small stocks on average generate greater risk adjusted returns than large stocks. (Chaibi et al. 2015, 35.) Patel (2012, 654−655, 659) examined the size effect in developed and emerging markets with Russell stock indices from 1996 to 2010 and finds that the size effect and the reverse size effect no longer exists in the stock market.

One common finding is that the size effect might have disappeared from developed markets, but still affects returns in developing markets. For example, Rutledge et al. (2008) find the size effect from the Chinese stock markets over the 6-year period of 1998 to 2003. On the contrary, Wu (2011) studies the size effect in the Chinese stock market in 1992–2009 and finds no significant size effect. Sehgal et al. (2014) examine the size effect in developing markets and find the size effect in India, South Korea and Brazil. Using data from 2003 to 2015, Pandey and Sehgal (2016) confirm the presence of a strong size effect in the Indian stock market as well.

### *2.2.2 Value effect*

The value effect is one of the most widely studied stock return anomalies, beside the size effect, and has received substantial attention from both academicians and investment practitioners. The value effect refers to the phenomenon of higher book-to-market equity, or value stocks, earning higher returns than lower book-to-market equity, or growth stocks, on average (Lee et al. 2014, 166). The value premium has been shown to be present, when various ratios are used: P/E (Basu 1977), book-to-market value (Fama & French 1992), and sales-to-price ratio (Barbee et al. 1996).

Basu's (1977) paper is one of the first researches that document the value effect. He measures value with E/P portfolios on a risk-adjusted basis in the US market. Later Fama and French (1992) find that the E/P factor actually consists of size and B/P premiums, and therefore, include the B/P factor in their famous three-factor model instead of E/P. More recently, Chan and Lakonishok (2004) use the MSCI EAFE index to find the value effect in developed non-U.S. countries. Dimson et al. (2003) discover a strong value premium in the UK for the period of 1955–2001 and that the value premium exists within both the small-cap and large-cap universe. Black and Fraser (2003) investigate the value effect in the US, UK and Japan and find the value effect in those markets for the time period of 1975−2000. Fama and French (2012) also find international evidence to support the value effect, when they examine four regions: the US, Europe, Japan and Asia Pacific in their fresher paper.

Explanations for the value effect fall into two main categories: security mispricing and risk compensation, though there is no clear consensus on which explanation dominates (Richardson et al. 2010, 50). Security mispricing is related to behavioral finance literature and implies market inefficiency. It argues that judgement errors made by irrational investors cause the systematic underpricing of value stocks and overpricing of growth stocks. (Lee et al. 2014, 166−167.) Lakonishok et al. (1994, 1542) suggest that investors often become over-excited about stocks that have done well in the past and overprice these "glamour" stocks, while overselling badly preforming stocks leading to an underpricing of value stocks. Hwang and Rubesam (2013, 2369, 2376) find also that value stocks are not riskier than growth stocks and state the value effect to be caused by investor overconfidence, which leads to overreactions, especially in noisy markets.

The second explanation, risk compensation, connects the value premium to risk and implies market efficiency. From this perspective, value stocks are riskier than growth stocks, and higher returns of value stocks are compensation to investors for bearing higher risk. (Lee et al. 2014, 166−167.) Fama and French (1995, 132) suggest that value stocks have persistently low profitability, which leads to higher risk. Chen and Zhang (1998, 501) state that distress risk, i.e., higher financial leverage and greater earnings uncertainty, causes the value effect while Campbell and Vuolteenaho (2004) suggest that value effect comes from the higher cash-flow betas of value stocks compared to growth stocks. Finally, Galsband (2012) suggests that value stocks are more sensitive to downside risk than growth stocks.

A third possible explanation for the existence of the value premium relies on the data snooping bias and other biases in data (Conrad et al. 2003). It argues that the persistence of the value effect is due to transaction costs. Because arbitrage is costly, any systematic mispricing cannot be quickly and completely traded away as arbitrage costs exceed arbitrage benefits. (Ali et al. 2003, 356.)

Loughran (1997) and Dhatt et al. (1999), among others, report that the value premium is strongest for small-cap firms. Houge and Loughran (2006, 16−17) however, examine value and growth index funds and find that small-cap value funds realize insignificantly lower annual returns than small-cap growth funds. Thus, there is no value premium for small-caps, which can result from the bid-ask spread, transaction costs, and/or the price impact of trading. Finally, Phalippou (2008, 46) finds that most of the value premium comes from stocks that have low levels of institutional ownership.

Although there is no clear consensus among researchers on the reasons behind the value effect, only few articles have questioned the existence of the anomaly. Some researchers, however, have argued that the value effect might have weakened after the 1990s or even disappeared. (Chung et al. 2016, 124.) For example, Li et al. (2009) do not find the value effect in the UK over the 1975 to 2001 period. Abhyankar et al. (2009) also indicate that no significant difference is found between the returns of value and growth portfolios in the UK, France, Germany and Italy. Chung et al. (2016) also discover that the value premium of the Australian and New Zealand markets has become weak in the sample period of 1998–2014.

### *2.2.3 Momentum effect*

The momentum effect is an anomaly based on the finding that stocks that have performed well in the past will continue to perform well in the future. The momentum effect has been first documented by Jegadeesh and Titman (1993). They examine the time period of 1965–1989 and find that the strategy of buying stocks that have had high returns in the previous 3–12 months (winners) and selling short stocks that have had low returns (losers), generated approximately 1% of abnormal return per month.

Later Jegadeesh and Titman (2001) confirm the existence of the momentum effect. Griffin et al. (2003) discover the momentum effect in the US and international markets. Scowcroft and Sefton (2005) examine the momentum effect in international markets by using the stocks of the MSCI World Index in 1992–2003 and find a strong premium. Fresh research done by Lim et al. (2018) states that the momentum effect still persists in the market, when examining the time-series momentum in US markets from 1927 to 2017.

The momentum anomaly has been found in a great number of researches, but some empirical results have also questioned its existence. Hwang and Rubesam (2015) use US data from 1927 to 2010 to state that the momentum premium has been profitable from the 1940s to the 1960s and again from the 1970s to the 1990s, but not since then. They also find that these results are robust to different momentum strategies and asset pricing models. Bhattacharya et al. (2017) do not find significant momentum effect in the US markets after the late 1990s either. By contrast, Barroso and Santa-Clara (2015, 112) suggest that the momentum has not disappeared, and that bad momentum performance is rather due to the high-risk episodes of the past ten years.

The momentum effect is one of the most difficult anomalies to explain rationally. Many other explanations, mostly related to behavioral finance, have been suggested during the years, but there is no clear consensus among researchers, what is driving the anomaly (Scowcroft & Sefton 2005, 64). Fama and French (1996) bring up the "embarrassment" of their three-factor model because it cannot explain momentum. In addition, they suggest three different explanations for the anomaly: the results are data specific (the anomaly disappears when many out-of-sample researches are made), investors' underreaction to information, or errors in the three-factor model.

The first explanation can be rejected because many researches have documented the anomaly in many different geographical areas and in different time periods. The second explanation, investors' underreaction, has been found in various studies (see e.g.

Jegadeesh & Titman 2001; Chen & Zhao 2012; Lim et al. 2018). Underreaction is related to the conservatism bias: information asymmetry makes investors react to good or bad news more conservatively, which means that investors are slow to update their prior beliefs when new information occurs. Therefore, new information does not completely reflect into prices at first but will adjust later. This leads to delayed price adjustment, underreaction and to the existence of the momentum premium. (Scowcroft & Sefton 2005, 77.)

Although most explanations are related to behavioral finance, some researchers have suggested rational explanations for the momentum effect. Barroso and Santa-Clara (2015, 113) state that momentum returns include a significant crash risk, as they have a very high excess kurtosis and pronounced left skew. This means that momentum returns are volatile and can drop very fast. In addition, Sadka (2006, 27) finds that liquidity risk could be one explanation for the momentum effect. Since momentum can be seen as an investors reaction to news about stock, momentum associated returns are sensitive to shocks in the market-wide information asymmetry environment: they outperform during months of positive liquidity shocks and underperform during months of negative liquidity shocks.

### *2.2.4 Macroeconomic factors*

Many researchers have investigated the relationship between the stock market and macroeconomic factors. Most of the previous literature supports the idea that movements in the stock market have an impact on the main macroeconomic factors and vice versa. (Jareño & Negrut 2016, 325.) For example, Flannery and Protopapadakis (2002) conclude that there is a clear relationship between the stock market and macroeconomic variables. One explanation behind macroeconomic risk factors is that certain macroeconomic news is released at a prescheduled time, so news release dates are known a long time beforehand, even though investors can obviously not know, what the news will be. Therefore, if stock prices are reacting to this news, the risk of holding stocks that are affected by the news is realized. (Savor & Wilson 2013, 343−344.)

The reason, why research has lately focused mostly on other than macroeconomic factors, is that while most of the existing literature supports the idea of a connection between macro factors and stock returns, there is only a little empirical support for this relationship (Flannery & Protopapadakis 2002, 751). For example, Maio and Philip (2014) examine six macroeconomic factors in 1964–2010 and find that including macro factors does not significantly improve the fit of multifactor models.

Many researches have also found that when fundamental factors are added to the model, macro factors are no longer significant. This indicates that macro effects are already included in fundamental factors. The reason for this can be that fundamental factors are expressed in portfolio returns so they are constructed to mimic economy wide risk factors and can be viewed then as factor-mimicking portfolios (FMPs). Based on the literature, a model with FMPs will almost always outperform a model with real economic factors. (Ericsson & Karlsson 2003, 12, 18.) For example, Hooker (2004, 382−383) examines two macro factors, the short rate and the expected GDP change and shows that when market premium is included, the short rate is no longer significant, and when a full set of financial variables is included, GDP change turns insignificant as well. Similar findings have also been found by Ferson and Harvey (1991) and Petkova (2006).

Many different macroeconomic factors have been suggested to have an effect on stock returns. Among the most examined factors are long and short interest rates, yield spread, default spread, inflation related factors, unemployment rate, industrial production and oil price related factors.

*Interest rate* factors include, for example, short and long-term interest rates, yield spreads and credit/default spreads and are perhaps the mostly commonly used macroeconomic factors in research. Short-term interest rates are often measured by one or three-month Treasury bill rates and calculated by the end-of-period return of the bill. Qi and Maddala (1999) examine the one-month Treasury bill rate as a macroeconomic factor in the US market and find it to be significant and negatively correlated with the stock market in 1954–1992. Jareño and Negrut (2016) use long-term interest rates as a macro factor to find that they are significantly priced in the time period of 2008–2014 in US markets.

The yield spread (or term spread) is calculated by the difference between short- and long-term interest rates, typically 10, 20 or 30-year T-bonds and one- or three-year T-bills. Kaneko and Lee (1995) find the yield spread to be significantly priced in the US market in 1975–1993. Earlier Chen et al. (1986) had made a similar finding in 1953–1983. Czaja and Scholz (2007) examine the relationship between term structure and stock returns in Germany in the period of 1974–2002 and find that term structure predicted stock returns in all industries, especially in financials and utilities. However, Kang et al. (2011) do not find it to be priced in the time period of 1963–2005.

The spread between high and low-graded bonds can be calculated as a difference between Baa and Aaa graded bonds (default spread) or as a difference between Baa and long-term government bonds, for example the 30y T-bond, (credit spread). Chen et al. (1986) examine the credit spread in 1953–1983 and find that the factor is significantly priced in the US market. Ando (2009) includes also credit spread in his factor model and find that it does not seem to be priced anymore in the US in the time period of 1990–2004. Kang et al. (2011) examine the default spread among other macroeconomic variables in 1963–2005 but do not find it to be priced in the markets.

*Industrial production* and *the growth rate of industrial production* are widely examined macroeconomic factors. The relationship between these variables and stock market returns has been found to be positive: higher prices in the stock markets are associated with higher values in industrial production. The finding seems to be very intuitive as good news to the financial economy often means good news to the real economy as well. (Jareño & Negrut 2016, 329.) The industrial production factor is often measured by the Industrial Production Index (IPI) that measures the productive activity of the industrial sector (excluding construction). For example, Cheng (1996) find that industrial production affects stock returns in the time period of 1965–1988 in the UK and US, and Jareño and Negrut (2016) confirm this finding in their study from 2008–2014 in US markets. Qi and Maddala (1999) examine the growth rate of industrial production (measured as a logarithmic number) and find a relationship to exist between industrial production and stock returns, but one that seemed to be negatively correlated. Kaneko and Lee (1995) find a significantly priced growth rate of industrial production in the US market in 1975–1993. Flannery and Protopapadakis (2002) do not find industrial production to be significant in the time period of 1980–1996.

*Inflation* related macro factors typically include different consumer price indexes as well as a pure inflation number. Their relationship with the stock market is uncertain, because it can vary according to the needs of the economy (Jareño & Negrut 2016, 329). Inflation is typically measured by the Consumer Price Index (CPI) or Producer Price Index (PPI). Qi and Maddala (1999) examine the relevance of the inflation growth rate with PPI, measured as a logarithmic number, and find it to be significant in US markets in 1954–1992. Flannery and Protopapadakis (2002) use both CPI and PPI in their study to find that both inflation measures are significant in capturing the stock market returns in US markets in 1980–1996. Chen et al. (1986) and Kaneko and Lee (1993) use CPI and find it significant in their studies. However, Jareño and Negrut (2016) who also use CPI in their study to predict stock market returns, do not find it to be significant in their fresher time period of 2008–2014.

*The Unemployment rate* is perhaps not as widely examined as other macro factors represented here, but it has been found to have some explanatory power in the stock market. Cheng (1996) finds the factor in US and UK stock markets in 1965–1988, Flannery and Protopapadakis (2002) in US markets in 1980–1996 and Jareño and Negrut (2016) in US markets in 2008–2014. The relationship between the unemployment rate and stock returns is more complex than with other macro factors. It can be seen that unemployment is negatively related to stock returns, because a rising unemployment rate is bad news for the economy and stock markets react to that by falling. However, because higher employment leads to a better economic situation, it can cause higher inflation and higher interest rates that may actually decrease the value of shares. This indicates that the relationship can also be positive. (Jareño & Negrut 2016, 328−329.) Jareño and Negrut (2016) find the relationship to be negative in their study.

*Oil price* and *oil price change* are also common international macro factors. One explanation behind their existence can be investors' under-reactions to oil news (Narayan

& Gupta 2015, 2). The oil price factor is often measured by changes in the crude petroleum producer price index. Cheng (1996) finds a significant oil price change effect in UK and US markets in 1965–1988. Narayan and Gupta (2015) use a long time period of 1859–2013 to find that oil price changes predict stock returns in US markets. Their results show that both positive and negative shocks are significant but negative changes more so in predicting stock returns. On the contrary, Chen et al. (1986) examine the oil price effect in 1953–1983 and do not find it significant in US markets, and neither do Kaneko and Lee (1995) in the time period of 1975–1993.

### *2.2.5 Multifactor models*

Based on the additional risk factors that have been found in the markets, researchers have tried to construct a multifactor model that could explain stock market returns more accurately than the simple one-factor CAPM. One of the earliest contributions to multifactor models is Roll and Ross's (1980) article on Arbitrage Pricing Theory (APT). They state that stock returns are affected by various firm-specific and macroeconomic factors instead of only one risk factor. APT model is derived from the following equation (Roll & Ross 1980, 1078, 1085):

$$E(r_i) - r_f = \lambda_1 b_{i1} + \cdots + \lambda_k b_{ik} + e_i,$$

where the left side of the equation is the excess return for the asset $i$ calculated as a difference of the expected return $E(r_i)$ and the risk-free return $r_f$. The right hand side of the equation consists of the non-zero constants $\lambda_1, \dots, \lambda_k$, which are factors in the asset pricing model, the coefficients $b_{i1}, \dots, b_{ik}$, which are factor loadings for these factors in all assets $i$, and $e_i$ is the error term. Roll and Ross do not specify the number of these factors or which they are.

The second important step in constructing multifactor models is Fama and French's (1992, 1993) researches, which extend the simple one-factor CAPM. They include the size and value effects that have been found earlier in the market with market premium in their asset pricing model. This three-factor model is perhaps the most examined multifactor model since its development and has been used by both academicians and practitioners. The three-factor model can be derived as follows:

$$E(r_i) - r_f = a + b_i(r_m - r_f) + s_i SMB + h_i HML + e_i,$$

where the excess return for the asset $i$, $E(r_i) - r_f$, is the sum of three factors multiplied with the factor loadings for the specific factor for the asset $i$ and the constant $a$ and the error term $e_i$. The first factor, $r_m - r_f$, is the market premium, $SMB$ the size premium and $HML$ the value premium. The coefficients $b_i$, $s_i$ and $h_i$ are the factor loadings for the asset $i$.

Carhart (1997) adds the momentum factor to the three-factor model and this four-factor model is another very widely examined model in research. The four-factor model is defined as follows:

$$E(r_i) - r_f = a + b_i(r_m - r_f) + s_i SMB + h_i HML + p_i MOM + e_i,$$

where $p_i$ is the factor loading for the momentum factor for asset $i$ and $MOM$ the momentum factor.

Both the three- and four-factor models have been examined in many different markets during the years and have received quite promising results. For example, Tai (2003) examines the four-factor model in US markets in 1953–2000 and finds that all the factors are significantly priced in the markets. Drew (2003) examines the three-factor model in the 90s in Hong Kong, Korea, Malaysia and the Philippines and finds that the three-factor model can predict stock returns in all markets. Das and Barai (2016) find that the four-factor model can predict stock returns in the time period of 2000–2013 in Indian markets. Xie and Qu (2016) use the time period of 2005–2012 to find that the three-factor model fit well into Chinese stock markets. Gaunt (2004) reports significant improvement in the explanatory power of the three-factor model in the Australian stock market in 1991–2000 compared to the CAPM.

More recently, however, many researchers have started to question these models and search for better multifactor models (e.g. Fama & French 2015; Skočir & Lončarski 2018). Many different factors, some perhaps more relevant than others, have been found in the research. Therefore, there are hundreds of different combinations of factors being used to form multifactor models in research. Some of the most popular fundamental factors as well as the set of macro-economic factors were described earlier and are tested later in this thesis. To date, there is still no consensus on a multifactor model that can predict stock market returns better than others.

# 3 BAYESIAN APPROACH IN MULTIFACTOR MODELS

## 3.1 Bayesian framework

### *3.1.1 Advantages and limitations*

If portfolio selection is implemented based on the multifactor model discussed in the previous chapters, a set of factors that capture the asset return distribution must be selected. Many different approaches have been introduced for this purpose, but there is no consensus in existing literature on what is the best way to select factors and evaluate their relevance. (Cremers 2002, 1223.)

Two competing approaches of statistical inference are mainly used in existing literature, the classical and Bayesian approach, which differ from each other by the notion of probability. In the classical framework (based for example on the Wald test or the likelihood ratio), which can be seen as a benchmark in existing asset pricing research, probability is defined as a *p*-value, which refers to the limit of relative frequency. In other words, how likely it is the model will be rejected with a given significance level. The competing framework, the Bayesian approach, defines probability as the degree of belief on the values of parameters. (Harvey & Zhou 1990, 221; Puga et al. 2015, 277.)

The Bayesian model selection approaches can provide several advantages compared to other methodologies (Berger et al. 2001). While the generally used classical methods to estimate linear multifactor models fail to lead to sensible conclusions, the Bayesian estimation approach allows for both parameter uncertainty and the instability of factors and risk premia deliver encouraging results (Bianchi et al. 2017, 111).

One classical method to test asset pricing models, which can be expanded to the multifactor world, is the two-stage procedure developed by Fama and MacBeth (1973). In the first stage, the factor betas are estimated for all the factors included in the model by using time-series regressions from historical excess returns on the assets and selected factors. In the second stage, the cross-sectional regressions are run for each of the periods by using ex-post realized excess returns to evaluate the equilibrium restrictions. Here, an intercept should be equal to zero if the model is correctly specified. (Bianchi et al. 2017, 112.)

In the classical method, the model is either rejected or accepted based on the result of a hypothesis test. However, it is not clear what this outcome tells about the usefulness of the model in decision making. If the model is not rejected, should it be seen as absolute truth? If the model is rejected, is it worthless for decisionmakers? Many researchers have criticized the standard Fama-MacBeth method because of its limitations to capture many aspects of both the model and the data, as well as its relatively simplistic view of the usefulness of models. (Pástor 2000, 179.)

Another difficulty associated with the classical method of hypothesis testing is the problem of determining an appropriate significance level (McCullock & Rossi 1991, 147). Searching for variables with the largest t-statistics puts all the weight on one specific model. This could be problematic, as it clearly ignores a very important issue, the researcher´s uncertainty about the correct model, which can lead to overconfident inferences and decisions that are riskier than one thinks they are. (Hoeting et al. 1999, 382; Cremers 2002, 1223; Hooker 2004, 380−381.)

Because of these difficulties, many researchers have stated that it can be reasonable to assume that asset pricing models are neither perfect nor useless. This is also constructed inside the definition of the model, which describes the model as a simplification of reality. Even if the data fails to reject the model, the decision maker may not automatically want to use the model as the absolute truth. At the same time, the view that rejected models are entirely worthless seems rather extreme. The decision maker may want to use the model at least to some degree even if the data rejects the model. (Pástor 2000, 179.)

A Bayesian approach has developed to avoid many of these difficulties with the direct calculation of posterior model probabilities. The decision maker is not forced to either accept or reject a null hypothesis but can represent the strength of sample evidence in a reasonable probability metric. (McCullock & Rossi 1991, 147.) When the decision maker has strong faith in a model's pricing ability, his optimal portfolio can exhibit significant differences from that of another individual with, for example, equally strong faith in an alternative pricing model (Pástor & Stambaugh 2000, 336).

Another question in the existing literature is which factors are relevant for asset pricing and should be chosen for multifactor models. Theory provides very little guidance to this and some widely used factors, like the Fama-French factors, have received a lot of critique nowadays as many researches have questioned their existence. This parameter uncertainty has been taken into account in the Bayesian approach and motivates the use of the framework. (Hooker 2004, 380.)

In addition to the parameter uncertainty, decision makers are also uncertain about which model to use. In the Bayesian approach, all the individual models are evaluated by their posterior model probabilities (Pástor & Stambaugh 2000, 361; Bianchi et al 2017, 123). In the Bayesian method, all the possible combinations of factors are calculated resulting in $2^k$ models, when $k$ different factors are considered (Hooker 2004, 380−381). This is the main advantage of the Bayesian framework compared to the classical approach, where the uncertainty of the models is ignored. If one wants to reduce the risk related to model uncertainty even more, the model averaging methodology can be used. One model averaging method is Bayesian model averaging (BMA), which calculates the model posterior probabilities for each competing model and uses these probabilities as weights in the overall multifactor model. (Hoeting et al. 1999, 383.)

The Bayesian framework is attractive from both a sample theoretic and frequentist standpoint. It is a well-known fact, that given a specific significance level, the probability of committing a Type II error approaches zero in the classical framework of hypothesis testing, as the sample size approaches infinity. However, the probability of committing a Type I error in the classical framework does not approach zero, even in large samples. By contrast, Bayesian tests can be shown to be completely consistent, so that the probability of both Type I and Type II errors disappears asymptotically. (Avramov & Chao 2006, 302.)

The classical methods, for example the two-stage Fama-MacBeth approach, have some other statistical drawbacks. It has been shown that second-stage multivariate regression, which is used to test equilibrium, suffers from obvious generated regressor (error-in-measurement) problems. This is caused by the rolling window beta estimates (estimated in the first state of time-varying regression) that are used for cross-sectional regressions in the second stage. If cross-sectional estimates for the betas covary with the underlying but unknown risk premia, they may easily yield biased and inconsistent estimates of the risk premia themselves. Unfortunately, this covariation is extremely plausible. For example, during business cycle downturns both the sizes of betas and the unit risk prices increase only because recessions are characterized by higher systematic uncertainty as well as by lower “risk appetite”. In addition, using rolling windows in two-stage regression to capture parameter instability is not only ad hoc but also inefficient. This results from the lack of specific parametric forms, which makes testing for time variation very dependent on hard-to-justify choices of rolling window length. (Bianchi et al. 2017, 112.)

Another possible problem with the classical framework is data snooping. This refers to the use of data analysis in finding statistically significant patterns from data, when in fact there are no real effects. This is caused by multiple separate statistical tests on data and by paying attention only to significant results, instead of stating a single hypothesis before the analysis and then testing it. Many researches have stated that the success of specific factors in existing multifactor research could partly result from data snooping. In the Bayesian approach, although the choice of variables (prior) still suffers from data snooping, the problem is minimized in posterior probabilities, because as demonstrated in the literature, the posterior of a specific model depends more on its Bayesian criterion than the prior, when the sample size is large. (Tsai et al. 2010, 110.) Moreover, the Bayesian approach limits the data-snooping problem that arises from using only the best model, which is a result of ignoring the uncertainty of models, as it takes uncertainty into account and compares all possible models simultaneously. (Cremers 2002, 1224; Bianchi et al. 2017, 123.)

One of the advantages in Bayesian methods is that they do not contain the problem of diluting data information as a result of including too many regressors, which is a common issue in other models (Fernandez et al. 2001, 1). Bayesian model selection approaches favor simpler models over more complex ones when the data provides roughly comparable fits for the models. Usually more complex models provide better fit to the data but at the same time the problem of overfitting increases. To avoid this problem, a penalty term is often used in non-Bayesian methodologies (for example in AIC), so that when the complexity of a model increases, the penalty term also increases. However, there is no consistency among researchers on which penalty term is the best. This is another reason why the Bayesian approach is favorable, as it does not need a penalty term. (Berger et al. 2001, 138−139.)

Bayesian model selection is consistent. This means that, if one of the tested models is actually the true model, then Bayesian model selection will (under very mild conditions) guarantee the selection of this true model if enough data is observed. This is a valuable feature, because the use of most classical model selection tools, such as *p*-values and AIC, does not guarantee consistency. It is sometimes argued, however, that consistency is not really a relevant concept, because none of the possible models are likely to be exactly true. Some researchers have shown though, that even when the true model is not among those being considered, Bayesian model selection will choose the one that is closest to the true model. (Berger et al. 2001, 138.)

In Bayesian model selection the method is straightforward and theoretically the same, regardless of the number of models that are under consideration and it can easily deal with even a large number of models (Berger et al. 2001, 139; Ericsson & Karlsson 2003, 4). The Bayesian method does not require nested models. Two models are nested, if the first model can always be obtained from the second model by constraining some of the parameters. This is not the case for Bayesian methods, as they consider all the possible combinations of potential factors. (Berger et al. 2001, 139.)

The Bayesian method does not require normal distributions but allows for many different distributions to be used (Berger et al. 2001, 139). This is relevant in finance as the asset returns are rare normally distributed. In addition, time-series variables in Bayesian methods can be either stationary processes or unit root processes. When utilizing the Bayesian approach, the pretesting of data for unit roots and co-integration, or the transformation of data to receive stationarity, does not have to be done. (Avramov & Chao 2006, 302.)

There are some problems with utilizing the Bayesian model. First, choosing the right distributions for priors can fail, which is a serious threat since posterior probabilities are very sensitive to the choice of prior distributions. With little, or under the absence of, prior information, choosing prior distributions is very tricky. Non-informative priors are one solution in these situations, but it is proved that the rules of probability no longer apply if improper non-informative priors are used on model-specific parameters. Another problem with non-informative priors is that they can yield indeterminate marginal likelihoods (Ericsson & Karlsson 2003, 6). These issues related to choosing the prior are discussed more carefully later in this paper. The influence of prior distribution can be difficult to identify for posterior model probabilities as well. (Fernandez et al. 2001, 1−3.)

The calculation of posterior probabilities can also be challenging, because of the high dimensional nature of parameter spaces, which is analytically intractable. However, some methods have been developed lately to answer this problem, for example, the development of Monte Carlo numerical integration. (Harvey & Zhou 1990, 222; Berger et al. 2001, 142.)

### *3.1.2 The Bayesian theorem*

The basic idea of the Bayesian approach is to first determine the probability distribution that reflects the current state of knowledge. When new data becomes available this probability distribution is then updated in light of the new data. (Puga et al. 2015, 277.) Probabilities are calculated for each variable (or factor) and all the possible combinations of these variables are evaluated simultaneously.

In Bayesian portfolio selection, several parameters are optimized within the asset return prediction model. First, a set of adequate factors is selected within the predictive distribution. These parameters are then used to optimize portfolio selection. In other words, an optimal multifactor model is constructed by finding the factors that contribute most significantly to the asset return prediction. (Ando 2009, 551.)

Bayesian statistics assume that population parameters are unknown quantifiable random variables, and that the uncertainty related to them can be described by probability distributions. First, subjective probability statements, or priors, of these parameters based on previous experience and knowledge, are made. Prior probability refers to the probability of the hypothesis before current evidence, or data, is observed. (Puga et al. 2015, 277.) In other words, the model can be seen as a point of reference around which an investor can center his prior beliefs specified with varying degrees of confidence in the model (Pástor 2000, 180; Pástor & Stambaugh 2000, 335).

When prior probabilities are formed, posterior model probabilities, that update the prior model probabilities, can be calculated. The posterior probability means the probability of a hypothesis given the observed evidence, usually new data. This provides the probability, where new information has been taken into account and priors have been updated. (Puga et al. 2015, 277.) The posterior probability of model basically tells the probability in which the model is the “true” model.

After that, posterior probabilities for individual factors, can be also examined. Posterior probability of factor shows the probability of including specific variables to the model and it is calculated by summing up the posterior probabilities from each of the possible combinations of variables in which it is included. If a variable's posterior probability is greater than its prior, it means that the data provides some support for the variable's significance. Respectively, if the posterior probabilities are below priors, there is evidence against the significance of that variable. (Hooker 2004, 381.)

The Bayesian approach is based on the Bayes' theorem which uses previously introduced concepts of prior odds and posterior probabilities:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

where $P(A|B)$ is the posterior probability, $P(A)$ is the prior probability, $P(B|A)$ is the likelihood, which is a conditional probability of B given A, and $P(B)$ is the marginal likelihood or "model evidence", which measures how well the model fits the data (Rouder & Morey 2018, 1). With the marginal likelihood of different models, it is possible to take into account the overall (in-sample) statistical performance of models, and not only their asset pricing plausibility (Bianchi et al. 2017, 122).

Bayesian methods of hypothesis testing and model selection, lead to a test statistic that takes into account an in-sample goodness-of-fit measure and favors simpler models over complex ones. Since a higher-dimensional model always fits the data at least as well as the models with fewer variables, it is important to take into account model complexity in the test statistic, because overfitting the data may be a problem if models are evaluated only on the basis of goodness of fit. Thus, the Bayesian approach that automatically penalizes model complexity (without a separate penalty term) helps reduce the inclusion of excess, useless factors in linear factor models. (Avramov & Chao 2006, 294.)

Many researchers have argued that using only the best subset selection method (including standard Bayesian methods), which means allocating all weight in one selected "optimal model", is too risky. It is better to combine, or average, a number of competitive models with appropriate weights. This has resulted in the development of one model combination method, Bayesian model averaging (BMA), where the overall model is the weighted average of the individual models with weights given by posterior probabilities. (Hooker 2004, 381; Tsai et al. 2010, 110.) For this reason, it works with more sophisticated return processes and can incorporate learning into portfolio choices. (Ando 2009, 551.) Although BMA has reached theoretical usefulness in factor selection and has been extensively investigated in literature, empirical studies in the field of finance are relatively limited (Tsai et al. 2010, 110).

In addition to BMA, there are some other non-Bayesian approaches in the existing literature that take model uncertainty into account. One of the most known approaches is Akaike's information criterion (AIC) that approximates model weights. Some additional methodologies, including methods related to neural networks, machine learning and COLT (computational learning theory), generally focus on point prediction, often in the context of supervised learning, which makes them very different from the Bayesian framework. (Hoeting et al. 1999, 398.) These optional methodologies and their applicability to factor selection are not, however, examined in this thesis.

### *3.1.3 Prior information*

The specification of appropriate priors is essential in Bayesian methodology as posterior model probabilities in the context of model uncertainty are very sensitive to prior information (Fernandez et al. 2001, 1, 5; Cremers 2002, 1228−1230). There are many different options on how to specify prior distribution in the existing literature, but in financial (and multifactor) research, informative priors are perhaps the most widely used.

Some researchers, however, for example Fernandez et al. (2001, 1, 5), use non-informative (or diffuse or weak) priors. The paper focuses on the situation where a large number of sampling models are considered and only a little prior information is available to provide the prior specification that could be used in these cases. The prior structure presented is based on improper priors that are widely accepted as non-informative (and not natural-conjugate) priors for scale parameters. The paper suggests that the prior specification examined is suitable for the purposes of model averaging, where prior information is lacking. (Fernandez et al. 2001, 5, 8−9, 20.) Bianchi et al. (2017, 116) also use non-informative priors for all their multifactor variables. Harvey and Zhou (1990) use both informative and non-informative priors in their multifactor tests.

The non-informative prior was first introduced into Bayesian multivariate analysis by Geisser and Cornfield (1963). It is a prior of "minimal prior information" and it produces results similar to those produced by a classical framework. (Harvey & Zhou 1990, 228.) When using non-informative priors, posteriors will be very similar to classical maximum-likelihood results. However, replacing non-informative priors by informative priors allows for additional inference of information in the data. (Cremers 2006, 2962.)

There is a quite clear consensus in the literature that it is better to use informative priors, where it is possible. It is shown, for example, that improper non-informative priors yield indeterminate marginal likelihoods. (Ericsson & Karlsson 2003, 5.) There are also other advantages for using informative priors: to avoid some other problematic properties of non-informative priors, to allow for prior model mispricing, and to adjust estimates of inefficiency measures for model size when comparing different models. In addition, instead of only testing "exact efficiency", which means that the model restrictions should hold exactly, informative priors can explicitly deal with cases where models are only expected to hold approximately, for example, due to measurement problems. (Cremers 2006, 2952, 2963.)

If an investor has a certain prior, for example, of expected returns, then applying the model to the data should lead to posterior probabilities of expected returns that are more in line with the evidence in the data. This is because of the fact that data generally moves priors of inefficiency statistics toward the values implied by the model. (Cremers 2006, 2977−2978.)

One way to form informative priors is that investors have certain views about models. Typically, the degree of confidence ranges from a dogmatic belief in the model to a belief that the model is useless. Naturally, when the degree of skepticism about the model grows, the optimal allocation moves further from benchmark portfolios toward data-based (factor) allocation. When an investor does not have the specific view about the goodness of competing models, an equal prior probability can be also set for all the models. (Pástor 2000, 181.)

Hierarchical (multilevel) models are dominant to modern Bayesian statistics for both conceptual and practical reasons. On the conceptual side, hierarchical models are a more "objective" approach because the parameters of prior distributions are estimated from the data rather than requiring them to be specified using investors' subjective information and views. On the practical side, hierarchical models are flexible tools for combining information and the partial pooling of inferences. When using hierarchical models, however, hyperparameters are required, and these must be given their own prior distribution. (Gelman 2006, 515−516.) Cremers (2002) states that because of the lack of consensus in literature on what the important predictors are, it is better to use hierarchical models and investigate the priors from the data.

There are several ways to build the hierarchical models in Bayesian statistics. Typically, conjugate priors are used for the parameters. In Bayesian probability theory, if the posterior distributions belong in the same probability distribution family as the prior probability distribution (for example, all the exponential distributions), the prior is called a conjugate prior for the likelihood function. One example of a prior which is hierarchical and conjugate, is the prior based on inverse-gamma distribution. The inverse-gamma distribution is the reciprocal of the gamma distribution and it belongs to the exponential family. It remains only marginally studied and used in practice but is, however, quite popular choice in the Bayesian statistics for the variance parameter $\sigma^2$ in a normal distribution. The inverse-gamma distribution has two positive parameters, the shape parameter, which controls the height, and scale parameter, which controls the spread. The main difference between the gamma and inverse-gamma distributions is that the inverse-gamma distribution is always positive, while the gamma distribution can be zero. (Llera & Beckmann 2016, 1.) Cremers (2002) and Bianchi et al. (2017, 116) use the conjugate inverse-gamma distribution for $\sigma^2$.

The Wishart and inverse-Wishart distributions are also widely used distributions in Bayesian hierarchical models and have hierarchical and conjugate priors for the variance-covariance matrix parameter $\mathbf{\Sigma}$ of a multivariate normal distribution. The inverse-Wishart distribution is basically the multivariate extension of the inverse-gamma distribution. This means that whereas the inverse-gamma distribution has the conjugate prior for the variance parameter $\sigma^2$ of a univariate normal distribution, the Inverse-Wishart distribution extends conjugacy to the multivariate normal distribution. Therefore, these distributions are used in situations, where a larger dimensionality is involved. (Gelman 2006, 516−518; Nydick 2012, 10−12.) The inverse-Wishart distribution is determined as follows:

$$\sum \sim W^{-1}(\boldsymbol{\Psi}, v),$$

where $\mathbf{\Psi}$ is a real-valued positive-definite (scale) matrix and $v$ the degree of freedom. The scale matrix can be thought as a sums-of-squares matrix from the multivariate normal distribution. The degree of freedom is the number of observations that have been observed prior to collect the data, or, the number of observations on which the sums-of-squares matrix $\mathbf{\Psi}$ is based. (Nydick 2012, 10−12.)

Cremers (2006, 2952, 2969) uses the normal inverse-Wishart priors in his multifactor model analysis. He finds that these types of informative priors greatly improve the clarity and controllability of posterior analysis, because it leads to the prior and the posterior to have the same distributional forms. Informative inverse-Wishart priors are also used in, for example, Pástor (2000) and Pástor and Stambaugh (2000), in both papers the degree of freedom $v$ is fixed at 15. This basically means that the prior contains only about as much information as a sample of fifteen observations. McCullock and Rossi (1991) also adopt the inverse-Wishart prior for their variance-covariance matrix. They consider, however, the varying degree of freedom $v$.

Ando (2009) uses the inverse-Wishart priors for the variance-covariance matrix $\mathbf{\Sigma}$ and another conjugate prior, the matric-variate normal distribution, for mean value $\mu$. The matric-variate normal distribution (also called matrix variate normal distribution) is a generalization of the multivariate normal distribution, like the inverse-Wishart distribution, and another widely used distribution in the Bayesian statistics. The parameter $\mathbf{X}$ is said to have a matric-variate normal distribution,

$$\boldsymbol{X} \sim MN_{KN}(\boldsymbol{A}, \boldsymbol{\Sigma}, \boldsymbol{V}), \tag{1}$$

where $\mathbf{A}$ is the $K \times N$ mean matrix, and $\mathbf{\Sigma}$ (with the dimension $N \times N$) and $\mathbf{V}$ $(K \times K)$ are positive definite symmetric matrices that are proportional to the variance matrix of the rows and columns of $\mathbf{X}$ respectively, if $vec(\mathbf{X})$ is multi-variate normal

$$vec(\boldsymbol{X}) \sim N(vec(\boldsymbol{A}), \boldsymbol{\Sigma} \otimes \boldsymbol{V}), \tag{2}$$

where $vec(\mathbf{A})$ denotes a vectorization of matrix $\mathbf{A}$ that stacks its columns one under another in a $K \times N \times 1$ vector, $\mathbf{\Sigma} \otimes \mathbf{V}$ is the covariance matrix for $vec(\mathbf{A})$, where $\otimes$ denotes the Kronecker product of matrices, and proportionality of two variables means the situation where two variables are connected to a constant, or more specifically, either their ratio or product yields a constant. For example, if $y = c \times x$, $y$ is said to be proportional to $x$ with the proportionality constant $c$. (Iranmanesh et al. 2010, 34; Woźniak 2016, 371−374.)

Table 1 shows the different types of priors used by different researchers. It clearly shows the relevance of informative priors as compared to non-informative. However, there are some studies that have received reliable results with non-informative priors. One way to deal with non-informative priors, has been presented by Avramov and Chao (2006). They split the total sample of data into two subsamples: a training sample and a primary sample. The idea is to combine a (perhaps improper) non-informative prior density with data from the training sample to obtain a posterior density. After that, this posterior density is used as a prior density and combined with the data from the primary sample to calculate model probabilities for model comparison. (Avramov & Chao 2006, 300.)

Table 1 Prior types in different papers

The table here describes the common prior types in the different Bayesian multifactor papers. The non-informative prior is a prior of "minimal prior information" and it produces results similar to those produced by a classical framework. Informative prior is used when the investor has a certain view about the prior information and it can be based on the certain distribution or specified separately for every competing model.

| | Informative conjugate | | | Non-informative |
|---|---|---|---|---|
| | Inverse-gamma distributed | Inverse-Wishart distributed | Matric-variate distributed | |
| Harvey & Zhou 1990 | | | | x |
| McCullock & Rossi 1991 | | x | | |
| Pástor 2000 | | x | | |
| Pástor & Stambaugh 2000 | | x | | |
| Fernandez et al. 2001 | | | | x |
| Cremers 2002 | x | | | |
| Cremers 2006 | | x | | |
| Ando 2009 | | x | x | |
| Bianchi et al. 2017 | x | | | x |

In addition to these priors, some other (mostly other conjugate) priors can be used. For example, Harvey et al. (2010, 8, 15−16) assume conjugate prior densities for the unknown parameters, which uses a priori normal for mean $\mu$ and a priori Wishart for the variance-covariance matrix $\mathbf{\Sigma}$.

## 3.2 Bayesian model averaging (BMA) for factor selection

This thesis is mostly focused on the Bayesian model averaging method (BMA), which is examined, for example, in Ando's (2009) paper. Compared to the other Bayesian models, this method takes into account not only the impact of parameter uncertainty but also uncertainties that arise from model specification itself. (Ando 2009, 551.)

The method of model averaging is relatively simple. In the context of stock market prediction, the idea is first to choose interesting factors that could have an impact on stock returns. After this, all the possible combinations of these potential factors (competing models) are evaluated. BMA methodology can be used in two ways: to find the best model based on the highest posterior probability of the model and use it as a multifactor model, or, use all the competing models instead of just the best model by weighting the models with their posterior probabilities. (Tsai et al. 2010, 110; Steel 2011, 33).

If the first approach is used, the model with the highest posterior probability is chosen, so it will most likely be the "true" model among the competing models. If the posterior probability is relatively high for the best model, and there is a big difference in posterior probabilities between the first- and second-best models, it is not as relevant to use all the models, as one model clearly stands out from the others. (Raftery et al. 1997, 182.)

In the second approach, all the competing models are used, not only the best one. The overall model is the average of these competing models. Basically, all the models have certain posterior probabilities that they are "true" models and these probabilities are used as the weights. The overall model is then the weighted average of these models. The number of these models in the overall model can be enormous, and the amount grows rapidly if the number of factors is increased. Because of this, the number of models in BMA can be restricted with Occam's Window. (Raftery et al. 1997, 180, 182.)

The steps of running a BMA algorithm are now described more carefully, mostly based on Ando's (2009) article. From a given set of $k$ factors, all $2^k$ different models are evaluated by the extent to which they describe the data as given by the posterior model probabilities. Hence, all the possible models of $M_j$ where $j = 1, \dots, 2^k$, and which all include $0 \leq k_j \leq k$ factors are considered.

First, suppose that an asset's returns for the model $M_j$ are given by the multifactor model

$$\boldsymbol{r_t} = \boldsymbol{\alpha_j} + \boldsymbol{\Gamma_j}'\boldsymbol{f_{jt}} + \boldsymbol{\varepsilon_{jt}}, \quad t = 1, \dots, n, \tag{3}$$

where $\boldsymbol{f_{jt}} = (f_{1t}, \dots, f_{p_j t})'$ is a $p_j$-dimensional vector of factors, and $\boldsymbol{\varepsilon_{jt}} = (\varepsilon_{j1t}, \dots, \varepsilon_{jmt})'$ is an $m$-dimensional noise vector with the mean $\mathbf{0}$ and variance $\mathbf{\Sigma_j}$. Vector $\boldsymbol{\alpha_j} = (\alpha_{j1}, \dots, \alpha_{jm})'$ and matrix $\boldsymbol{\Gamma_j} = (\boldsymbol{\beta_{j1}}, \dots, \boldsymbol{\beta_{jm}})$ consist of unknown parameters, where $\boldsymbol{\beta_{jk}} = \left(\beta_{jk1}, \dots, \beta_{jkp_j}\right)'$ is a $p_j$-dimensional vector of factor loadings.

The previous model can also be written in a matrix form

$$\boldsymbol{R} = \boldsymbol{X_j B_j} + \boldsymbol{E_j},$$

where $\mathbf{R} = (r_1, \dots, r_n)'$, $\mathbf{X_j} = (\mathbf{1}_n, \mathbf{F})$, $\mathbf{F} = (f_1, \dots, f_n)'$, $\mathbf{B_j} = (\alpha, \mathbf{\Gamma}')'$, and $\mathbf{E_j} = (\boldsymbol{\varepsilon_1}, \dots, \boldsymbol{\varepsilon_n})'$ with $\boldsymbol{\varepsilon_t} \sim N(\mathbf{0}, \mathbf{\Sigma})$.

To calculate the posterior probabilities for all the models $M_j$, the prior distributions and certain likelihood functions are needed. First, the prior distributions for the factor sensitivities $\mathbf{B_j}$ and the covariance matrix $\mathbf{\Sigma_j}$ are formed to get the prior distribution

$$\pi(\boldsymbol{\theta_j} | M_j),$$

where $\boldsymbol{\theta}_{\mathbf{j}} = (vec(\mathbf{B_j})', vec(\boldsymbol{\Sigma}_{\mathbf{j}})')'$ is the parameter vector of model $M_j$. There are many distributions that have been previously used as priors in BMA literature in stock markets as described earlier. In Ando's (2009) research, prior distribution is formed by using the matric-variate normal distribution for mean $\mathbf{B_j}$ and inverse-Wishart distribution for the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{j}}$, which is a quite common way in the literature. Based on the Equation (1), the mean parameter $\mathbf{B_j}$ is said to have the matric-variate normal distribution

$$\pi(\boldsymbol{B}_j|\boldsymbol{B}_0, \boldsymbol{\Sigma}_j, \boldsymbol{A}) = (2\pi)^{-\frac{(p+1)\times m}{2}} |\boldsymbol{\Sigma}_j|^{-\frac{p+1}{2}} |\boldsymbol{A}|^{-\frac{m}{2}} \times e^{-\frac{1}{2}tr\{\boldsymbol{\Sigma}_j{}^{-1}(\boldsymbol{B}_j - \boldsymbol{B}_0)'\boldsymbol{A}^{-1}(\boldsymbol{B}_j - \boldsymbol{B}_0\}},$$

if $vec(\mathbf{B_j})$ is multivariate normal with the mean $vec(\mathbf{B_0})$ and the variance $\boldsymbol{\Sigma}_{\mathbf{j}} \otimes \mathbf{A}$, as described in the Equation (2). Then, $\mathbf{B_0}$ is the $(p+1) \times m$ mean matrix and $\boldsymbol{\Sigma}_{\mathbf{j}}$ (with dimensions $m \times m$) and $\mathbf{A}$ (with dimensions $(p+1) \times (p+1)$) are the positive definite symmetric matrices that are proportional to the variance matrix of the rows and columns of $\mathbf{B_j}$. (Woźniak 2016, 377.) Finally, $m$ and $p$ are dimensions of matrices in Equation (3).

The prior distribution for the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{j}}$ is inverse-Wishart distribution and defined as follows:

$$\pi(\boldsymbol{\Sigma}_j|\boldsymbol{\Lambda}_0, v_0) = \frac{|\boldsymbol{\Lambda}_0|^{\frac{v_0}{2}}}{2^{\frac{mv_0}{2}}\boldsymbol{\Gamma}_{\boldsymbol{m}}\left(\frac{v_0}{2}\right)} |\boldsymbol{\Sigma}_j|^{-\frac{v_0+m+1}{2}} \times e^{-\frac{1}{2}tr(\boldsymbol{\Lambda}_0\boldsymbol{\Sigma}_j{}^{-1})},$$

where $\boldsymbol{\Sigma}_{\mathbf{j}}$ is again the $m \times m$ matrix proportional to the variance matrix of $\mathbf{B_j}$, $\boldsymbol{\Lambda}_{\mathbf{0}}$ is a positive definite $m \times m$ scale matrix, $\boldsymbol{\Gamma}_{\boldsymbol{m}}$ is the multivariate gamma function, $v_0$ is the degree of freedom, and $m \geq v_0$ and $v_0 > p - 1$. The matrix $\boldsymbol{\Lambda}_{\mathbf{0}}$ can be understood as a sums of squares matrix with a multivariate normal distribution and $v_0$ as the number of observations that are observed prior to collecting data, or, alternatively, the number of observations on which the prior sums-of-squares matrix $\boldsymbol{\Lambda}_{\mathbf{0}}$ is based. (see, e.g., Nydick 2012).

After determining the prior probabilities, posterior probabilities are calculated based on the priors and the new information $\mathbf{D}$ which becomes available, usually the new data updates the information. To get the posterior probabilities, the likelihood and marginal likelihood functions are then calculated for model $M_j$. The likelihood function is

$$L(\boldsymbol{D}|\boldsymbol{\theta}_j, M_j) = \prod_{t=1}^{n} f(r_t|\boldsymbol{f}_{jt}, \boldsymbol{B}_j, \boldsymbol{\Sigma}_j)$$
$$= (2\pi)^{-\frac{nm}{2}} |\boldsymbol{\Sigma}_j|^{-\frac{n}{2}} \times e^{-\frac{1}{2}tr\{|\boldsymbol{\Sigma}_j|^{-1}(\boldsymbol{R}_j - \boldsymbol{X}_j\boldsymbol{B}_j)'(\boldsymbol{R}_j - \boldsymbol{X}_j\boldsymbol{B}_j)\}},$$

where $\mathbf{D} = (r_1, \dots, r_n)$ is the new data available, and all parameters are known, so it is easy to calculate the likelihood function for the model $M_j$. Marginal likelihood, sometimes called model evidence, measures, how well the model (and the prior) fits the data. It is called marginal, because it marginalizes (integrates) the variable $\boldsymbol{\theta}$. It is obtained by means of the likelihood function and prior distribution as follows:

$$\pi(\boldsymbol{D}|M_j) = \int L(\boldsymbol{D}|\boldsymbol{\theta_j}, M_j)\pi(\boldsymbol{\theta_j}|M_j)d\boldsymbol{\theta_j}, \tag{4}$$

where the marginal likelihood $\pi(\boldsymbol{D}|M_j)$ is calculated by integrating the joint distribution of the data and the parameters over the whole parameter space, $L(\boldsymbol{D}|\boldsymbol{\theta_j}, M_j)$ is the likelihood function of model $M_j$ and $\pi(\boldsymbol{\theta_j}|M_j)$ is the prior distribution. Both parameters are conditional on $M_j$, the set of all models being considered. The role of marginal likelihood is to make sure that the posterior is a valid probability by making its area sum to 1. Therefore, it only affects to the posterior by scaling it up or down, but does not change the shape of the posterior. (Woźniak 2016, 368.)

The posterior probability for the model $M_j$ can be determined as follows:

$$\pi(M_j|\boldsymbol{D}) = \frac{\pi(\boldsymbol{D}|M_j)\pi(M_j)}{\sum_{j=1}^{J}\pi(\boldsymbol{D}|M_j)\pi(M_j)}, \quad j = 1, \dots, 2^k, \tag{5}$$

where $\pi(M_j)$ is the prior probability for the model $M_j$, and $\pi(\boldsymbol{D}|M_j)$ its marginal likelihood. The denominator of equation is the sum of these prior probabilities and marginal likelihoods of all models $M_j$. This probability is calculated separately for each competing model $M_j$ and it tells the probability of a specific $M_j$ being the "true" model. The model with the highest probability can be selected.

In addition to the models, posterior probabilities can be calculated for different factors. This is useful if one wants to examine the relevance of selected factors and their features. The posterior probability of a factor tells the probability for which this specific factor is included in the multifactor model and it is calculated by

$$\Pr(factor\ f_k\ is\ included \mid \boldsymbol{D}) = \sum_{j=1}^{K} \pi(M_j|\boldsymbol{D}) \times \delta(M_j, k), \tag{6}$$

where $\pi(M_j|\boldsymbol{D})$ is the posterior probability of a specific model and $\delta(M_j, k)$ is 1 if factor $k$ is included in model $M_j$ and 0 otherwise. These separate model probabilities $\pi(M_j|\boldsymbol{D})$ for factor $f_k$ are summed through all the models $M_j$, where $j = 1, \dots, K$. As a result, the posterior probability for a specific factor $f_k$ is achieved. After that, the equation is calculated separately for all other factors $f_k$. (Ando 2012, 94.)

If one wants to reduce model uncertainty, posterior probabilities of different models can be used as weights, resulting in an overall model, which is the weighted average of individual models. The implementation of this methodology is quite difficult, as all models contain different factors, and because the number of models can be enormous. For example, if we have ten explanatory variables, $2^{10} = 1024$ different models need to be combined. A common solution to this problem in BMA is the use of Occam's window algorithm, which was originally developed by Madigan and Raftery (1994).

The basic idea of the algorithm is to average over a set of parsimonious, data-supported models with two basic principles: first, if a model predicts the data far worse than the model which provides the best predictions, then it is effectively discredited and should no longer be considered. Second, the algorithm excludes complex models which receive less support from the data than their simpler counterparts. (Hoeting et al. 1999, 384.)

Occam's window averages then over a set of likely models and ignores unlikely models. Unlikely models are models that are much less likely than the most likely model, in other words, all models $M_j$ satisfying

$$\frac{\pi(M_{max}|\boldsymbol{D})}{\pi(M_j|\boldsymbol{D})} > C,$$

where $M_{max}$ is the most likely single model. This most likely model is compared separately with all models $M_j$ to find the models that lie in Occam's window. Basically, if the model $M_j$ differs from this most likely model by more than the window size $C$, it is excluded. Finally, all these unlikely models are then excluded from the posterior distribution. The number of models in Occam's window increases as the value $C$ decreases. This means that more models are excluded from posterior distribution, so the overall number of models decreases when $C$ decreases. (Bollen & Long 1993, 172−173; Ando 2009, 555.)

There is no clear consensus on the optimal size of Occam's window among researchers, but many studies use a number less than 25 (Raftery et al. 1997, 182). Often the size is set to 20, which is consistent with the popular .05 cutoff for $p$-values (Madigan & Raftery 1994, 1536). This means that if the value is greater than 20, when comparing the model $M_j$ with the most likely model $M_{max}$, the model $M_j$ is included in Occam's window, and therefore excluded from the posterior distribution. It also means that the excluded model $M_j$ is 20 times less likely than the most likely model $M_{max}$ (Bollen & Long 1993, 170).

There are also several other ways to determine the size of Occam's window $C$. For example, in Ando's (2009) article the Bayesian predictive information criterion (BPIC) is developed for this purpose. BPIC can be calculated with the following equation:

$$BPIC_{BMA} = \sum_{j=1}^{K} \pi(M_j|\boldsymbol{D}) \times \left[-2 \int \log L\left(D|\boldsymbol{\theta_j}, M_j\right) \times \pi(\boldsymbol{\theta_j}|\boldsymbol{D}, M_j) d\boldsymbol{\theta_j} + 2q_j]\right],$$

where the parameters $\boldsymbol{\theta_j}$ are integrated over the posterior distribution, $\pi(M_j|\boldsymbol{D})$ is posterior probability, $L(\boldsymbol{D}|\boldsymbol{\theta_j}, M_j)$ is the likelihood function of model $M_j$, $\pi(\boldsymbol{\theta_j}|\boldsymbol{D}, M_j)$ is the posterior distribution, and $q_j$ is the dimension of $\boldsymbol{\theta}_\mathbf{j}$. When we search for the optimal size of Occam's window $C$, various values of $C$ are chosen and then a BPIC-score is calculated for each of them. The optimal value of $C$ is the one which achieves the smallest BPIC-score among the candidates.

Another commonly used method to reduce the number of models, is a Markov chain Monte Carlo model composition ($MC^3$), which uses the standard MCMC method for the purpose of determining an optimal set of models (Hoeting et al. 1999, 385).

## 3.3 Previous findings and literature

### *3.3.1 Previous literature on factor selection with the Bayesian model averaging method*

Various Bayesian methodologies have been developed by statisticians and mathematicians but their implementation into finance is quite limited in the literature. Only a few papers have utilized the Bayesian approach specifically to the problem of factor selection in stock markets. However, previous research has recommended Bayesian methods for use in factor selection and has introduced the various advantages of the methodology.

Bayesian factors can be used in many ways (for example, the choice of the prior) and there are different methods based on the Bayesian approach. Although in this thesis the Bayesian model averaging method is utilized in the empirical research, the previous research is introduced from a slightly wider perspective to get a better understanding of the Bayesian research field. First, the previous literature related directly to BMA is discussed. After this, the Bayesian research field from the perspective of selecting factors for multi-factor models is examined from a wider perspective.

Ando (2009) examines two Bayesian methods (the empirical Bayes method and the Bayesian model averaging method) and compares them with each other and with the traditional mean-variance approach. The sample period is from 1990 to 2004 and all NYSE, AMEX and NASDAQ stocks are divided into twelve industry portfolios and included in the empirical research. Three Fama-French and three other factors are included as the potential factors: market premium, size- and value factors, the difference between

Moody’s Baa corporate yield and the long-term U.S. government bond yield, growth of the S&P’s common stock price index and growth of U.S. Consumer Price Index (CPI).

Monte Carlo simulations are used to investigate the performance of the proposed portfolio selection methods. Both Bayesian approaches outperform the mean-variance method and the BMA method achieves the best performance and the largest Sharpe ratio. Forecasting accuracy is also examined for each method by calculating the out-of-sample mean squared forecasting error (MSFE), the mean forecasting error (MFE) and the mean absolute error (MAE) of each method. Among the three methods examined, the BMA method predicts future returns best, as it has the smallest out-of-sample MSFE in each sub-period. Similar results are achieved with the mean absolute error (MAE). Finally, the MFE indicates that BMA gives the least biased results. (Ando 2009, 558−561.)

Ando (2012) repeats the previous study for the time period of 1990−2006. This time the same Bayesian methods, empirical Bayes method and BMA method, are used for factor selection in US and Japanese markets. In this article, three Fama-French factors are used as potential factors and ten industry portfolios as test portfolios.

Through the whole time period, both Bayesian methods outperform out-of-sample the standard mean-variance method based on the Sharpe ratio. The difference between these Sharpe ratios are, however, not significant, when analyzed using the paired *t*-test. (Ando 2012, 90, 98.) In addition, the results show that selected factors are not constant over time, except for the market premium, which’s posterior probability approaches 100% throughout the time period in both US and Japanese markets. Posterior probabilities for value- and size factors depends strongly on the time period and market. (Ando 2012, 94−95.)

Tsai et al. (2010) compare the forecasting accuracy of three Bayesian model selection approaches: the Bayesian information criterion (BIC), Bayesian model averaging, and model mixing. The two latter methods are compared to BIC, which is the “traditional” Bayesian approach. The method of model mixing splits the data into two subsamples of equal size and then uses the former subsample to build up the model and the latter to evaluate the model’s forecasting accuracy with the appropriate criterion to assign different combination weights to the competing models according to their estimated forecasting accuracy. (Tsai et al. 2010, 111.) The difference to Ando’s (2009 & 2012) research is that the Bayesian methods are not compared to the traditional mean-variance method.

The value-weighted CRSP stock return index is used as the dependent variable of future stock returns. In addition, twelve predictive variables, that have been shown to have forecasting ability on future stock returns in the existing literature, are used as potential factors (see e.g. Cremers 2002). These variables are: dividend yield, E/P-ratio, share volume/price, credit spread (Baa–Aaa), the yield of a 3-month treasury bill, the change in the yield of a 3-month treasury bill, term spread (the difference between the yield on a 10-year treasury bond and a 3-month treasury bill), yield spread (the difference between the Fed Fund Rate and a 3-month treasury bill), inflation, change in inflation, the change in CPI, and the logarithm of change in industrial production. The datasets range from 1954 to 2005 and three time horizons (3-, 5-, and 8-year periods) are used so that the data is not too short to fit the linear regression model. (Tsai et al. 2010, 113.)

Tsai et al. (2010) find that all three models lead to a better prediction than a purely random process. Furthermore, the averaging and mixing methods yield a slightly higher hit rate in predicting next period returns when compared to BIC. For the 3-year time period, for example, the averaging and mixing methods, produce correct forecasts for 62.0% and 62.6% of the hit rates, which are 5.0% and 5.6% higher than the BIC approach. The improvements are even greater for the 5- and 8-year periods: both approaches yield about 7.8% better forecasts than the BIC. It is clear that the averaging and mixing methods are superior to BIC in out-of-sample predictions of hit rates. (Tsai et al. 2010, 114.)

Portfolios based on the averaging and mixing methods however, do not lead to a statistically better performance than the portfolio based on BIC. While more recent approaches to model averaging and model mixing beat the BIC in their out-of-sample hit rates, the formed test portfolios do not significantly outperform the portfolio obtained via the BIC. (Tsai et al. 2010, 109, 115.)

### *3.3.2 Previous literature on factor selection with other Bayesian methods*

Pástor and Stambaugh (2000) examine the impact of prior information with three different pricing models. The test portfolios are formed by sorting stocks by size and book-to-market. In their paper, priors are highly informative, as the investors are expected to have specific prior beliefs for all three pricing models on whether they are true or not. The purpose in this article is not to choose one pricing model over another but instead to pay attention to the economic importance of considering such a choice (Pástor & Stambaugh 2000, 335).

Cremers (2002) investigates S&P 500 excess returns with fourteen macroeconomic factors, based on the previous literature, using the Bayesian model selection approach and strongly informative priors like Pástor and Stambaugh (2000). The potential factors are lagged returns, dividend yield, earnings yield, share volume/price, credit spread, yield on a short-term Treasury bill, change in the yield on a short-term Treasury bill, term spread, yield spread between the yield on an overnight fixed income security and the short-term treasury bill, the January dummy, growth rate of industrial production, inflation and change in inflation. The research is done for the time period of 1954−1998. (Cremers 2002, 1224−1226.)

Five statistical model selection criteria are compared to each other: $R^2$, Akaike's information criterion (AIC), Schwarz’s criterion, the Bayesian information criterion (BIC), and the posterior information criterion (Cremers 2002, 1238). The prior views are set to range from skeptic to confident (Cremers 2002, 1223−1224). When looking at the optimal multifactor model, for the skeptic investor, seven variables stand out: the liquidity variable, the credit spread, the yield on a 3-month Treasury bill and its first difference, the yield spread, the first difference in the inflation variable, and the January dummy. For the confident investor also seven variables stand out. Compared to the best-performing variables for the skeptic investor, the first difference in the inflation variable is replaced with the term spread. (Cremers 2002, 1239.)

An out-of-sample forecasting analysis suggests that the Bayesian model selection criteria outperforms classical methods. The paper suggests that this is due to the stringent nature of classical tests to put 100% weight on inclusion or exclusion. In other words, conditioning on one individual model fails to take model uncertainty into account. Compared to this, the in-sample and out-of-sample results for the Bayesian analysis are consistent and show more (though only a little) evidence of predictability. (Cremers 2002, 1225.)

Pástor (2000) examines sample evidence on value and size effects and home bias from an asset-allocation perspective and uses the Bayesian framework that incorporates the investor’s prior degree of confidence in an asset pricing model, like Cremers (2002) and Pástor and Stambaugh (2000). The degree of confidence is again allowed to range from complete confidence to complete skepticism. Basically, the research examines what happens between these two extremes. (Pástor 2000, 181, 196−197.)

The results show that compared to the traditional data-based approach, optimal portfolio weights are less sensitive to sampling error and tend to have less extreme values in the Bayesian approach than in the traditional mean-variance approach. (Pástor 2000, 208−209). The results of home bias testing in the time period of 1973−1996 show that the optimal weights in the World-Except-U.S. portfolio tend to be large. Therefore, a typical U.S. investor's confidence in the domestic CAPM must be very strong to justify low holdings of foreign stocks. (Pástor 2000, 198−200.) The paper also investigates size and value effects from 1927 to 1996. Value premium in the U.S. is very robust and estimates are positive throughout the time period, while the size effect and market premium are more unstable and sometimes even negative. (Pástor 2000, 200−202.)

Cremers (2006) develops a new general framework based on the Bayesian statistical approach to examine whether the Fama-French factors are priced in an Intertemporal Capital Asset Pricing Model (ICAPM) setting. The data from 1927 to 1953 is used as a training sample to determine the choice of informative priors. The sample period is then set from 1954 to 2001. Both sets contain all stocks listed in NYSE, AMEX and NASDAQ.

The purpose of the paper is to evaluate different models by comparing how the data changes the informative prior views with four different inefficiency measures into posterior distributions. The results provide little evidence that either of the two additional factors in the Fama-French model or the momentum factor are priced in an ICAPM framework. In addition, the pricing performance of the Fama-French model is not robust to the choice of test portfolios, as it works only with BE/size-sorted portfolios. However, the data gives clear evidence of the usefulness of the market portfolio for pricing, as using the CAPM will lead to lower pricing errors than what was expected in the priors. (Cremers 2006, 2952−2954, 2993.)

Ericsson and Karlsson (2003, 2−8) examine Bayesian factor selection with fifteen factors: size, value and momentum premiums, market excess return, dividend yield, credit risk spread, change in yield on a 3-month T-bill, long term spread, short term spread, and monthly and yearly growth rate in consumption and in disposable income. The time period of the research is 1963−2002.

To summarize, strong evidence is found that a general multifactor pricing model should include the market excess return and the size and value factors. The evidence for including the momentum factor depends more on the sample used and the prior specification. In addition, the credit risk spread should be included as an additional factor and industrial production may also have some relevance in the market. (Ericsson & Karlsson 2003, 18.)

Hooker (2004) examines the predictive power of macroeconomic factors in emerging markets' equity returns using the Bayesian model selection approach developed by Cremers (2002). Eight prior combinations that cover a sensible range of views from very skeptical about predictability to quite confident, are calculated (Hooker 2004, 381). The data sample includes all MSCI Emerging Markets Free index constituent countries, with a monthly frequency from 1992 to 2002.

Three sets of factors are analyzed: one with only macroeconomic factors, one with macroeconomic factors and beta, and the third by adding several financial variables that are often mentioned in the existing literature. Six country-specific macroeconomic factors are considered as potential factors in regressions: the change in the foreign currency exchange rate vs. the US dollar, a local interest rate, the real short-term rate, the change in expected GDP growth, inflation, and the sovereign credit risk (JP Morgan Emerging Market Bond Index (EMBI) spreads). The financial factors are price momentum, P/E, P/B, downside risk, and size. (Hooker 2004, 381−382.)

The results provide strong evidence against most of the considered macroeconomic factors except for exchange rate changes. Consistent with the findings of the existing literature, however, the evidence shows strong support for several financial factors, but not beta, as significant predictors of excess returns. Momentum, P/E, and downside risk appeared to have explanatory power for emerging market equity returns. (Hooker 2004, 379−380.)

Young and Lenk (1998, 114, 116) also examine portfolio selection with the Bayesian approach. They use monthly data on 500 randomly selected stocks from the CRSP database, which is obtained for nineteen four-year intervals: [1955−1959, … ,1991−1994]. The two-factor model (market and size factor) is used for the analysis. The data from the first two years is used to form parameter estimates for a linear factor model using hierarchical Bayes and multiple shrinkage techniques. The parameter estimates are then compared to the ordinary least-square estimates to evaluate the parameter estimation accuracy of the two methods. As a result, the hierarchical Bayesian method leads to improved estimation accuracy and therefore to improved portfolio selection.

Hall et al. (2002, 2307−2308) use the Bayesian approach to test the global multifactor model. The empirical research is done with MSCI stocks from 1988 to 1998, and the number of sectors and countries is reduced by pooling them into nine geographical groups and nine sectors. Four factors are used: value, growth, total debt/book value ratio and size. (Hall et al. 2002, 2310−2314.) Summarizing the results, equities seem to not be well explained by any combination of styles (Hall et al. 2002, 2319−2320).

Harvey et al. (2010, 8, 15−16) also use the Bayesian method for factor selection and form two different models that take skewness into account. Two sets of portfolios are considered in the paper: a portfolio of four stocks (General Electric, Lucent Technologies, Cisco Systems and Sun Microsystems), and a portfolio of four equity (Russell 1000, Russell 2000, MSCI EAFE and MSCI EMF) and three fixed income (government bonds, corporate bonds and mortgage packed bonds) portfolios, all from Lehman Brothers. The first set considers daily returns from 1996−2002 and the second weekly returns from 1989−2002.

To find out which model best fit the data, Bayes factors are computed for the multivariate normal model, the skew normal model with a diagonal $\Delta$ matrix, and the skewed normal model with both a diagonal and a full $\Delta$ matrix. Summarizing the results, both skew normal models fit the data better than the multivariate normal model. Skewed models also accommodate heavy tails and can then be considered as realistic models for portfolio returns. (Harvey et al. 2010, 16, 19.)

McCullock and Rossi (1991) quantify differences between restricted and unrestricted models by testing APT with the Bayesian method. A collection of ten size portfolios for all listed AMEX and NYSE stocks from 1964−1983 are used for testing. The resulting odds ratios show that the sample evidence is quite weak and do not favor the factor model based on APT. (McCullock & Rossi 1991, 141.)

Bianchi et al. (2017) examine a stochastic volatility asset pricing model with Bayesian methods, where both betas and the prices of risk factors are time varying. The performance of Bayesian time varying betas and the stochastic volatility framework (B-TVB-SV) is compared to a benchmark Fama-MacBeth approach as well as to two other Bayesian models (Bianchi et al. 2017, 116). The time period for the research is from 1972 to

2011. Monthly data of excess returns on 23 portfolios of securities traded in the U.S., is used. Two sets of portfolios are tested, industry and size portfolios. (Bianchi et al. 2017, 111, 114.)

The superiority of the B-TVB-SV model is proved by three different ratios: the marginal likelihood, variance ratios and pricing error. The B-TVB-SV model shows high log-marginal likelihood ratio across all of the portfolios under attention and dominates in every case, especially in the case of bonds and medium and large cap portfolios. Not surprisingly, given its ad hoc nature, the classical Fama-MacBeth approach ranks last with an overall marginal likelihood. (Bianchi et al. 2017, 123.)

A correctly specified multi-factor model's variance ratio should explain most or all of the predictable variation in excess returns of test portfolios. The unexplained portion should be as small as possible. B-TVB-SV dominates other models in variance ratios as well: 80% of the predictable variation in excess returns is captured on average. (Bianchi et al. 2017, 124.) The B-TVB-SV model also leads to both the lowest average pricing error and the lowest median posterior error. Moreover, B-TVB-SV consistently dominates all the other models in all sub-samples. (Bianchi et al. 2017, 126.) Soyer and Tanyeri (2006) confirm the finding of Bianchi et al. (2017), that stock returns are described by time-varying stochastic variance models.

Avramov and Chao (2006) also test asset pricing models with time-varying risk premia by using Bayesian posterior probabilities. In this paper the method is applied to test the International Asset Pricing Model (ICAPM) and four conditional ICAPM versions for the time period of 1975−2000. Four global ICAPM models are the return differential between (1) high and low book- to-market stocks, (2) high and low earnings-yield stocks, (3) high and low cash-flow-to-price stocks, and (4) high and low dividend-yield stocks. Each model have two versions, the restricted and unrestricted, so in total, ten different models are considered. Each of the ten data generating models is assigned an equal (10%) prior probability. (Avramov & Chao 2006, 295.)

Five potential factors are considered: the excess rate of return on the world index lagged by one month, the January dummy, the term spread, the dividend yield, and the 1-month rate of interest on a Eurodollar deposit. Some of the instruments used are local. (Avramov & Chao 2006, 306.) The cross-model comparison shows that the best performing model is the ICAPM with a value premium constructed based on earnings yield. (Avramov & Chao 2006, 295.) However, when examining the country-specific market and characteristic-sorted portfolios, ICAPM with the E/P-factor is found to be the best model. The results are similar for the global portfolios. (Avramov & Chao 2006, 310−311.)

In Table 2 all the studies described previously that compare the standard mean-variance method and Bayesian methods in testing multifactor models are collected. As we can see from the table, the results for using Bayesian methodology instead of traditional methods are promising: in every case the Bayesian method outperforms the mean-variance method. This supports the usage of the Bayesian approach in factor selection for multifactor models. However, the research on this topic has been done mostly in the 90s and early 2000s, so it would be interesting to test the methodology with fresher data as well.

Table 2 Previous studies that compare the standard *p*-values based and Bayesian methodology in testing multifactor models

| | Bayesian method | Time period | Market | Result |
|---|---|---|---|---|
| **Ando (2009)** | Empirical Bayesian method, BMA | 1990−2004 | US | Bayesian method outperforms |
| **Ando (2012)** | Empirical Bayesian method, BMA | 1990−2006 | US Japan | Bayesian method outperforms, no significant difference |
| **Bianchi et al. (2017)** | Three Bayesian models with time-varying betas and stochastic volatility | 1972−2011 | US | Bayesian method outperforms |
| **Cremers (2002)** | BIC, the posterior information criterion | 1954−1998 | US | Bayesian method outperforms |
| **Pástor (2000)** | Empirical Bayesian method | 1927−1996 1973−1996 | US | Bayesian method outperforms |
| **Young & Lenk (1998)** | Hierarchical Bayesian method | Nineteen 4-year time periods from 1955−1994 | US | Bayesian method outperforms |

# 4 DATA AND RESEARCH METHODOLOGY

## 4.1 Data

The problem of constructing optimal multifactor models is examined through empirical research in U.S. stock markets. The idea is to determine several potential fundamental and macroeconomic factors and use the Bayesian model averaging methodology to form the optimal combination of these factors. Potential fundamental factors are the three factors from the Fama-French three factor model, market, size and value premium, and the momentum factor. As macroeconomic factors, changes in long- and short-term interest rates, changes in inflation rate, unemployment rate and industrial production, oil price change, yield spread, and default spread, are included as potential factors. These twelve different factors are shown in Table 3 below. All the potential factors are included based on their relevance in existing multifactor research (see, e.g., Ericsson & Karlsson 2003; Ando 2009; Tsai et al. 2010; Jareño & Negrut 2016).

Table 3 BMA factors included in the empirical research

| **Potential factor** | **Factor name in analysis** |
|---|---|
| **Panel 1: Fundamental factors** | |
| Market premium | MKT-RF |
| Size premium | SMB |
| Value premium | HML |
| Momentum | MOM |
| **Panel 2: Macroeconomic factors** | |
| Long-term interest rate change | LONG_TERM |
| Short-term interest rate change | SHORT_TERM |
| Yield spread | YIELD_SPREAD |
| Default spread | DEFAULT_SPREAD |
| Inflation rate change | INFL |
| Unemployment rate change | UNEMPLOY |
| Industrial production change | IP |
| Oil price change | OIL |

All the NYSE, AMEX and NASDAQ stocks, that have the historical data available, are divided into twelve industry portfolios and used for the empirical testing. The time period for the research is set from 2007−2018 (twelve years). The first ten years (2007−2016) are used to find the optimal factors for the multifactor model in-sample and last the two years (2017−2018) are used for testing the predictability of the formed BMA multifactor model out-of-sample. This time period is chosen because the purpose of this thesis is to examine, which factors have been relevant lately and also, because there is not as much previous research from this time period. This time period is also selected because it includes both a financial crisis and a period of expansionary monetary policy by central banks, as well as a period when the markets were recovering from a recession. Based on the existing literature, it is also beneficial to keep the time period of the data relatively short, as both academics and practitioners have found that stock returns are unstable and too long-term forecasts are meaningless (for example, a twenty years' time-period). (Tsai et al. 2010, 113.)

The data for the fundamental factors is provided by Kenneth French's data library and the data for the macroeconomic variables is collected from Bloomberg and Eikon. Monthly data is used, and portfolios are rebalanced every month-end.

To test the potential multifactor models, stocks are allocated into twelve industry portfolios to reduce idiosyncratic risk, which is a typical way of doing it in the literature (Bianchi et al. 2017, 114). The data for industry portfolios is provided by Kenneth French's data library, where the portfolios are constructed at the end of June of year *t*. Each NYSE, AMEX, and NASDAQ stock is assigned to an industry portfolio based on its four-digit SIC code. The twelve industries are: consumer non-durables; consumer durables; manufacturing; oil, gas, and coal extraction and products; chemicals and allied products; business equipment; telephone and television transmission; utilities; wholesale, retail and some services; healthcare, medical equipment and drugs; finance; and; other (mines, construction, BldMt, transport, hotels, bus service and entertainment). The names for the industry portfolios that are used in the models are described in Table 4.

Table 4 Industry portfolios and their names in the analysis

| **Industry portfolio** | **Name** |
|---|---|
| Business equipment | BUSEQ |
| Chemicals and allied products | CHEMS |
| Consumer durables | DURBL |
| Oil, gas and coal extraction and products | ENRGY |
| Healthcare, medical equipment and drugs | HLTH |
| Manufacturing | MANUF |
| Finance | MONEY |
| Consumer non-durables | NODURBL |
| Other | OTHER |
| Wholesale, retail and some services | SHOPS |
| Telephone and television transmission | TELCM |
| Utilities | UTILS |

The data for fundamental factors (market, size and value premium) as well as the momentum factor is collected from Kenneth French's data library. The size premium (SMB) and value premium (HML) are constructed using six value-weight portfolios formed on size and book-to-market. SMB (Small Minus Big) is the average return on three small portfolios minus the average return on three big portfolios:

$$SMB = \frac{1}{3}(Small\ value + Small\ neutral + Small\ growth) - \frac{1}{3}(Big\ value + Big\ neutral + Big\ Growth).$$

HML (High Minus Low) is formed by calculating the average return of two value portfolios and subtracting the average return of two growth portfolios from it:

$$HML = \frac{1}{2}(Small\ value + Big\ value) - \frac{1}{2}(Small\ growth + Big\ growth).$$

SMB and HML factors include all NYSE, AMEX, and NASDAQ stocks for which market equity data for December of $t - 1$ and June of $t$, and (positive) book equity data for $t - 1$ is available.

Market premium is the excess return on the market, which is calculated as a value-weight return of all CRSP firms incorporated in the U.S. and listed on NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month $t$, good shares and price data at the beginning of $t$, and good return data for $t$ minus the one-month Treasury bill rate.

Momentum premium is also provided by Kenneth French's data library. It is constructed by using six value-weighted portfolios formed on size and prior (2−12) returns. The portfolios are the intersections of two portfolios formed on size and three portfolios formed on prior (2−12) return. The monthly size breakpoint is the median NYSE market equity. The monthly prior (2−12) return breakpoints are the $30^{th}$ and $70^{th}$ NYSE percentiles. The momentum is calculated as the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios:

$$Momentum = \frac{1}{2}(Small\ high + Big\ high) - \frac{1}{2}(Small\ low + Big\ low).$$

All the NYSE, AMEX, and NASDAQ stocks with prior return data are included. To be included in a portfolio for month $t$ (formed at the end of last month), a stock must have a price for the end of month $t - 13$ and a good return for $t - 2$. In addition, each included stock must also have a market equity for the end of month $t - 1$ to calculate the size portfolios.

The data for macroeconomic factors is collected from Bloomberg or Eikon, depending on the availability and reliability of the data. For all interest rate related factors, the data is collected from Bloomberg (short- and long-term interest rate changes, yield spread and default spread). The Bloomberg prices are also used for the oil price change factor. Data of the unemployment rate, inflation and industrial production is collected from Eikon. Monthly data is again used to make sure that there is the same amount of observations in the time series than for the fundamental factors (120).

Because of the lag associated with the publication of some macroeconomic indicators, three variables are included in the analysis with a one-month time lag (e.g. Qi & Maddala 1999, 163-164). These three factors are changes in unemployment rate, inflation and industrial production. The rest of the variables are included in the analysis without lag.

The sources and descriptions for all the macroeconomic factors are represented in Table 5. As the yield spread (or term spread) is used the difference between the yields in maturity of 10-year treasury bond and 3-month treasury bill of US, like in Kang's et al. (2011) research. The changes in these two interest rates are also included as separate factors, long-term and short-term interest rate changes. Default spread is defined as the difference between the yields of Moody's Baa and Aaa municipal bond yield averages. Baa−Aaa spread is a quite common way to determine default spread in the literature (see, e.g., Kang et al. 2011). Inflation rate change is determined as the change in consumer price index (CPI), which is a quite common definition in previous studies (Chen et al. 1986; Kaneko & Lee 1993; Jareño & Negrut 2016). Unemployment rate change is simply the change in unemployment rate and industrial production the change in industrial production. Inflation rate change, unemployment rate change and industrial production change are all lagged by one month, so the changes in values are calculated as the difference between the index values in the time period $t-2$ and $t-1$. Lastly, Oil price change is defined as the crude oil (brent) price change.

Table 5 The sources and definitions for macroeconomic factors

| **Macroeconomic factor** | **Lag (months)** | **Description** | **Data source** | **Bloomberg ticker** |
|---|---|---|---|---|
| Long-term interest rate change | 0 | The change in the end of period returns of 10y Treasury bond from the time period *t*-1 to *t* | Bloomberg | USGG10YR Index |
| Short-term interest rate change | 0 | The change in the end of period returns for 3-month Treasury bill from the time period *t*-1 to *t* | Bloomberg | USGG3M Index |
| Yield spread | 0 | The difference in the yields to maturity of the 10y Treasury bond and the 3-month Treasury bill in the time period *t* | Bloomberg | USGG10YR Index − USGG3M Index |
| Default spread | 0 | The difference in the yields to maturity between Moody's Baa and Aaa municipal bond yield averages in the time period *t* | Bloomberg | MMBABAA2 Index – MMBAAAA2 Index |
| Inflation rate change | 1 | The change in Consumer Price Index from the time period *t*-2 to *t*-1 | Eikon | - |
| Unemployment rate change | 1 | The change in unemployment rate from the time period *t*-2 to *t*-1 | Eikon | - |
| Industrial production change | 1 | The change in Industrial Production Index from the time period *t*-2 to *t*-1 | Eikon | - |
| Oil price change | 0 | The change in crude oil (brent) price from the time period *t*-1 to *t* | Bloomberg | CO1 Comdty |

## 4.2 Methodology

The Bayesian model averaging method (BMA) is used to find the optimal multifactor model in-sample, which is constructed as a combination of twelve potential factors. First, the data for the factors is collected from Kenneth French’s data library (fundamental factors) and Bloomberg and Eikon (macroeconomic factors) and then modified to be suitable for the analysis. There are 120 observations for each potential factor, as the time period for the BMA is ten years (2007−2016) and monthly data is used. Second, twelve industry test portfolios are constructed by using Kenneth French’s data library for the same time period. After that, the BMA approach is run in eViews with the BMA add-in, which provides a front end to the R package and was written by Raftery et al. (2018).

In the BMA algorithm, the prior probabilities are first calculated by using specific prior distributions. In the BMA add-in, informative priors are used and set to 50%. This means that the prior probability is same for all competing models, so in other words, all models are "true" models in the probability of 50% before the data is added. Because twelve factors are considered as potential factors, there are 4096 different models of $M_j$ in the BMA algorithm, as $j = 1, \dots, 2^{12} = 1, \dots, 4096$.

The BMA algorithm is run separately for each of the twelve industry portfolios by including all the previously determined factors into the algorithm. For every industry portfolio, the posterior probabilities for all competing models are calculated based on the prior probabilities and marginal likelihood

$$\pi(M_j|\boldsymbol{D}) = \frac{\pi(\boldsymbol{D}|M_j)\pi(M_j)}{\sum_{j=1}^{J} \pi(\boldsymbol{D}|M_j)\pi(M_j)}, \quad j = 1, \dots, 2^k, \tag{5}$$

where $k = 12$ is the number of potential factors, $\pi(M_j)$ is the prior probability for the model $M_j$, and $\pi(\boldsymbol{D}|M_j)$ its marginal likelihood, which is calculated with Equation (4). Among a set of competing models, the preferred model is the one with the highest posterior model probability.

Because the number of competing models can be huge, the number of competing models is reduced with Occam's window. The number of models in Equation (5) decreases, when the value of Occam's window *C* decreases. Occam's window *C* can be determined in several ways and because there is no clear consensus among the researchers, what the optimal level is, the size of Occam's window is set to 20. This is a quite common level in the literature and consistent with the popular .05 cutoff for *p*-values and also the default setting of the BMA add-in. (Madigan & Raftery 1994, 1536; Raftery et al. 1999.)

Now, the optimal multifactor model for the sample period is found for all industries. It is also interesting to examine the factor-specific posterior probabilities, which tell the probability of including the specific factor in the multifactor model. The posterior probability of a factor is basically a sum of those model's posterior probabilities, where the factor is included. This means that the overall factor posterior probability takes into account the factor inclusion probability of all the models in which the specific factor is included. The posterior probability of a specific factor can be derived from the equation

$$\Pr(factor\ f_k\ is\ included \mid \boldsymbol{D}) = \sum_{j=1}^{K} \pi(M_j|\boldsymbol{D}) \times \delta(M_j, k), \tag{6}$$

where $\pi(M_j|\boldsymbol{D})$ is the posterior probability of a specific model and $\delta(M_j, k)$ is 1 if factor $k$ is included in the model $M_j$ and 0 otherwise, and $k = 12$. These factor probabilities tell, which factors seem to be relevant outside the most optimal model as well. Taking these factor-specific probabilities into account, it is easier to construct the overall multifactor model based on all of the industry portfolios.

Next, the overall multifactor model is constructed based on the best models in every industry. If the factor is included in the best model in four or more industries out of twelve, it is included in the overall model. The average posterior probabilities are also evaluated. This limit is set because results show that factors can be clearly divided into two classes: factors that seem to have at least some explanatory power, which are the ones that are included in the best model in four or more factors out of twelve, and factors that do not seem to have any or only very little explanatory power. The limit is then very naturally set at this level.

When the overall BMA multifactor model has been formed, its predictability ability out-of-sample can be tested and compared to other multifactor models. The important question is, whether all the variables are relevant to stock market predictions out-of-sample, because many studies acknowledge that the observed predictability is robust only within the sample period and not very useful in making out-of-sample predictions. (Tsai et al. 2010, 110.) The out-of-sample period is from 2017 to 2018 and since monthly data is again used, every benchmark model and the constructed BMA multifactor model have 24 observations. The benchmark models are the Fama-French three (FF3) and four factor models (FF4) as well as the Capital Asset Pricing model (CAPM). Table 6 describes the structures of the different factor models.

Table 6 An overview of the factor models under comparison

| Factor model | Number of factors | Description of factors |
|---|---|---|
| CAPM | 1 | market premium |
| FF3-factor model | 3 | previous one + size and value factors |
| FF4-factor model | 4 | previous ones + momentum factor |
| BMA multifactor model | 1−12 | possible factors: previous ones and eight macroeconomic factors |

The predictability ability one period ahead is examined for the different models using the mean squared error (MSE) criterion, which calculates the difference between actual returns and the expected return from the factor model. The formula for MSE for the industry portfolio $p$ is

$$MSE_p = \frac{1}{n}\sum_{i=1}^{n}(Y_{ip} - \hat{Y}_{ip})^2,$$

where $Y_{ip}$ is the actual return for the time period $i$ and the industry portfolio $p$, $\hat{Y}_{ip}$ is the expected return calculated for the time period $i$ and the industry portfolio $p$, $n$ is 24, which is the number of months in out-of-sample period, and $p = 1, \ldots, 12$, because twelve industry portfolios are used. MSE value is calculated separately for all four competing factor models.

To calculate the expected returns of the four different factor models, the beta coefficients, which express factor loadings, or more specifically, sensitivities for different factors, are calculated. The betas can be calculated with time-series regressions, which are done separately for all the factors in all four models and for all twelve industries $p$ separately. Rolling 10-year betas are used here, because typically in research the number of months for rolling the betas is the same as in the in-sample time period (see e.g. Ando 2009). The beta coefficients are calculated for all the 24 months in the out-of-sample time period.

First, a regression equation is introduced for the one factor model (CAPM). The Sharpe-Lintner CAPM states that the expected value of an asset's excess return, which is the difference between the asset's, or in this case the portfolio's, return and the risk-free rate, is completely explained with the market premium. The regression formula is

$$R_{pt} - R_{ft} = \alpha_p + \beta_{pM}(R_{Mt} - R_{ft}) + \varepsilon_{pt},$$

where $R_{pt} - R_{ft}$ is the excess return for the specific industry $p$ in the time period $t$, $\alpha_p$ is the intercept, $\beta_{pM}$ is the beta and $R_{Mt} - R_{ft}$ is the market premium in the time period $t$. (Fama & French 2004, 32.)

The regression equations are formed similarly for the three and four factor models:

$$R_{pt} - R_{ft} = \alpha_p + \beta_{pM}[R_{Mt} - R_{ft}] + \beta_{ps}SMB_t + \beta_{ph}HML_t + \varepsilon_{pt}.$$

$$R_{pt} - R_{ft} = \alpha_p + \beta_{pM}[R_{Mt} - R_{ft}] + \beta_{ps}SMB_t + \beta_{ph}HML_t + \beta_{pm}MOM_t + \varepsilon_{pt}.$$

Here, $SMB_t$ (Small minus Big) is the difference between the returns on diversified portfolios of small and big stocks in the time period $t$, $HML_t$ (High minus Low) is the difference between the returns on diversified portfolios of high and low B/M stocks in the time period $t$, $MOM_t$ is the difference between the returns on high prior return portfolios and low prior return portfolios in the time period $t$, betas are the slopes in multiple regression for the different factors and $R_{pt} - R_{ft}$ the excess return for the specific industry portfolio $p$ in the time period $t$. (Carhart 1997, 61; Fama & French 2004, 39.)

Finally, the regression equation is formed for the BMA multifactor model based on the regression equations above and Roll and Ross (1980) paper:

$$R_{pt} - R_{ft} = \alpha_p + \beta_{1p}f_{1t} + \cdots + \beta_{xp}f_{xt} + \varepsilon_{pt},$$

where $x = [1, k]$, $k = 12$ is the number of selected factors in the BMA multifactor model, in other words, the number of factors can be anything between 1 and 12, $f_{1t}$ ... $f_{xt}$ are the selected factors, $\beta_{1p}$ ... $\beta_{xp}$ are the slopes in multiple regression of excess return $R_{pt} - R_{ft}$ on $x$ factors and $p$ is the specific industry portfolio.

Time-series regressions to get the beta coefficients are run separately for all 24 months in the out-of-sample period. First, the beta is calculated for the first month of out-of-sample, which is January 2017. The time-period in the regression is then the previous ten years (120 months), which is from January 2007 to December 2016. Next, the time period is rolled one month forward. The beta for the second month in out-of-sample period, which is February 2017, is calculated from the time-series regression of February 2007−January 2017. The rolling of months is continued until the beta is calculated for the last month of the out-of-sample time period (December 2018).

When all the betas are calculated, the expected returns for all the factor models in the all industries $p$ can be calculated. Basically, the expected return for the factor model in period $t$ is the individual factor's beta coefficient times its expected factor realization and taking a sum of these to add all factors in the model, added to the expected risk-free return. The beta coefficients are calculated in the previous stage and the previous month's risk-free rate (the risk-free rate in the period $t - 1$) is used to calculate the expected risk-free rate. This can be done, because the risk-free rates do not change very fast, as they are highly depended on monetary policy, so the previous month's rate is the best guess for the following month's risk-free rate. The expected return for the specific factor model is then derived from the equation

$$E(R_{pt}) = r_{f,t-1} + \beta_{1pt}f_{1t} + \cdots + \beta_{xpt}f_{xt}, \tag{7}$$

where $x$ is the number of factors in the specific factor model, $r_{f,t-1}$ is the risk-free rate in the previous month, $\beta_{1pt}$ ... $\beta_{xpt}$ are the beta coefficients for the time period $t$ and factor $x$ for the industry $p$, and $f_{1t}$ ... $f_{xt}$ are the expected factor realizations for the time period $t$ and factor $x$. The number of factors in the factor models is $x = 1$ for the CAPM, $x = 3$ for the FF3-factor model, $x = 4$ for the FF4-factor model, and $x = [1, k]$, where $k = 12$ for the BMA multifactor model. Basically, the expected return $E(R_{pt})$ is calculated separately for all four factor models, as well as all industries $p$.

The expected factor realizations (or factor "returns") are calculated in two ways. First, by taking the average of the previous ten years factor realizations. This is perhaps the more reliable way of estimating expected factor realizations, because they can change very fast, so it might be better to use the average of a longer time period as the guess for the next period's factor realization. In this scenario, therefore, the expected factor realization for the time period $t + 1$ is calculated as the average of the previous ten years factor realizations up to time $t$, and for the next time period $t + 2$, the time period of ten years is rolled one month forward. Then the expected factor realizations are determined for all out-of-sample time periods and can be used in Equation (7).

Second way to calculate the expected factor realizations is to use the factor realization from the previous period $t - 1$ for the expected factor realization for the time period $t + 1$. So, the calculation is similar to the risk-free rate. However, this way of determining the expected factor realization is expected to be worse than the first way, because factor realizations seem to be quite volatile.

This method to calculate the factor realizations, or risk premiums, assumes that there is predictability in the premiums, which is against the efficient market hypothesis. In other words, historical premiums are used for predicting the risk premiums in the future. Another way could be, for example, to estimate the risk premiums using Fama-MacBeth-approach, but it is not examined in this thesis.

When the expected factor realizations and beta coefficients are calculated, the expected returns for all comparable factor models can be derived with the Equation (7). After that, the MSE values can finally be calculated to compare the forecasting ability of the competing factor models. The MSE values are calculated first for the separate industries and the overall score is the average of these values.

# 5 EMPIRICAL RESULTS

## 5.1 Descriptive analysis

In Table 7 the most important descriptive statistics are represented for twelve factors. The time period is from 2007 to 2018, so each time-series has 144 monthly observations. The mean, medium, variance and standard deviation values are multiplied with 1 000 for the short-term interest rate change variable (*SHORT_TERM*), and with 10 000 for the long-term interest rate change variable (*LONG_TERM*), to show more decimals.

The market premium (*MKT-RF*) has the highest mean (0.6%) among the fundamental factors. The size premium (*SMB*) and momentum premium (*MOM*) have little bit lower means (both 0.1%) in the time period. The lowest mean is with the value premium (*HML*), -0.2%. The value premium is negative in the sample period, which indicates that growth stocks actually performed better than value stocks. Standard deviation for *MOM* is 4.7%, which means that it is the most volatile fundamental factor. The lowest standard deviation among the fundamental factors is 2.3% for *SMB*. The standard deviation for *MKT-RF* is 4.4% and *HML* 2.7%.

Among the macroeconomic factors, Oil price change (*OIL*) has the lowest minimum value (-33%), and *DEFAULT_SPREAD* the highest (0%). *OIL* also has the highest maximum value (29.0%), making it the most volatile macroeconomic factor. The same observation can be made with the standard deviation, which is very high, 8.9%, for *OIL*.

The highest mean among the macro factors is with *YIELD_SPREAD* (2.0%), and the lowest with the unemployment rate (*UNEMPLOY*) (-1.0%). The sign of the unemployment factor has been under discussion in the literature. Here, the mean for unemployment rate change is negative, which is similar to some previous studies (e.g. Jareño & Negrut 2016). The means are also quite high for *DEFAULT_SPREAD* (1.2%), *OIL* (0.3%), and the change in inflation (*INFL*) (0.2%). The means for industrial production change (*IP*) and short- and long-term interest rate changes (*SHORT_TERM* and *LONG_TERM*) are closer to zero (0.1%, 0.0019% and 0.00001%).

As already mentioned, the highest standard deviation among the macro factors is with *OIL* (8.9%). The second highest is with *UNEMPLOY* (2.8%). The lowest standard deviation is with *LONG_TERM* (0.00095%), so it is the least volatile factor among the macro factors.

Table 7 Descriptive statistics for potential fundamental and macro-economic factors

In this table, the descriptive statistics for twelve factors are provided. The first column N is the number of observations, min and max are minimum and maximum values of factors, mean is the average value, median is the middle of the set of numbers, var is the variance and std.dev. is the standard deviation. Skewness, kurtosis and Jarque-Bera are related to the normality test.

| | N | Min | Max | Median | Mean | Var | Std.dev | Skew | Kurt | J-B |
|---|---|---|---|---|---|---|---|---|---|---|
| MKT-RF | 144 | -0.172 | 0.114 | 0.011 | 0.006 | 0.002 | 0.044 | 0.746 | 1.665 | 0.000 |
| SMB | 144 | -0.047 | 0.061 | 0.001 | 0.001 | 0.001 | 0.023 | 0.226 | -0.303 | 0.394 |
| HML | 144 | -0.111 | 0.083 | -0.003 | -0.002 | 0.001 | 0.027 | 0.148 | 2.456 | 0.000 |
| MOM | 144 | -0.344 | 0.125 | 0.003 | 0.001 | 0.002 | 0.047 | -2.848 | 19.684 | 0.000 |
| UNEMPLOY | 144 | -0.075 | 0.080 | 0.000 | -0.001 | 0.001 | 0.028 | 0.524 | 0.401 | 0.028 |
| IP | 144 | -0.043 | 0.015 | 0.002 | 0.001 | 0.000 | 0.007 | -2.112 | 9.593 | 0.000 |
| SHORT_TERM (*) | 144 | -0.000 | 0.001 | 0.003 | 0.019 | 0.000 | 0.144 | 4.657 | 31.408 | 0.000 |
| LONG_TERM (**) | 144 | -0.000 | 0.000 | 0.003 | 0.001 | 0.000 | 0.095 | 0.187 | 1.360 | 0.005 |
| YIELD_SPREAD | 144 | -0.006 | 0.038 | 0.020 | 0.020 | 0.000 | 0.009 | 0.360 | -0.050 | 0.214 |
| DEFAULT_SPREAD | 144 | 0.000 | 0.022 | 0.011 | 0.012 | 0.000 | 0.005 | 0.182 | -0.852 | 0.072 |
| OIL | 144 | -0.335 | 0.290 | 0.007 | 0.003 | 0.008 | 0.089 | 0.372 | 1.599 | 0.000 |
| INFL | 144 | -0.018 | 0.010 | 0.002 | 0.002 | 0.000 | 0.003 | -2.137 | 12.798 | 0.000 |

* Median, Mean, Var and Std. dev values are multiplied with 1 000 to show more decimals

** Median, Mean, Var and Std. dev values are multiplied with 10 000 to show more decimals

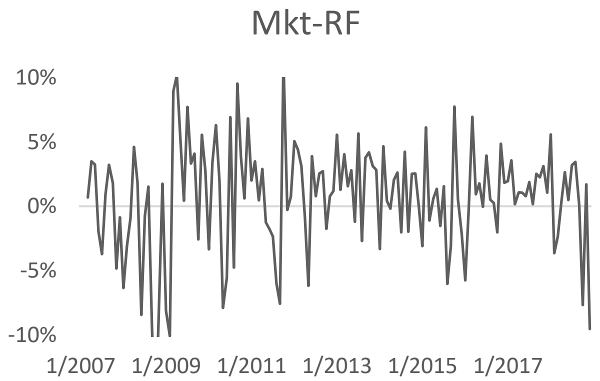
The normality of time-series is examined with skewness and extra kurtosis as well as with the Jarque-Bera normality test. Skewness for *MKT-RF* is 0.75 and kurtosis 1.67, so the distribution is a little bit skewed to the right and higher than normal distribution. In addition, the Jarque-Bera *p*-value is smaller than 0.1%, so the hypothesis of normality is rejected at a 0.1% significance level, which indicates that *MKT-RF* is not normally distributed. The Jarque-Bera *p*-values for *HML* and *MOM*, which has a lot of extra kurtosis (19,7) are both smaller than 0.1%, so their distributions also differ from the normal distribution. They are also skewer than the normal distribution: *HML* is right-skewed (0.15) and *MOM* left-skewed (-2.85).

On contrast, *SMB* has a Jarque-Bera *p*-value higher than 5%, so the null hypothesis of normality can be accepted. This means that the size premium follows the normal distribution, although its skewness is 0.23 and kurtosis -0.30, making the distribution a little bit flatter than normal distribution and slightly right-skewed.

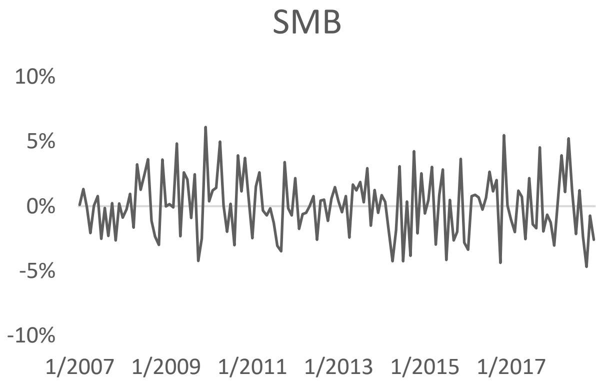
Among macro variables, all factors are right-skewed with skewness ranging from 0.18 (*DEFAULT_SPREAD*) to 4.66 (*SHORT_TERM*), except *INFL* and *IP*, which are left-skewed with respective skewnesses of -2.14 and -2.11. Almost all macro factors have extra kurtosis, ranging from 0.40 (*UNEMPLOY*) to 31.41 (*SHORT_TERM*). The kurtosis of the *YIELD_SPREAD* and *DEFAULT_SPREAD* are -0.05 and -0.85, so they are "flatter" than the normal distribution.

These two macro factors, *YIELD_SPREAD* and *DEFAULT_SPREAD*, are normally distributed based on the Jarque-Bera test, as their *p*-values exceed 5%, so the hypothesis of normality can be accepted at a 5% significance level. The Jarque-Bera *p*-value for *UNEMPLOY* is 2.8%, so the null hypothesis of normality can be accepted for it as well at a 1% significance level. The least factor that follows the normal distribution is *LONG_TERM*, which has a *p*-value of 0.5%, so the hypothesis of normality can be accepted at a 0.1% significance level. Bayesian methodology does not require normal distributions, so the non-normality of some factors does not reduce the reliability of the study (see, e.g., Berger et al. 2001, 139).

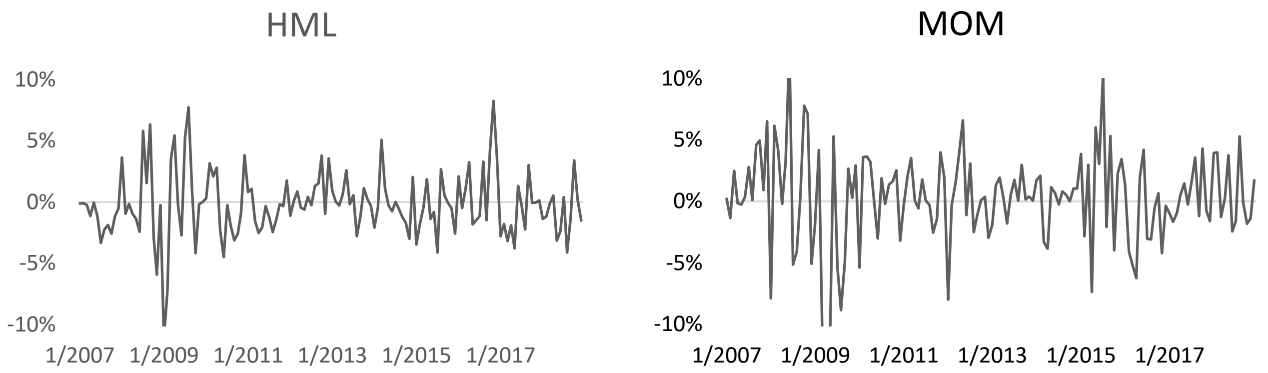
In Figure 1, the time-series variations of the fundamental factors are shown for the whole sample period. It can be seen from the Figure 1, that the realized premiums are quite volatile. The most volatile are the market premium (*MKT-RF*) and the momentum premium (*MOM*), which could also be seen from the standard deviations in Table 7. The premiums are the most volatile during the financial crisis in 2008−2009, especially the market, value (*HML*) and momentum premiums. The size premium (*SMB*) is more stable, even during the financial crisis, which is quite surprising, because one would expect that negative conditions in the economy would have a larger effect on small firms.




(a)



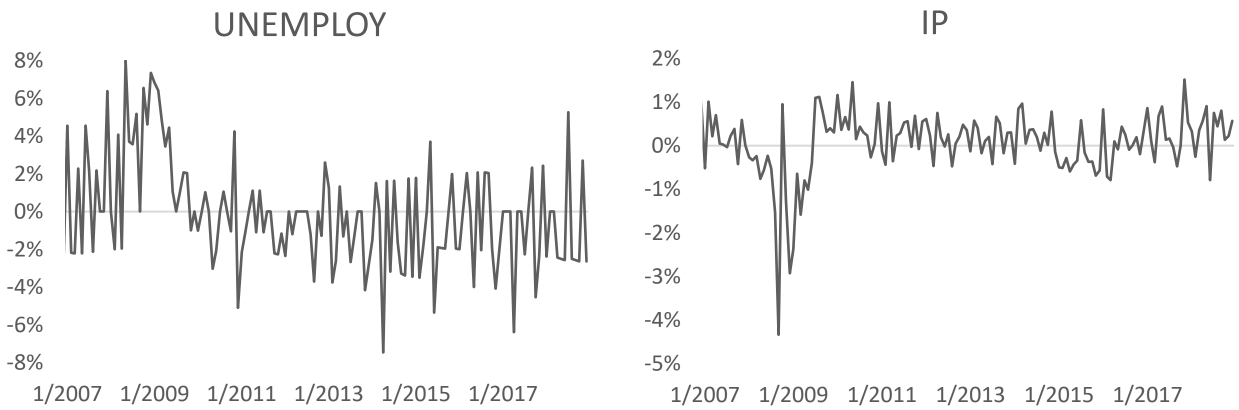

(b)




(c) (d)

Figure 1 Premiums for fundamental factors in 2007−2018

In Figure 2, the time-series variation of the macroeconomic factors are shown. As we can see from panels a to h, oil price change (*OIL*) has the highest, and also the most volatile, premiums throughout the time period. Unemployment change (*UNEMPLOY*) is also very volatile, and because it has a negative correlation with stock returns, the premium is very high during the financial crisis. The premiums of the industrial production factor (*IP*) and inflation change (*INFL*) drop during the financial crisis, but are otherwise quite small and stable. Short- and long-term interest rate changes (*SHORT_TERM* and *LONG_TERM*) are very stable and close to zero throughout the time period. This means that changes are small in interest rates between two subsequent months, in other words interest rates do not change fast, which is a quite intuitive result. The yield and default spread premiums (*YIELD_SPREAD* and *DEFAULT_SPREAD*) are the highest during the financial crisis but then start to lower.




(a) (b)

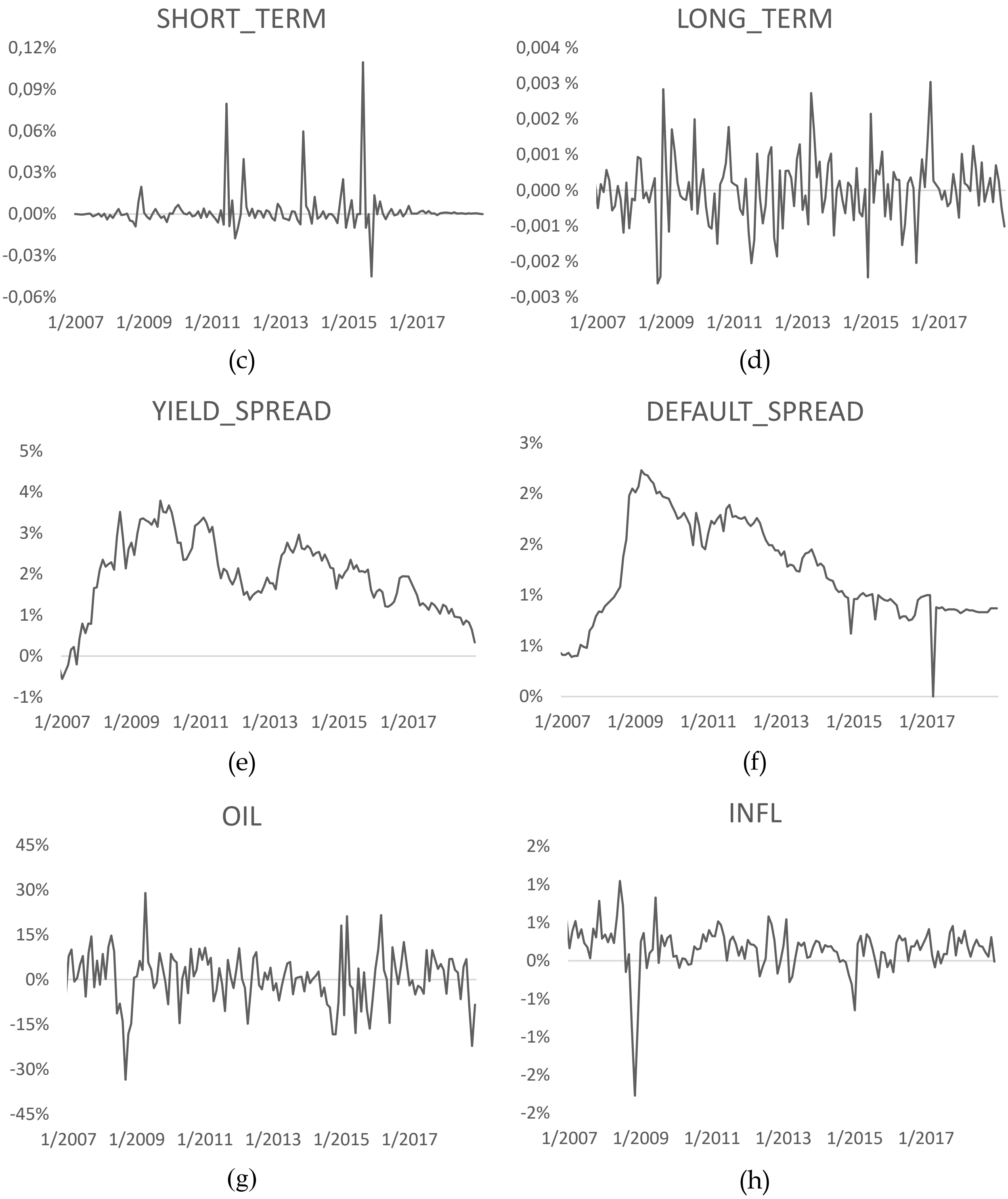



Figure 2 The premiums for the macroeconomic factors in 2007−2018

## 5.2 Results from Bayesian model averaging

The BMA algorithm is run separately for all twelve industries with the in-sample time period (2007−2016) to examine, which factors out of the twelve are dominating the US market and in order to construct the optimal multifactor model. The posterior probabilities for the best models in every industry as well as all the individual factors are represented in Table 8.

The numbers in the Table 8 are posterior probabilities for the individual factors in every industry, or in other words, the sum of those model posterior probabilities, where the factor is included, in the specific industry. For example, in the BUSEQ industry there are twelve competing models after the number is reduced with Occam's window. In BUSEQ industrial production change (*IP*) is included in two of the competing twelve models. The posterior probabilities for these two models are 5.2% and 2.6% (neither of these is the best model), and the sum of them is 7.8%, which is the overall factor-specific posterior probability for *IP* in the BUSEQ industry. With this method, factor-specific posterior probabilities are calculated for every industry and summed in Table 8. Only the final factor-specific posterior probabilities are shown in the Table 8, not the number of competing models in every industry, or their structure.

It can be easily seen, that six factors stand out from the others: market premium, size factor, value factor, momentum factor, long-term interest rate change, and oil price change. The most dominant factor is the market premium (*MTK-RF*), which has the posterior probability of one in all industries. This means that the probability of inclusion of the market factor approaches 100%. The second-best factors are the momentum factor (*MOM*), long-term interest rate change (*LONG_TERM*) and oil price change (*OIL*), which are included in the optimal model in five out of twelve industries with over 50% probability. In addition, *MOM* and *OIL* are included in the optimal model with 100% probability in three out of twelve industries. The Fama-French factors, *SMB* and *HML*, are included in the optimal model in four out of twelve industries.

The model posterior probability (the bottom line in the Table 8) is the probability of the best model being the true model, and the best model means the combination of factors that gives the highest model posterior probability in the certain industry. For example, for the industry BUSEQ, the best model includes three factors: *MKT-RF*, *HML* and *LONG_TERM* factor (marked with the star), and its posterior probability is 44.6%. The average of the model posterior probability of the best models is 23.1%, and it ranges from 6.6% to 50.9% between the different industries, which means that these are the probabilities that the model is the true model among the competing candidates.

Adjusted R-squared is high for all the best models and it varies from 52.3%−93.1% and the average is 80.9%. This indicates that the best models of all industries seem to be able to capture the characteristics of stock returns quite well.

Table 8 Posterior probabilities for the best models in every industry and for the individual factors

The numbers in the table are posterior probabilities for the individual factors in every industry. The value 1 means that the probability that the factor is included in the model approaches 100%. These factor-specific posterior probabilities are the sum of those model posterior probabilities, which include the factor. The star behind the value means that the specific factor is included in the best model in that industry. Under the line are the numbers related to the best model. No of variables tells how many factors are in the optimal model in the specific industry. Adjusted R-squared is the best model's adjusted R-squared. Finally, the posterior probability of the best model is shown in the bottom. In the right panel are averages of the all industries.

| | BUSEQ | CHEMS | DURBL | ENRGY | HLTH | MANUF | MONEY | NODURBL | OTHER | SHOPS | TELCM | UTILS | AVERAGE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MKT_RF | 1* | 1* | 1* | 1* | 1* | 1* | 1* | 1* | 1* | 1* | 1* | 1* | 1 |
| SMB | 0.042 | 0.685* | 1* | 0.033 | 0.026 | 1* | 0.158 | 1* | 0.130 | 0.330 | 0.040 | 0.058 | 0.375 |
| HML | 1* | 0.019 | 0.047 | 0.041 | 0.918* | 0.048 | 1* | 0.052 | 1* | 0.034 | 0.038 | 0.287 | 0.374 |
| MOM | 0.069 | 0.058 | 1* | 0.632* | 0.100 | 1* | 1* | 0.200 | 0.072 | 0.048 | 0.120 | 0.870* | 0.431 |
| DEFAULT_SPREAD | 0.042 | 0.021 | 0.050 | 0.159 | 0.027 | 0.135 | 0.036 | 0.081 | 0.039 | 0.124 | 0.453 | 0.038 | 0.100 |
| INFL | 0.055 | 0.584* | 0.054 | 0.533* | 0.073 | 0.067 | 0.040 | 0.096 | 0.040 | 0.103 | 0.061 | 0.044 | 0.146 |
| IP | 0.078 | 0.025 | 0.054 | 0.034 | 0.101 | 0.711* | 0.046 | 0.414 | 0.190 | 0.090 | 0.053 | 0.052 | 0.154 |
| LONG_TERM | 0.889* | 0.329 | 0.049 | 0.140 | 0.379 | 0.026 | 0.854* | 0.953* | 0.377 | 0.54* | 0.322 | 1* | 0.488 |
| OIL | 0.042 | 0.154 | 0.083 | 1* | 0.824* | 0.362 | 1* | 0.035 | 0.324 | 1* | 0.045 | 0.593* | 0.455 |
| SHORT_TERM | 0.051 | 0.666* | 0.054 | 0.133 | 0.044 | 0.263 | 0.165 | 0.035 | 0.040 | 0.365 | 0.093 | 0.039 | 0.162 |
| UNEMPLOY | 0.047 | 0.076 | 0.053 | 0.039 | 0.043 | 0.146 | 0.046 | 0.123 | 0.043 | 0.537* | 0.562* | 0.040 | 0.146 |
| YIELD_SPREAD | 0.043 | 0.041 | 0.047 | 0.060 | 0.027 | 0.320 | 0.036 | 0.073 | 0.042 | 0.049 | 0.078 | 0.038 | 0.071 |
| No of Variables | 4 | 5 | 4 | 5 | 4 | 5 | 6 | 4 | 3 | 4 | 4 | 5 | 4.417 |
| Adjusted R-squared | 0.897 | 0.856 | 0.828 | 0.736 | 0.686 | 0.929 | 0.924 | 0.740 | 0.931 | 0.844 | 0.814 | 0.523 | 0.809 |
| Posterior probability | 0.446 | 0.106 | 0.509 | 0.141 | 0.280 | 0.110 | 0.386 | 0.222 | 0.156 | 0.066 | 0.100 | 0.250 | 0.231 |

The number of possible models is reduced with Occam's window and its size is set at 20. Posterior probabilities cannot be shown here for all the competing models, because there are dozens of them (depending on the industry) even after adjusting Occam's window. The posterior probabilities for the three best models are shown in Table 9. As it can be seen from Table 9, the posterior probabilities start mainly to lower quite fast when moving from the best model to the second best and third best models. The posterior probability in the second-best models is under 10% in half of the industries. In addition, the average posterior probability drops from 23.1% to 10.6%, when moving from the best model to the second best. This indicates, that model averaging is not absolutely necessary for this study, as the relevance of other than the best models is not very significant (Steel et al. 2011, 33).

Table 9 Posterior probabilities for the second and third best models

The posterior probabilities are calculated for the individual industries with BMA add-in. The best model has the highest posterior model probability, whereas the second-best model has the second highest and the third model the third highest.

| | **The best** | **The second best** | **The third best** |
|---|---|---|---|
| | 0.446 | 0.085 | 0.069 |
| CHEMS | 0.106 | 0.089 | 0.052 |
| DURBL | 0.509 | 0.083 | 0.054 |
| ENRGY | 0.141 | 0.108 | 0.107 |
| HLTH | 0.280 | 0.128 | 0.075 |
| MANUF | 0.110 | 0.077 | 0.068 |
| MONEY | 0.386 | 0.128 | 0.117 |
| NODURBL | 0.222 | 0.129 | 0.060 |
| OTHER | 0.156 | 0.149 | 0.110 |
| SHOPS | 0.066 | 0.065 | 0.061 |
| TELCM | 0.100 | 0.084 | 0.080 |
| UTILS | 0.250 | 0.153 | 0.097 |
| Average | 0.231 | 0.106 | 0.079 |

Six factors out of twelve seem to have some significance in the US market. These factors are the market premium (*MKT-RF*), size factor (*SMB*), value factor (*HML*), momentum factor (*MOM*), long-term interest rate change (*LONG_TERM*), and oil price change (*OIL*). The average posterior probabilities for these factors range from 37.4% to 48.8% (and 100% for the market premium). These factors are all close to each other when measured by average posterior probability and they are also included in the best model in four or more industries out of twelve. So to answer the first research question, these six factors are included in the BMA multifactor model. The next best factor, short-term interest rate change (*SHORT_TERM*), is very far from these factors with a posterior probability of 16.2%. Because of these reasons, all other factors besides these six factors are left out of the BMA multifactor model.

The most dominant factor is the market premium, which has the highest posterior probability (approaches 100%) in all industries. This means that the test portfolios move quite in the line with the market portfolio, which is not very surprising as both the test portfolios and the market portfolio consist of NYSE, AMEX and NASDAQ stocks. The size premium has been questioned lately, for example Dijk (2011) states that the premium has disappeared from the market since the 80s. However, the results indicate that it can still have some explanatory power for stock market returns. The size premium has been found earlier in US markets by for example Ericsson and Karlsson (2003) in the time period of 1963−2002 and Marquering et al. (2006) in the time period of 1960−2003.

The value premium seems to also have some explanatory power in-sample. The premium has been found previously from US markets by many researchers, for example Black and Fraser (2003) and Ericsson and Karlsson (2003), and more lately, Fama and French (2012).

Industrial production, unemployment rate and inflation seem not be priced in the time period. Jareño and Negrut (2016) examine all of these three factors with the freshest data, 2008−2014 and find that industrial production and the unemployment rate are priced in the US markets, whereas inflation is not. The results here conflict a little with this finding, because the posterior probabilities are very low with all of these factors. The differences could be due to differences in the data and/or methodology. For example, Jareño and Negrut use quarterly, not monthly, data and the factors to predict only the S&P 500 and Dow Jones returns as a whole. They do not use the test portfolios, which can be problematic in terms of reliability of the results.

The momentum premium has a higher posterior probability than the size and value premiums and it is included in the multifactor model in five out of twelve industries. The factor has been found previously for example by Griffin et al. (2003) and Lim et al. (2018) in the US markets, so the results here are consistent with these findings. It seems that momentum has some explanatory power, which conflicts with Hwang and Rubesam's (2015) and Bhattacharya's et al. (2017) findings that the momentum has not been significant after the late 90s.

The predictability of long-term interest rates, as well as other macroeconomic effects, for stock returns has not been a great interest of researchers lately. The results suggest, however, that the long-term interest rate has the second highest posterior probability of inclusion after the market premium, and therefore it can explain stock returns in-sample. This result is consistent with Jareño and Negrut (2016), who find it to be significantly priced in 2008–2014 in US markets.

Oil price change is another macro variable, which is included in the model. It has been previously found for example by Cheng (1996) and more recently by Narayan and Gupta (2015) in the long time period of 1859–2013. Oil price change has the third highest posterior probability after the market premium and long-term interest rate change, which indicates that macroeconomic factors are also relevant to consider when constructing multifactor models, and in this sample period actually have a higher probability to be priced than most fundamental factors.

## 5.3 Performance of BMA multifactor model by out-of-sample forecasting ability

After constructing the BMA multifactor model, the second research question of the out-of-sample forecasting ability can be examined. Forecasting accuracy is measured by MSE (Mean Standard Error) values. To calculate the MSE values, the expected portfolio returns for all multifactor models are needed as noted before. The expected portfolio returns are calculated using the CAPM, the Fama-French 3-factor model and Fama-French 3-factor plus momentum model as well as for the BMA multifactor model constructed in the previous section. To get the expected portfolio returns, the beta coefficients and factor realizations are required.

Beta coefficients are calculated for each factor in each model separately for all the portfolios. The betas used are rolling 10-year betas and they are determined with the linear time-series regressions. The out-of-sample estimated time-series averages of beta coefficients are shown in Table 10. Because they are average beta coefficients through the whole out-of-sample time period, they are not used for calculating expected portfolio returns but showed here to give a better understanding about the process of calculating MSE.

The beta for the realized market premium is positive in every factor model and quite close to one. This means that the returns for the industry portfolios move quite in line with market returns. Betas for other factors depend a lot on the factor model and industry. The betas for long-term interest rate change are very high, because the beta coefficient is the portfolio returns volatility divided by factor volatility, and the factor volatilities are extremely low for the long-term interest rate factor. Dividing by these very low values leads to large numbers.

Table 10 Beta coefficients for the all factors in multifactor models in every industry

Average beta coefficients for all multifactor models through the out-of-sample time period of 2017−2018. The betas are calculates as the rolling 10-year betas with the time series regressions.

| | BUSEQ | CHEMS | DURBL | ENRGY | HLTH | MANUF | MONEY | NODURBL | OTHER | SHOPS | TLCM | UTILS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **PANEL 1 CAPM** | | | | | | | | | | | | |
| MKT−RF | 1.072 | 0.919 | 1.615 | 0.991 | 0.732 | 1.295 | 1.285 | 0.642 | 1.160 | 0.830 | 0.921 | 0.493 |
| **PANEL 2 Fama-French 3-factor model** | | | | | | | | | | | | |
| MKT−RF | 1.153 | 0.944 | 1.437 | 0.962 | 0.785 | 1.232 | 1.137 | 0.721 | 1.095 | 0.828 | 0.942 | 0.580 |
| SMB | -0.006 | -0.161 | 0.663 | 0.047 | 0.049 | 0.265 | -0.024 | -0.319 | 0.065 | 0.103 | -0.071 | -0.207 |
| HML | -0.377 | 0.018 | 0.272 | 0.089 | -0.292 | 0.072 | 0.723 | -0.098 | 0.248 | -0.076 | -0.037 | -0.239 |
| **PANEL 3 Fama-French 3-factor & momentum model** | | | | | | | | | | | | |
| MKT−RF | 1.139 | 0.927 | 1.332 | 0.975 | 0.804 | 1.199 | 1.111 | 0.731 | 1.089 | 0.838 | 0.953 | 0.619 |
| SMB | -0.005 | -0.161 | 0.670 | 0.045 | 0.048 | 0.267 | -0.022 | -0.320 | 0.065 | 0.103 | -0.072 | -0.210 |
| HML | -0.407 | -0.021 | 0.038 | 0.118 | -0.250 | 0.001 | 0.665 | -0.076 | 0.234 | -0.054 | -0.013 | -0.154 |
| MOM | -0.047 | -0.063 | -0.371 | 0.043 | 0.067 | -0.114 | -0.092 | 0.034 | -0.023 | 0.037 | 0.037 | 0.135 |
| **PANEL 4 Bayesian multifactor model** | | | | | | | | | | | | |
| MKT−RF | 1.124 | 0.911 | 1.388 | 0.686 | 0.869 | 1.183 | 1.156 | 0.743 | 1.129 | 0.920 | 0.966 | 0.616 |
| SMB | -0.039 | -0.124 | 0.636 | 0.095 | 0.068 | 0.283 | -0.097 | -0.275 | 0.080 | 0.111 | -0.040 | -0.091 |
| HML | -0.419 | -0.006 | 0.026 | 0.135 | -0.242 | 0.007 | 0.636 | -0.059 | 0.240 | -0.050 | -0.000 | -0.108 |
| MOM | -0.040 | -0.064 | -0.381 | 0.100 | 0.051 | -0.113 | -0.094 | 0.028 | -0.033 | 0.019 | 0.031 | 0.124 |
| OIL | -0.002 | 0.033 | -0.072 | 0.306 | -0.053 | 0.022 | -0.080 | 0.010 | -0.032 | -0.075 | 0.003 | 0.060 |
| LONG_TERM | 404.547 | -358.137 | 231.162 | 208.578 | -378.548 | -124.193 | 688.422 | -516.409 | -265.730 | -303.611 | -395.301 | -1277.451 |

Next, the expected factor realizations (factor "returns") are computed for all of the factors used in the competing models. The expected values are calculated for all the months in the out-of-sample time period, so that they can be used in MSE calculation. Two different estimates for factor realizations are used: the first is a rolling average of the previous 10 year's factor realizations and the second is simply the factor realization of the previous month (rolling 1 month). In Table 11 the results from the first method are shown in Panel 1 and the second in Panel 2. In the first method, the average expected factor realizations are positive for the market premium (0.8%), size premium (0.1%), long-term interest rate change factor (0.00001%), and oil price change factor (0.2%). The expected factor realizations are negative for the value premium (-0.1%) and momentum premium (-0.09%) on average.

The second method gives slightly higher expected factor realizations on average for the market premium (1.1%), momentum premium (0.5%), long-term interest rate change factor (0.00011%), and oil price change factor (0.9%). The size (-0.2%) and value premiums (-0.7%) factor realizations are slightly lower than in the first method on average.

Table 11 Expected factor realizations for 2017−2018 and average values

The expected factor realizations $f_{1t} \dots f_{xt}$ for all factors $x$ ($x = 6$) are calculated for all months $t$ in the out-of-sample time period. The calculations are done in two different ways, which are represented in different panels. In Panel 1, the values are calculated as the rolling 10-year averages. In Panel 2, the values are simply the factor realizations of the previous month $t$-1.

| **Panel 1: Factor realizations (10-year average)** | | | | | | |
|---|---|---|---|---|---|---|
| | MKT-RF | SMB | HML | MOM | LONG_TERM | OIL |
| 31.1.2017 | 0.0064 | 0.0013 | -0.0008 | -0.0000 | -0.0000 | 0.0037 |
| 28.2.2017 | 0.0064 | 0.0012 | -0.0010 | -0.0001 | -0.0000 | 0.0040 |
| 31.3.2017 | 0.0069 | 0.0009 | -0.0012 | -0.0001 | -0.0000 | 0.0034 |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |
| 31.10.2018 | 0.0104 | 0.0014 | -0.0026 | -0.0012 | 0.0000 | 0.0025 |
| 30.11.2018 | 0.0112 | 0.0012 | -0.0020 | -0.0020 | 0.0000 | 0.0045 |
| 31.12.2018 | 0.0120 | 0.0014 | -0.0015 | -0.0027 | 0.0000 | 0.0042 |
| Average | 0.0081 | 0.0014 | -0.0012 | -0.0009 | 0.0000 | 0.0017 |

| **Panel 2: Factor realizations (previous month)** | | | | | | |
|---|---|---|---|---|---|---|
| | MKT-RF | SMB | HML | MOM | LONG_TERM | OIL |
| 31.1.2017 | 0.0182 | 0.0007 | 0.0360 | -0.0036 | 0.0000 | 0.1258 |
| 28.2.2017 | 0.0194 | -0.0102 | -0.0278 | -0.0097 | 0.0000 | -0.0197 |
| 31.3.2017 | 0.0357 | -0.0200 | -0.0180 | -0.0166 | -0.0000 | -0.0020 |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |
| 31.10.2018 | 0.0006 | -0.0237 | -0.0134 | -0.0009 | 0.0000 | 0.0685 |
| 30.11.2018 | -0.0768 | -0.0468 | 0.0341 | -0.0182 | 0.0000 | -0.0876 |
| 31.12.2018 | 0.0169 | -0.0074 | 0.0019 | -0.0140 | -0.0000 | -0.2221 |
| Average | 0.0105 | -0.0016 | -0.0071 | 0.0052 | 0.0000 | 0.0092 |

Expected portfolio returns can be calculated as the expected factor betas times the expected factor realizations and by summing these with the expected risk-free return. The expected portfolio returns are calculated for every month in the out-of-sample time period from 2017 to 2018, so 24 expected portfolio returns are calculated for every competing model: CAPM, FF3, FF4 and the BMA multifactor model. In addition, expected portfolio returns are calculated in two ways: first by using the rolling average of the previous ten year's factor realizations and then by using only the previous month's factor realizations as an estimate of the period's factor realization.

After that, Mean Standard Error (MSE) values can be calculated to answer the second research question on the performance of the BMA multifactor model compared to the benchmark models. The MSE values are calculated based on two different expected factor realization estimates and they are simply the averages of the squared difference between the actual and expected portfolio returns. The expected portfolio returns and MSE values based on the rolling averages of 10 year's factor realizations are shown in Table 12. The values based on using the previous month's factor realization as the estimate for the factor realization of the period, are examined in Table 13.

From Table 12, it can be seen that the average expected returns range from 0.97% to 1.08% in the business equipment industry (BUSEQ), from 1.11% to 1.15% in the manufacturing industry (MANUF), from 1.30% to 1.41% in the consumer durables industry (DURBL) and from 0.97% to 1.04% in the other industry (OTHER), which are the highest industry-specific expected returns among the twelve industries considered. The lowest expected portfolio returns on average are in the utilities industry (UTILS), from 0.50% to 0.58%, depending on the factor model.

The MSE values are represented in Panel 3. The values are multiplied by 1000, so that more decimals can be seen. The MSE values are quite small for each model in every industry and they vary mostly between 0.001 and 0.002. The highest MSE values are in the manufacturing (MANUF), around 0.0021, and oil, gas and coal extraction and products industry (ENRGY), around 0.0037. The lowest values are in the utilities (UTILS) industry, around 0.0007, so the models predict stock returns best in that industry.

The differences between the competing models are relatively small. When looking at the average MSE values in all twelve industries, the BMA multifactor model has the lowest MSE value, 0.001707. The second lowest values are those of the FF3 and FF4 factor models, both 0.001709. The FF4 model has a slightly lower value, when more decimals are used. CAPM has the highest value, 0.001719.

The results show, that the multifactor model constructed with the BMA method can predict stock market returns slightly better than the traditional Fama-French models in the time period of 2017–2018. The BMA multifactor model has the lowest MSE score in four out of twelve industries and the second lowest in three out of twelve industries. The second research question can be answered: the BMA multifactor model can predict stock returns better than the benchmark models out-of-sample. This means that it can be confirmed that the additional factors, long-term interest rate change and the oil price change, have at least some explanatory power in the US market. As it was seen in the previous section, these additional macro factors actually had higher posterior probabilities of inclusion than the Fama-French factors, which indicates that they can be even more relevant compared to fundamental factors in predicting stock returns. The Fama-French models can predict stock returns better than CAPM, so value, size and momentum factors still seem to have at least some explanatory power in the US market. However, the differences between the BMA and benchmark models are very small, so it cannot be said that the BMA model can predict stock returns significantly better than the benchmarks.

Because the constructed BMA multifactor model performed better out-of-sample, it can be confirmed that the Bayesian model averaging approach is a useful tool for factor selection in stock markets. Researchers would find it beneficial to consider the Bayesian approach alongside the traditional mean-variance methods in multifactor research, as it provides many benefits compared to those methods (see e.g. Hoeting et al. 1999; Cremers 2002; Hooker 2004; Bianchi et al. 2017). The results are consistent with Ando (2009), who examined factor selection in US with the BMA approach in 1990−2004, and Tsai et al. (2010), who used a longer time period, 1954−2005. From the time period that is used in this study, 2007−2018, other researches of the topic are not available from the US (or elsewhere). For this reason, the results are compared to those from the early 2000s.

In Table 13 the results of using the simpler previous month's factor realization as the estimate for the factor realization of the period are shown. Again, the expected portfolio returns are calculated for every month in the out-of-sample time period in every industry and for all factor models. The averages of these values are represented in the table. The expected returns are higher than previously. For example, in the business equipment industry (BUSEQ) the expected return ranges between 1.24% and 1.59% on average, when previously it was 0.97−1.08%. The same effect is seen in every industry. The manufacturing (MANUF) and consumer durables (DURBL) industries have the highest expected returns on average, which vary between 1.28−1.46% (MANUF) and 1.19−1.82% (DURBL).

The MSE values of these factor realizations are shown in Panel 3, and to show more decimals, they are multiplied again by 1000. The MSE values are slightly higher in this calculation than previously and vary approximately between 0.001 and 0.005, depending on the model and industry. The utilities industry (UTILS) once again has the lowest MSE values, from 0.0011 to 0.0015. The oil, gas and coal extraction and products industry (ENRGY) has the highest MSE values again, ranging this time from 0.0044 to 0.0051. Manufacturing (MANUF) also has one of the highest MSE values once more, 0.0030−0.0033, but this time the second highest value is in consumer durables (DURBL), 0.0035−0.0043.

Because MSE values are now slightly higher in every portfolio, it seems that the second way to determine factor realizations is not as good as the previous method, where the rolling 10-year average is used. The result was also assumed before the calculations and it feels quite intuitive: factor realizations (or factor returns) can move very fast in the market, much faster than risk-free returns, where the previous month's level is also used. For that reason, it is better to use a longer time average to predict future factor realizations.

Compared to the previous calculation, the average MSE values through all the industries seem similar: the BMA multifactor model achieves the lowest MSE score again, 0.0025860, and the FF3 and FF4 slightly higher values, 0.0026132 and 0.0025864. CAPM once again has the highest MSE value, 0.0026352. The BMA multifactor model has the lowest MSE score in four out of twelve industries and the second lowest in five out of twelve industries. The results are similar between the two different calculation methods, so the conclusions made earlier about the predictability of factors are consistent between the two calculation methods.

Table 12 The expected returns and MSE values with 10-year rolling factor realizations for all factor models

The expected returns (Panel 1) are calculated as the sum of expected risk-free return and the expected factor betas times the expected factor realizations. Here, only the average over the out-of sample period is shown. The expected risk-free return is the risk-free return in the previous month $t$-1 (not shown in the table). The factor realizations are calculated here as the rolling average of factor realizations in the previous 10 years (not shown in the table). The beta coefficients are calculated as the rolling 10-year betas with the time-series regression (not shown in the table). Actual returns for all industry portfolios (Panel 2) are collected from Kenneth French's data library. MSE values (Panel 3) are calculated as the difference between the actual industry portfolio returns and the expected portfolio returns in every month in the out-of-sample time period. Here, only the averages are shown and they are multiplied by 1000 to show more decimals.

| Panel 1: Expected portfolio returns | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **BUSEQ** | **TELCM** | **UTILS** | **SHOPS** | **HLTH** | **MONEY** | **OTHER** | **NODURBL** | **DURBL** | **MANUF** | **ENRGY** | **CHEMS** | **Average** |
| CAPM | 0.0097 | 0.0085 | 0.0050 | 0.0078 | 0.0070 | 0.0115 | 0.0104 | 0.0062 | 0.0141 | 0.0115 | 0.0091 | 0.0085 | 0.0091 |
| FF3 | 0.0108 | 0.0085 | 0.0056 | 0.0080 | 0.0079 | 0.0093 | 0.0097 | 0.0065 | 0.0132 | 0.0113 | 0.0087 | 0.0084 | 0.0090 |
| FF4 | 0.0108 | 0.0086 | 0.0057 | 0.0080 | 0.0079 | 0.0092 | 0.0097 | 0.0066 | 0.0130 | 0.0112 | 0.0088 | 0.0084 | 0.0090 |
| BMA | 0.0107 | 0.0087 | 0.0058 | 0.0086 | 0.0083 | 0.0095 | 0.0100 | 0.0067 | 0.0134 | 0.0111 | 0.0070 | 0.0083 | 0.0090 |

| Panel 2: Actual returns for the industry portfolios | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **BUSEQ** | **TELCM** | **UTILS** | **SHOPS** | **HLTH** | **MONEY** | **OTHER** | **NODURBL** | **DURBL** | **MANUF** | **ENRGY** | **CHEMS** | **Average** |
| Actual return | 0.0140 | 0.0001 | 0.0051 | 0.0114 | 0.0104 | 0.0056 | 0.0035 | -0.0004 | -0.0035 | 0.0070 | -0.0062 | 0.0059 | 0.0044 |

| Panel 3: MSE values*1000 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **BUSEQ** | **TELCM** | **UTILS** | **SHOPS** | **HLTH** | **MONEY** | **OTHER** | **NODURBL** | **DURBL** | **MANUF** | **ENRGY** | **CHEMS** | **Average** |
| CAPM | 1.8997 | 1.5357 | 0.7213 | 1.9762 | 1.7078 | 1.6621 | 1.3275 | 1.2018 | 1.7511 | 2.0996 | 3.7359 | 1.0076 | 1.7189 |
| FF3 | 1.9111 | 1.5354 | 0.7209 | 1.9844 | 1.7164 | 1.6230 | 1.3071 | 1.2152 | 1.7066 | 2.0816 | 3.6940 | 1.0072 | 1.7086 |
| FF4 | 1.9120 | 1.5358 | 0.7208 | 1.9840 | 1.7162 | 1.6236 | 1.3073 | 1.2152 | 1.7024 | 2.0825 | 3.6957 | 1.0071 | 1.7085 |
| BMA | 1.9153 | 1.5360 | 0.7196 | 1.9776 | 1.7130 | 1.6293 | 1.3105 | 1.2136 | 1.7216 | 2.0834 | 3.6601 | 1.0049 | 1.7071 |

Table 13 The expected returns and MSE values with previous month´s factor realizations for all factor models

The expected returns (Panel 1) are calculated as the sum of expected risk-free return and the expected factor betas times the expected factor realizations. Here, only the averages over the out-of sample period is shown. The expected risk-free return is the risk-free return in the previous month $t$-1 (not shown in the table). The factor realizations are here simply the previous month's factor realizations (not shown in the table). The beta coefficients are calculated as the rolling 10-year betas with the time-series regression (not shown in the table). Actual returns for all industry portfolios (Panel 2) are collected from Kenneth French's data library. MSE values (Panel 3) are calculated as the difference between the actual industry portfolio returns and the expected portfolio returns in every month in the out-of-sample time period. Here, only the averages are shown and they are multiplied by 1000 to show more decimals.

**Panel 1: Expected portfolio returns**

| | BUSEQ | TELCM | UTILS | SHOPS | HLTH | MONEY | OTHER | NODURBL | DURBL | MANUF | ENRGY | CHEMS | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CAPM | 0.0124 | 0.0109 | 0.0066 | 0.0096 | 0.0085 | 0.0144 | 0.0131 | 0.0078 | 0.0182 | 0.0146 | 0.0112 | 0.0106 | 0.0115 |
| FF3 | 0.0158 | 0.0117 | 0.0099 | 0.0096 | 0.0109 | 0.0077 | 0.0107 | 0.0097 | 0.0134 | 0.0133 | 0.0107 | 0.0112 | 0.0112 |
| FF4 | 0.0156 | 0.0118 | 0.0105 | 0.0097 | 0.0112 | 0.0073 | 0.0106 | 0.0099 | 0.0119 | 0.0128 | 0.0111 | 0.0110 | 0.0111 |
| BMA | 0.0159 | 0.0113 | 0.0094 | 0.0095 | 0.0110 | 0.0079 | 0.0104 | 0.0092 | 0.0125 | 0.0130 | 0.0109 | 0.0106 | 0.0110 |

**Panel 2: Actual returns for the industry portfolios**

| | BUSEQ | TELCM | UTILS | SHOPS | HLTH | MONEY | OTHER | NODURBL | DURBL | MANUF | ENRGY | CHEMS | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Actual return | 0.0140 | 0.0001 | 0.0051 | 0.0114 | 0.0104 | 0.0056 | 0.0035 | -0.0004 | -0.0035 | 0.0070 | -0.0062 | 0.0059 | 0.0044 |

**Panel 3: MSE values*1000**

| | BUSEQ | TELCM | UTILS | SHOPS | HLTH | MONEY | OTHER | NODURBL | DURBL | MANUF | ENRGY | CHEMS | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CAPM | 1.9792 | 2.1706 | 1.1076 | 2.4169 | 2.3153 | 3.0349 | 2.5135 | 1.4549 | 4.3461 | 3.2875 | 5.0725 | 1.9239 | 2.6352 |
| FF3 | 2.4221 | 2.1808 | 1.4201 | 2.5107 | 2.5735 | 2.6318 | 2.2576 | 1.6797 | 3.8299 | 3.0925 | 4.8210 | 1.9389 | 2.6132 |
| FF4 | 2.3928 | 2.1956 | 1.4696 | 2.5456 | 2.6280 | 2.5705 | 2.2456 | 1.6944 | 3.5228 | 3.0286 | 4.8472 | 1.8962 | 2.5864 |
| BMA | 2.3719 | 2.1507 | 1.4193 | 2.6827 | 2.7267 | 2.7089 | 2.3173 | 1.6244 | 3.7501 | 3.0371 | 4.3664 | 1.8767 | 2.5860 |

## 5.4 Robustness analysis

In this section the reliability of this study is examined with a robustness check. In this paper, twelve industry portfolios were used as test portfolios. To make sure that the results are reliable, they have to be robust throughout the different sets of portfolios (Cremers 2006, 2993). For this reason, MSE calculations are done for the different sets of portfolios, which are ten portfolios formed based on the firms' size.

The data for the second set of test portfolios is also collected from Kenneth French's data library. All the size portfolios are value-weighted and consist of 10% of stocks from NYSE, AMEX and NASDAQ, so that the first portfolio has the smallest 10% and the $10^{th}$ portfolio the biggest 10%. The portfolios are rebalanced at the end of June and all the stocks that have market equity data in June are included.

The time period of the out-of-sample period is the same, 2017−2018. MSE values are calculated in the same way as before, as the squared difference between the test portfolio returns and expected portfolio returns and taking the average of that. The only difference is then the test portfolio returns, which were previously industry portfolio returns and now size portfolio returns. In addition, both methods for calculating the expected factor realizations are used this time, and the results are represented for both methods separately.

The results are shown in Table 14. From the table it can be seen that the BMA multifactor model has the lowest average MSE value no matter which factor realizations are used. When rolling 10-year factor realization is used (Panel 1), the average MSE value is 0.0018155 for the BMA model, 0.0018159 for the FF4 model, 0.0018157 for the FF3 model and 0.0018398 for CAPM. CAPM performed the worst in out-of-sample again and the FF3 model a little bit better than the FF4 model. The MSE value for the BMA model is the lowest in four out of ten size portfolios. The values are multiplied with 1000 in the table so that more decimals can be shown.

In Panel 2, the MSE values are calculated based on the estimates of factor realizations that are simply the previous month's factor realizations (one month rolling factor realization). The results stay the same: the BMA model has the lowest average MSE value, 0.0029073 and it is the lowest in five out of ten size portfolios. This time the FF4 model has a lower MSE value, 0.0029130, than the FF3 model, 0.0029215, on average. CAPM has the highest MSE value, 0.0029537, once again.

The results are robust between two different sets of test portfolios in the out-of-sample time period of 2017−2018. This indicates that the results are consistent no matter what portfolios are used and the results can be universalized to a degree.

There are, however, some methodological issues in the research that can affect the reliability of the results. First, the time period of the study was fixed, 2007−2018, so it is not guaranteed that the optimal factors would stay the same outside of the test period. Additionally, only the best models from the Bayesian model averaging were taken into account when forming the optimal multifactor model. This was done, because the differences between the best and second best models were relatively large, however, if model uncertainty is fully taken into account, all models should be used to optimize the overall multifactor model.

Another problem related to the reliability of results is related to the calculation of the factor realizations. It was assumed that the factor realizations, or premiums, can be predicted from historical factor realizations, which is against the EMH. This problem was attempted to be negated by using the previous ten year's rolling average for predicting the next period's factor realization. Another way would be to calculate the factor realizations with Fama-MacBeth regressions.

Finally, the MSE values of competing models were compared by using the industry averages. The BMA multifactor model outperformed in four out of twelve industries and it was the second best in three to five out of twelve industries. Therefore, there is some variation between industries on how well the BMA multifactor model performs compared to the benchmark models. This problem was reduced with the robustness check using different sets of test portfolios. There was, however, some variation between these test portfolios as well.

Table 14 MSE values for ten size portfolios

MSE values calculated for 10 size portfolios from Kenneth French's data library. The smallest 10 stocks are in the Lo10 portfolio and biggest in the Hi10 portfolio. All other portfolios are middle quantiles. MSE values are multiplied with 1000. In Panel 1, the 10 year rolling factor realizations are used for the factor realization estimates for the next period. In the Panel 2, the estimate for the next period's factor realization is simply the previous months factor realization. Betas are calculated as the 10-year rolling betas in both methods. The expected risk-free return is the previous period's risk-free return.

| Panel 1: MSE values for 10 size portfolios with 10 year rolling factor realizations | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lo10 | Qnt2 | Qnt3 | Qnt4 | Qnt5 | Qnt6 | Qnt7 | Qnt8 | Qnt9 | Hi10 | **Average** |
| CAPM | 2.3525 | 2.5012 | 2.1687 | 2.2105 | 2.0372 | 1.6565 | 1.7318 | 1.3988 | 1.1931 | 1.1472 | **1.8398** |
| FF3 | 2.2986 | 2.4412 | 2.1339 | 2.1807 | 2.0039 | 1.6430 | 1.7209 | 1.3920 | 1.1898 | 1.1529 | **1.8157** |
| FF4 | 2.2995 | 2.4411 | 2.1339 | 2.1807 | 2.0043 | 1.6430 | 1.7213 | 1.3924 | 1.1904 | 1.1527 | **1.8159** |
| BMA | 2.2944 | 2.4409 | 2.1357 | 2.1823 | 2.0050 | 1.6414 | 1.7205 | 1.3929 | 1.1899 | 1.1525 | **1.8155** |
| **Panel 2: MSE values for 10 size portfolios with 1 month rolling factor realizations** | | | | | | | | | | | |
| | Lo10 | Qnt2 | Qnt3 | Qnt4 | Qnt5 | Qnt6 | Qnt7 | Qnt8 | Qnt9 | Hi10 | **Average** |
| CAPM | 3.3305 | 3.8205 | 3.6454 | 3.5996 | 3.4818 | 2.7411 | 2.9115 | 2.2982 | 2.0566 | 1.6513 | **2.9537** |
| FF3 | 3.3241 | 3.9209 | 3.6466 | 3.5193 | 3.2313 | 2.7135 | 2.8113 | 2.2459 | 2.0106 | 1.7909 | **2.9215** |
| FF4 | 3.3121 | 3.9196 | 3.6428 | 3.5181 | 3.2199 | 2.7153 | 2.7995 | 2.2355 | 1.9628 | 1.8046 | **2.9130** |
| BMA | 3.2679 | 3.8819 | 3.6441 | 3.5405 | 3.2323 | 2.7022 | 2.7970 | 2.2352 | 1.9606 | 1.8113 | **2.9073** |

# 6 CONCLUSIONS

In this thesis the problem of selecting the factors for a multifactor model was examined with the Bayesian model averaging approach. The constructed multifactor model was then evaluated and compared to the benchmark models. Factor selection was done in the in-sample time period of 2007−2016. The results showed that six factors out of twelve seemed to have some predictive power in the US markets based on their posterior probabilities.

Answering the first research question, the optimal set of factors consisted of six factors: the market premium, size, value, momentum, changes in long-term interest rates and oil price change. The last two, long-term interest rate and oil price change, received the highest factor-specific posterior probabilities of inclusion after the market premium, which indicates that those were the most relevant factors in explaining stock returns in-sample. This is an interesting result, because multifactor research has lately been focused more on fundamental factors and macroeconomic factors have received only a little attention.

The connection between long-term interest rates and stock returns is not a surprise, because typically when the stock market drop, investors allocate more wealth into short- and long-term bonds. They are considered as “safe havens”, which means that they are expected to retain, or increase, their value during times of market turbulence. Despite this connection, the long-term interest rate (measured by 10-year T-bond yield) is not examined much in the previous multifactor literature. Changes in oil price are more commonly used to predict stock returns in the existing literature but results have been mixed. However, the results here show that macroeconomic variables can be priced in US along the fundamental factors. For this reason, the research should not forget the macro factors when examining the multifactor models.

The market, size and value premium have been of great interest to researchers during the last few decades. In this thesis all three of these variables had relatively high posterior probabilities on average. The results here confirm that the market premium is the most dominant factor and the value premium still has some predictive power in the US. Also, regardless that the size premium has been strongly questioned lately, and many researchers have stated that it has disappeared from the markets since the 80s, the result here is that the size premium still exists in the US market, at least on some level. Original Fama-French factors therefore stay strong in predicting the stock returns and form the good base for constructing more complex multifactor models.

The momentum premium also had a relatively high posterior probability, so it seems to have some explanatory power in the US. Lim et al. (2018) used data up to 2017, and got the same results as represented here. The findings are a little bit mixed with all the factors considered here, but momentum is perhaps not as questioned as Fama-French factors. Perhaps it is because momentum is difficult to explain rationally, and relates rather to the irrational behavior of investors, which keep it priced in the markets.

The changes in short interest rates, the yield spread, default spread, industrial production, inflation rate and unemployment rate do not seem to have been priced in the US markets in 2007−2016. The findings are mostly in line with the previous literature. However, only eight macro factors were considered in this thesis, and there are many other factors in addition to them examined in the literature.

To answer the second research question, the MSE (Mean Squared Error) values were calculated for all of the competing multifactor models. The question was if the constructed BMA multifactor model could predict stock returns better than the benchmark models in the out-of-sample time period of 2017−2018. The results show that the BMA model has the lowest MSE values among the competing models. The Fama-French models also had higher MSE values than CAPM. Based on these results it can be said that the BMA multifactor model can predict stock returns out-of-sample better than traditional models. The results were robust throughout the set of portfolios.

The differences in MSE values are quite small between the competing models however. For that reason, the question needs more examination and there are many extensions that can be applied to this study. For example, in this study the same BMA multifactor model was used for every industry. To achieve better results, the multifactor model can be constructed separately for all industries (or other sets of test portfolios), so industry specific weightings between the factors can be taken into account more carefully. Secondly, more fundamental and macroeconomic factors could be considered as potential factors to find out, if more predictability power can be received. These factors could be the dividend yield, as an example of an additional fundamental factor, and GDP or credit spread, as examples of additional macroeconomic factors. Lastly, if model uncertainty is wanted to fully be taken into account by not only choosing the best model, Bayesian model averaging can be used as the method to weigh all competing models with their posterior probabilities and then use them all in an overall multifactor model.

## REFERENCES

Abhyankar, A. − Ho, K. Y. − Zhao, H. (2009) International value versus growth: evidence from stochastic dominance analysis. *International Journal of Finance & Economics*, Vol. 14 (3), 222−232.

Al-Rjoub, S. A. − Varela, O. − Kabir Hassan, M. (2005) The size effect reversal in the USA. *Applied Financial Economics*, Vol. 15 (17), 1189−1197.

Ali, A. − Hwang, L. S. − Trombley, M. A. (2003) Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics*, Vol. 69 (2), 355−373.

Amihud, Y. (2002) Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial markets*, Vol. 5 (1), 31−56.

Amihud, Y. − Mendelson, H. (1986) Asset pricing and the bid-ask spread. *Journal of Financial Economics*, Vol. 17 (2), 223−249.

Ando, T. (2009) Bayesian portfolio selection using a multifactor model. *International Journal of Forecasting*, Vol. 25 (3), 550−566.

Ando, T. (2012) Bayesian portfolio selection under a multifactor asset return model with predictive model selection. *Global Business and Economics Review*, Vol. 14 (1−2), 77−101.

Ang, A. (2014) *Asset management: A systematic approach to factor investing.* Oxford University Press, Oxford.

Avramov, D. − Chao, J. C. (2006) An exact Bayes test of asset pricing models with application to international markets. *The Journal of Business*, Vol. 79 (1), 293−324.

Avramov, D. − Chordia, T. (2006) Asset pricing models and financial market anomalies. *The Review of Financial Studies*, Vol. 19 (3), 1001−1040.

Banz, R. (1981) The relationship between return and market value of common stocks. *Journal of Financial Economics,* Vol. 9 (1), 3−18.

Barbee Jr, W. C. − Mukherji, S. − Raines, G. A. (1996) Do sales–price and debt–equity explain stock returns better than book–market and firm size? *Financial Analysts Journal*, Vol. 52 (2), 56−60.

Barnard, G. A. (1963) New methods of quality control. *Journal of the Royal Statistical Society. Series A (General)*, Vol. 126 (2), 255−258.

Barroso, P. − Santa-Clara, P. (2015) Momentum has its moments. *Journal of Financial Economics*, Vol. 116 (1), 111−120.

Basu, S. (1977) Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*, Vol. 32 (3), 663−682.

Bauman, W. S. − Conover, C. M. − Miller, R. E. (1998) Growth versus value and large-cap versus small-cap stocks in international markets. *Financial Analysts Journal*, Vol. 54 (2), 75−89.

Bender, J. − Briand, R. − Melas, D. − Subramanian, R. (2013) Foundations of factor investing. *MSCI Research Insight*, 1−33.

Berger, J. O. − Pericchi, L. R. − Ghosh, J. K. − Samanta, T. − De Santis, F. (2001) Objective Bayesian methods for model selection: introduction and comparison. *Lecture Notes-Monograph Series*, Vol. 38, 135−207.

Bhattacharya, D. − Li, W. H. − Sonaer, G. (2017) Has momentum lost its momentum? *Review of Quantitative Finance and Accounting*, Vol. 48 (1), 191−218.

Bianchi, D. − Guidolin, M. − Ravazzolo, F. (2017) Macroeconomic factors strike back: A Bayesian change-point model of time-varying risk exposures and premia in the US cross-section. *Journal of Business & Economic Statistics*, Vol. 35 (1), 110−129.

Black, A. J. − Fraser, P. (2003) The value premium: rational, irrational or random. *Managerial Finance*, Vol. 29 (10), 57−75.

Blume, M. E. − Stambaugh, R. F. (1983) Biases in computed returns: An application to the size effect. *Journal of Financial Economics*, Vol. 12 (3), 387−404.

Bollen, K. A. − Long, J. S. (1993) *Testing structural equation models*. Sage Publications Inc, California.

Campbell, J. Y. − Vuolteenaho, T. (2004) Bad beta, good beta. *American Economic Review*, Vol. 94 (5), 1249−1275.

Carhart, M. (1997) On persistence in mutual fund performance. *The Journal of Finance,* Vol. 52 (1), 57−82.

Cederburg, S. − O'Doherty, M. S. (2015) Asset-pricing anomalies at the firm level. *Journal of Econometrics*, Vol. 186 (1), 113−128.

Chaibi, A. − Alioui, S. − Xiao, B. (2015) On the impact of firm size on risk and return: Fresh evidence from the American stock market over the recent years. *Journal of Applied Business Research*, Vol. 31 (1), 29−36.

Chan, K. C. − Chen, N. F. − Hsieh, D. A. (1985) An exploratory investigation of the firm size effect. *Journal of Financial Economics*, Vol. 14 (3), 451−471.

Chan, L. K. − Karceski, J. − Lakonishok, J. (2000) New paradigm or same old hype in equity investing? *Financial Analysts Journal*, Vol. 56 (4), 23−36.

Chan, L. K. − Lakonishok, J. (2004) Value and growth investing: Review and update. *Financial Analysts Journal*, Vol. 60 (1), 71−86.

Chen, N. F. (1983) Some empirical tests of the theory of arbitrage pricing. *The Journal of Finance,* Vol. 38 (5), 1393−1414.

Chen, N. F. − Roll, R. − Ross, S. A. (1986) Economic forces and the stock market. *The Journal of Business*, Vol. 59 (3), 383–403.

Chen, N. F. − Zhang, F. (1998) Risk and return of value stocks. *The Journal of Business*, Vol. 71 (4), 501−535.

Chen, Y. − Zhao, H. (2012) Informed trading, information uncertainty, and price momentum. *Journal of Banking & Finance*, Vol. 36 (7), 2095−2109.

Cheng, A. C. (1996) Economic factors and stock markets: Empirical evidence from the UK and the US. *International Journal of Finance & Economics*, Vol. 1 (4), 287−302.

Chung, Y. T. − Hsu, C. H. − Ke, M. C. − Liao, T. L. − Chiang, Y. C. (2016) The weakening value premium in the Australian and New Zealand stock markets. *Pacific-Basin Finance Journal*, Vol. 36 (C), 123−133.

Connor, G. (1995) The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal*, Vol. 51 (3), 42−46.

Conrad, J. − Cooper, M. − Kaul, G. (2003) Value versus glamour. *The Journal of Finance*, Vol. 58 (5), 1969−1995.

Cremers, K. M. (2002) Stock return predictability: A Bayesian model selection perspective. *The Review of Financial Studies*, Vol. 15 (4), 1223−1249.

Cremers, K. M. (2006) Multifactor efficiency and Bayesian inference. *The Journal of Business*, Vol. 79 (6), 2951−2998.

Czaja, M. G. − Scholz, H. (2007) Sensitivity of stock returns to changes in the term structure of interest rates − evidence from the German market. In: *Operations Research Proceedings 2006*, eds. K-H. Waldmann – U. M. Stocker. Springer, Berlin, Heidelberg.

Daniel, K. − Hirshleifer, D. − Subrahmanyam, A. (1998) Investor psychology and security market under-and overreactions. *The Journal of Finance*, Vol. 53 (6), 1839−1885.

Das, S. − Barai, P. (2016) Size, value and momentum in stock returns: Evidence from India. *Macroeconomics and Finance in Emerging Market Economies*, Vol. 9 (3), 284−302.

Dhatt, M. S. − Kim, Y. H. − Mukherji, S. (1999) The value premium for small-capitalization stocks. *Financial Analysts Journal*, Vol. 55 (5), 60−68.

van Dijk, M. A. (2011) Is size dead? A review of the size effect in equity returns. *Journal of Banking & Finance,* Vol. 35 (12), 3263−3274.

Dimson, E. − Marsh, P. (1999) Murphy's law and market anomalies. *Journal of Portfolio Management*, Vol. 25 (2), 53−69.

Dimson, E. − Nagel, S. − Quigley, G. (2003) Capturing the value premium in the United Kingdom. *Financial Analysts Journal*, Vol. 59 (6), 35−45.

Dissanaike, G. (2002) Does the size effect explain the UK winner-loser effect? *Journal of Business Finance & Accounting*, Vol. 29 (1−2), 139−154.

Drew, M. (2003) Beta, firm size, book-to-market equity and stock returns. *Journal of the Asia Pacific Economy*, Vol. 8 (3), 354−379.

Ericsson, J. − Karlsson, S. (2003) Choosing factors in a multifactor asset pricing model: A Bayesian approach. *SSE/EFI Working Paper Series in Economics and Finance,* No. 524, 1−46.

Fama, E. F. (1970) Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, Vol. 25 (2), 383−417.

Fama, E. F. (1998) Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, Vol. 49 (3), 283−306.

Fama, E. F. − French, K. R. (1992) The cross-section of expected stock returns. *The Journal of Finance*, Vol. 47 (2), 427−465.

Fama, E. F. − French, K. R. (1993) Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, Vol. 33 (1), 3−56.

Fama, E. F. − French, K. R. (1995) Size and book-to-market factors in earnings and returns. *The Journal of Finance,* Vol. 50 (1), 131−155.

Fama, E. F. − French, K. R. (1996) Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, Vol. 51 (1), 55−84.

Fama, E. F. − French, K. R. (2004) The capital asset pricing model: Theory and evidence. *The Journal of Economic Perspectives*, Vol. 18 (3), 25−46.

Fama, E. F. − French, K. R. (2012) Size, value, and momentum in international stock returns. *Journal of Financial Economics*, Vol. 105 (3), 457−472.

Fama, E. F. – French, K. R. (2015) A Five-factor Asset Pricing Model. *Journal of Financial Economics,* Vol. 116 (1), 1–22.

Fama, E. F. – MacBeth, J. D. (1973) Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy,* Vol. 81 (3), 607–636.

Fernandez, C. − Ley, E. − Steel, M. F. (2001) Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, Vol. 100 (2), 381−427.

Ferson, W. E. − Harvey, C. R. (1991) The variation of economic risk premiums. *Journal of Political Economy*, Vol. 99 (2), 385−415.

Findlay, M. C. − Williams, E. E. (2000) A fresh look at the efficient market hypothesis: How the intellectual history of finance encouraged a real "fraud-on-the-market". *Journal of Post Keynesian Economics*, Vol. 23 (2), 181−199.

Flannery, M. J. − Protopapadakis, A. A. (2002) Macroeconomic factors do influence aggregate stock returns. *The Review of Financial Studies*, Vol. 15 (3), 751−782.

Galsband, V. (2012) Downside risk of international stock returns. *Journal of Banking & Finance*, Vol. 36 (8), 2379−2388.

Gaunt, C. (2004) Size and book to market effects and the Fama French three factor asset pricing model: evidence from the Australian stock market. *Accounting & Finance*, Vol. 44 (1), 27−44.

Geisser, S. − Cornfield, J. (1963) Posterior distributions for multivariate normal parameters. *Journal of the Royal Statistical Society, Series B (Methodological),* Vol. 25 (2), 368−376.

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis,* Vol. 1 (3), 515−534.

Griffin, J. M. − Ji, X. − Martin, J. S. (2003) Momentum investing and business cycle risk: Evidence from pole to pole. *The Journal of Finance*, Vol. 58 (6), 2515−2547.

Guerard, J. B. − Xu, G. − Gültekin, M. (2012) Investing with momentum: The past, present, and future. *The Journal of Investing*, Vol. 21 (1), 68−80.

Hall, A. D. − Hwang, S. − Satchell, S. E. (2002) Using Bayesian variable selection methods to choose style factors in global stock return models. *Journal of Banking & Finance*, Vol. 26 (12), 2301−2325.

Harvey, C. R. − Zhou, G. (1990) Bayesian inference in asset pricing tests. *Journal of Financial Economics*, Vol. 26 (2), 221−254.

Harvey, C. R. − Liechty, J. C. − Liechty, M. W. − Müller, P. (2010) Portfolio selection with higher moments. *Quantitative Finance*, Vol. 10 (5), 469−485.

Hoeting, J. − Madigan, D. − Raftery, A. − Volinsky, C. (1999) Bayesian model averaging: A tutorial. *Statistical Science*, Vol. 14 (4), 382−401.

Hooker, M. A. (2004) Macroeconomic factors and emerging market equity returns: A Bayesian model selection approach. *Emerging markets review*, Vol. 5 (4), 379−387.

Houge, T. − Loughran, T. (2006) Do Investors Capture the Value Premium? *Financial Management*, Vol. 35 (2), 5−9.

Hwang, S. − Rubesam, A. (2013) A behavioral explanation of the value anomaly based on time-varying return reversals. *Journal of Banking & Finance*, Vol. 37 (7), 2367−2377.

Hwang, S. − Rubesam, A. (2015) The disappearance of momentum. *The European Journal of Finance*, Vol. 21 (7), 584−607.

Hwang, T. − Gao, S. − Owen, H. (2014) Markowitz efficiency and size effect: Evidence from the UK stock market. *Review of Quantitative Finance and Accounting*, Vol. 43 (4), 721−750.

Iranmanesh, A. − Arashi, M – Tabatabaey, S. M. M. (2010) On conditional applications of matrix variate normal distribution. *Iranian Journal of Mathematical Sciences and Informatics*, Vol. 5 (2), 33−43.

Jareño, F. − Negrut, L. (2016) US stock market and macroeconomic factors. *Journal of Applied Business Research*, Vol. 32 (1), 325−340.

Jegadeesh, N. − Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, Vol. 48 (1), 65−91.

Jegadeesh, N. − Titman, S. (2001) Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance*, Vol. 56 (2), 699−720.

Kaneko, T. − Lee, B. S. (1995) Relative importance of economic factors in the US and Japanese stock markets. *The Journal of the Japanese and International Economies,* Vol. 9 (3), 290−307.

Kang, J. − Kim, T. S. − Lee, C. − Min, B. K. (2011) Macroeconomic risk and the cross-section of stock returns. *Journal of Banking & Finance*, Vol. 35 (12), 3158−3173.

Kartašova, J. − Remeikiene, R. − Gaspareniene, L. − Venclauskienė, D. (2014) Transformations of efficient market hypothesis under the influence of behavioral finance. *Mediterranean Journal of Social Sciences*, Vol. 5 (13), 327−333.

Keim, D. B. (1983) Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics*, Vol. 12 (1), 13−32.

Kenneth French's data library, *URL: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html*

Kim, M. K. − Burnie, D. A. (2002) The firm size effect and the economic cycle. *Journal of Financial Research*, Vol. 25 (1), 111−124.

Lakonishok, J. − Shleifer, A. − Vishny, R. W. (1994) Contrarian investment, extrapolation, and risk. *The Journal of Finance*, Vol. 49 (5), 1541−1578.

Lee, E. − Strong, N. − Zhu, Z. J. (2014) Did the value premium survive the subprime credit crisis? *The British Accounting Review*, Vol. 46 (2), 166−178.

Li, X. − Brooks, C. − Miffre, J. (2009) The value premium and time-varying volatility. *Journal of Business Finance & Accounting*, Vol. 36 (9−10), 1252−1272.

Lim, B. Y. − Wang, J. G. − Yao, Y. (2018) Time-series momentum in nearly 100 years of stock returns. *Journal of Banking & Finance*, Vol. 97, 283−296.

Lintner, J. (1965) The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, Vol. 47 (1), 13−37.

Liu, W. (2006) A liquidity-augmented capital asset pricing model. *Journal of Financial Economics*, Vol. 82 (3), 631−671.

Llera, A. − Beckmann, C. F. (2016) Estimating an inverse gamma distribution. *Donders Institute Technical Report*, 1−16.

Loughran, T. (1997) Book-to-market across firm size, exchange, and seasonality: Is there an effect? *Journal of Financial and Quantitative Analysis*, Vol. 32 (3), 249−268.

Madigan, D. − Raftery, A. E. (1994) Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, Vol. 89 (428), 1535−1546.

Maio, P. − Philip, D. (2015) Macro variables and the components of stock returns. *Journal of Empirical Finance*, Vol. 33, 287−308.

Malkiel, B. G. (2003) The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, Vol. 17 (1), 59−82.

Malkiel, B. G. (2005) Reflections on the efficient market hypothesis: 30 years later. *Financial Review*, Vol. 40 (1), 1−9.

Markowitz, H. (1952) Portfolio selection. *The Journal of Finance*, Vol. 7 (1), 77−91.

Marquering, W. − Nisser, J. − Valla, T. (2006) Disappearing anomalies: A dynamic analysis of the persistence of anomalies. *Applied Financial Economics*, Vol. 16 (4), 291−302.

McCulloch, R. − Rossi, P. E. (1991) A Bayesian approach to testing the arbitrage pricing theory. *Journal of Econometrics,* Vol. 49 (1−2), 141−168.

Merton, R. C. (1973) An intertemporal capital asset pricing model. *Econometrica*, Vol. 41 (5), 867−887.

Mills, T. C. − Jordanov, J. V. (2003) The size effect and the random walk hypothesis: Evidence from the London Stock Exchange using Markov chains. *Applied Financial Economics*, Vol. 13 (11), 807−815.

Mossin, J. (1966) Equilibrium in a Capital Asset Market. *Econometrica*, Vol. 34 (4), 768−783.

Narayan, P. K. − Gupta, R. (2015) Has oil price predicted stock returns for over a century? *Energy Economics*, Vol. 48, 1−23.

Nydick, S. W. (2012) The Wishart and inverse Wishart distributions. *Electronic Journal of Statistics,* Vol. 6, 1−19.

Pandey, A. − Sehgal, S. (2016) Explaining size effect for Indian stock market. *Asia-Pacific Financial Markets*, Vol. 23 (1), 45−68.

Pástor, L. (2000) Portfolio selection and asset pricing models. *The Journal of Finance*, Vol. 55 (1), 179−223.

Pástor, L. − Stambaugh, R. F. (2000) Comparing asset pricing models: An investment perspective. *Journal of Financial Economics,* Vol. 56 (3), 335−381.

Patel, J. B. (2012) A further analysis of small firm stock returns. *Managerial Finance*, Vol. 38 (7), 653−659.

Petkova, R. (2006) Do the Fama–French factors proxy for innovations in predictive variables? *The Journal of Finance*, Vol. 61 (2), 581−612.

Phalippou, L. (2008) Where is the value premium? *Financial Analysts Journal*, Vol. 64 (2), 41−48.

Puga, J. L. − Krzywinski, M. − Altman, N. (2015) Bayes' theorem: Incorporate new evidence to update prior information. *Nature Methods*, Vol. 12 (4), 277−279.

Polson, N. G. − Tew, B. V. (2000) Bayesian portfolio selection: An empirical analysis of the S&P 500 index 1970–1996. *Journal of Business & Economic Statistics*, Vol. 18 (2), 164−173.

Pätäri, E. − Leivo, T. (2017) A closer look at value premium: Literature review and synthesis. *Journal of Economic Surveys*, Vol. 31 (1), 79−168.

Qi, M. − Maddala, G. S. (1999) Economic factors and the stock market: A new perspective. *Journal of Forecasting*, Vol. 18 (3), 151−166.

Raftery, A. E. − Madigan, D. − Hoeting, J. A. (1997) Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, Vol. 92 (437), 179−191.

Raftery, A. E – Hoeting, J. A. – Volinsky, C − Painter, I − Yeung, K. Y. (2018) BMA R package, version 3.18.9, *URL: http://stats.research.att.com/volinsky/bma.html*

Reinganum, M. R. (1981) Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. *Journal of Financial Economics*, Vol. 9 (1), 19−46.

Reinganum, M. R. (1982) A direct test of Roll's conjecture on the firm size effect. *The Journal of Finance*, Vol. 37 (1), 27−35.

Richardson, S. − Tuna, I. − Wysocki, P. (2010) Accounting anomalies and fundamental analysis: A review of recent research advances. *Journal of Accounting and Economics*, Vol. 50 (2−3), 410−454.

Roll, R. (1977) A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics*, Vol. 4 (2), 129−176.

Roll, R. − Ross, S. A. (1980) An empirical investigation of the arbitrage pricing theory. *The Journal of Finance*, Vol. 35 (5), 1073−1103.

Rosenberg, B. − Reid, K. − Lanstein, R. (1985) Persuasive evidence of market inefficiency. *The Journal of Portfolio Management,* Vol. 11 (3), 9−16.

Rouder, J. N. − Morey, R. D. (2018) Teaching Bayes’ theorem: Strength of evidence as predictive accuracy. *The American Statistician*, 1−5.

Rubinstein, M. (2002) Markowitz's “portfolio selection”: A fifty-year retrospective. *The Journal of Finance*, Vol. 57 (3), 1041−1045.

Rutledge, R. W. − Zhang, Z. − Karim, K. (2008) Is there a size effect in the pricing of stocks in the Chinese stock markets? The case of bull versus bear markets. *Asia-Pacific Financial Markets*, Vol. 15 (2), 117−133.

Sadka, R. (2006) Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *Journal of Financial Economics*, Vol. 80 (2), 309−349.

Savor, P. − Wilson, M. (2013) How much do investors care about macroeconomic risk? Evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis*, Vol. 48 (2), 343−375.

Scowcroft, A. − Sefton, J. (2005) Understanding momentum. *Financial Analysts Journal*, Vol. 61 (2), 64−82.

Sehgal, S. − Subramaniam, S. − Deisting, F. (2014) Tests of equity market anomalies for select emerging markets. *The International Journal of Business and Finance Research*, Vol. 8 (3), 27−46.

Sharpe, W. (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, Vol. 19 (3), 425–442.

Sharpe, W. (1994) The Sharpe ratio. *The Journal of Portfolio Management*, Vol. 21 (1), 49−58.

Shiller, R. J. (2003) From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, Vol. 17 (1), 83−104.

Skočir, M. − Lončarski, I. (2018) Multi-factor asset pricing models: Factor construction choices and the revisit of pricing factors. *Journal of International Financial Markets, Institutions and Money*, Vol. 55, 65−80.

Soyer, R. − Tanyeri, K. (2006) Bayesian portfolio selection with multi-variate random variance models. *European Journal of Operational Research*, Vol. 171 (3), 977−990.

Steel, M. F. (2011) Bayesian model averaging and forecasting. *Bulletin of EU and US Inflation and Macroeconomic Analysis*, Vol. 200, 30−41.

Tai, C. S. (2003) Are Fama–French and momentum factors really priced? *Journal of Multinational Financial Management*, Vol. 13 (4-5), 359−384.

Tsai, C. L. − Wang, H. − Zhu, N. (2010) Does a Bayesian approach generate robust forecasts? Evidence from applications in portfolio investment decisions. *Annals of the Institute of Statistical Mathematics*, Vol. 62 (1), 109−116.

Vassalou, M. − Xing, Y. (2004) Default risk in equity returns. *The Journal of Finance*, Vol. 59 (2), 831−868.

Woźniak, T. (2016) Bayesian vector autoregressions. *Australian Economic Review*, Vol. 49 (3), 365−380.

Wu, H. (2011) The value and size effect − are there firm-specific-risks in China's domestic stock markets? *International Journal of Economics and Finance*, Vol. 3 (3), 26−37.

Xie, S. − Qu, Q. (2016) The three-factor model and size and value premiums in China's stock market. *Emerging Markets Finance and Trade*, Vol. 52 (5), 1092−1105.

Young, M. R. − Lenk, P. J. (1998) Hierarchical Bayes methods for multifactor model estimation and portfolio selection. *Management Science*, Vol. 44 (11, part 2), 111−124.