Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Using Model Averaging Techniques to Examine Determinants of Stock Returns

Author: **Bc. Miriama Tóthová**

Supervisor: **PhDr. Zuzana Havránková, Ph.D.**

Academic Year: **2017/2018**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 31, 2018

Signature

Acknowledgments

I would like to express my biggest gratitude especially to my thesis supervisor PhDr. Zuzana Havránková, Ph.D. and doc. PhDr. Tomáš Havránek, Ph.D. for their valuable comments and suggestions throughout the process of writing this thesis as well as for the help with R code.

I am also very grateful to my parents and my sister for their endless support, provided during all my studies.

Abstract

The predictability of stock returns has been a widely discussed topic in the financial literature. In the presented thesis, we examine the effect of 20 possible predictors on S&P 500 excess returns in the time period from June 1998 till December 2016. However, traditional models examining stock returns usually ignore the issue of model uncertainty. In order to explicitly incorporate uncertainty about the model into the analysis, we employ two model averaging techniques, in particular Bayesian model averaging (BMA) and frequentist model averaging (FMA). As a robustness check we use three different combinations of priors within BMA framework. We assess the quality of their predictions and compare the results with the traditional methods based on model selection criteria. We find out that among the most important variables explaining excess returns on S&P 500 stock index are three-month Treasury bill rate, dividend yield, term premium, payout ratio, excess returns lagged twice, and default risk premium. These are robust across all models we have estimated. Although frequentist model averaging provides in-sample predictions superior to BMA as the literature suggests and it also performs better than models selected according to popular statistical criteria, it fails to outperform the Bayesian model averaging out-of-sample. Moreover, among the three BMA specifications, the data-dependent hyper-g prior performs the best in-sample and out-of-sample.

JEL Classification	C11, C52, G12
Keywords	stock return predictability, model uncertainty, Bayesian model averaging, frequentist model averaging

Author's e-mail	`tothovamiriama@gmail.com`
Supervisor's e-mail	`zuzana.havrankova@fsv.cuni.cz`

Abstrakt

Predpovedateľnosť výnosov akcií je veľmi často diskutovaná téma v súčasnej finančnej literatúre. V prezentovanej práci skúmame efekt dvadsiatich možných determinantov na nadmerné výnosy S&P 500 indexu v období od júna 1998 do decembra 2016. Tradičné modely skúmajúce výnosy akcií však často ignorujú problém modelovej neistoty. Preto, aby sme do analýzy zahrnuli neistotu ohľadom správneho modelu, používame dve metódy priemerovania modelov, konkrétne Bayesovské priemerovanie modelov (BMA) a frekventistické priemerovanie modelov (FMA). Ako kontrolu robustnosti používame vrámci BMA tri rôzne kombinácie priorov. Hodnotíme kvalitu ich predpovedí a porovnávame výsledky s tradičnými metódami, ktoré vyberajú najlepší model na základe štatistických kritérií. Zistili sme, že medzi najdôležitejšie premenné, ktoré vysvetľujú nadmerné výnosy na akciovom indexe S&P 500, sú trojmesačná úroková sadzba pokladničných poukážok, výnosy z dividend, časová prémia, výplatný pomer, nadmerné výnosy oneskorené o dve obdobia a riziková prémia. Tieto determinanty sú robustné naprieč všetkými odhadovanými modelmi. Napriek tomu, že frekventistické priemerovanie modelov prekonáva Bayesovské priemerovanie modelov na základe testovania in-sample ako literatúra naznačuje a tiež prekonáva modely vybrané podľa populárnych štatistických kritérií, jeho predikcie nie sú lepšie ako Bayesovské na základe out-of-sample testovania. Navyše, medzi tromi BMA špecifikáciami, hyper-g prior prekonáva ostatné BMA alternatívy in-sample a tiež out-of-sample.

Klasifikace JEL	C11, C52, G12
Klíčová slova	predpovedateľnosť výnosnosti akcií, modelová neistota, Bayesovské priemerovanie modelov, frekventistické priemerovanie modelov

E-mail autora	`tothovamiriama@gmail.com`
E-mail vedoucího práce	`zuzana.havrankova@fsv.cuni.cz`

Contents

List of Tables **viii**

List of Figures **ix**

Acronyms **x**

Thesis Proposal **xi**

Introduction **1**

1 Literature Review **4**

1.1 Determinants of stock returns 4

1.2 Model averaging techniques 10

1.2.1 Bayesian Model Averaging 11

1.2.2 Frequentist Model Averaging 13

2 Data **16**

2.1 Data construction 16

2.2 Data sources 17

2.3 Descriptive statistics 19

3 Methodology **21**

3.1 Bayesian Model Averaging 21

3.1.1 Priors 23

3.1.2 MCMC sampler 25

3.2 Frequentist Model Averaging 27

3.3 Traditional model selection 28

3.4 Predictive performance 30

4 Results **32**

4.1 Baseline results 32

4.2 Alternative BMA specifications 42
4.3 Frequentist Model Averaging 48
4.4 Model Selection Criteria 50
4.5 Predictive Performance 53

Conclusion **57**

Bibliography **66**

A Additional tables and figures **I**

List of Tables

1.1 Determinants examined in published articles 9

2.1 Description of predictive variables 18
2.2 Descriptive statistics 19

4.1 Model diagnostics, UIP + uniform model prior 33
4.2 Top 3 models inclusion, UIP + uniform model prior 37
4.3 Results from BMA, baseline specification 38
4.4 Results from BMA, alternative specifications 46
4.5 Results from frequentist model averaging 49
4.6 Inclusion of predictors according to model selection criteria . . . 52
4.7 Root Mean Square Error - different models 54
4.8 Mean Absolute Error - different models 55
4.9 Logarithmic predictive score - different models 56

A.1 List of predictive variables and their abbreviations I
A.2 Correlation matrix of examined variables II
A.3 Correlation matrix of examined variables - continued III
A.4 BMA baseline specification using reverse-jump sampler IV
A.5 Frequentist check (OLS with variables having PIP>0.5) V
A.6 Comparison of two studies - Avramov (2002) and Cremers (2002) VI
A.7 BMA with term spread and term premium separately VII
A.8 Model diagnostics, all three BMA specifications VII
A.9 Sign certainty of predictors XII
A.10 Top 3 models inclusion - BMA alternative specifications XII
A.11 Summary of best submodels in model selection XIII
A.12 OLS with all variables lagged XIV
A.13 Summary of model averaging techniques (Coefficient, StDev) . . XV

List of Figures

4.1 Model size and convergence, UIP + uniform model prior 34
4.2 Model inclusion in BMA, UIP + uniform model prior 35
4.3 Model inclusion in BMA, BRIC + random model prior 44
4.4 Model inclusion in BMA, hyper-g + random model prior 44
4.5 Model comparison of different BMA alternatives 47
4.6 Plot of best submodels ordered according to BIC 51

A.1 PMP convergence based on 20 best models IV
A.2 Model size and convergence, BRIC + random model prior . . . VIII
A.3 Model size and convergence, hyper-g + random model prior . . IX
A.4 Plot of best submodels ordered according to Adjusted R2 X
A.5 Plot of best submodels ordered according to Cp X
A.6 Choosing best submodel according to statistical criteria (full sample) XI

Acronyms

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

BMA Bayesian Model Averaging

DJIA Dow Jones Index Average

FIC Fischer Information Criterion

FMA Frequentist Model Averaging

LPS Log-predictive score

MAE Mean Absolute Error

MCMC Markov Chain Monte Carlo

MDC Markov Dimension Criterion

MMA Mallows Model Averaging

PIC Posterior Information Criterion

PLS Predictive Least Squares Criterion

PIP Posterior Inclusion Probability

PMP Posterior Model Probability

UIP Unit Information Prior

RMSE Root Mean Square Error

WALS Weighted-average Least Squares

Master's Thesis Proposal

Author	Bc. Miriama Tóthová
Supervisor	PhDr. Zuzana Havránková, Ph.D.
Proposed topic	Using Model Averaging Techniques to Examine Determinants of Stock Returns

Motivation In recent years, the predictability of stock returns has become one of the most discussed issue of empirical asset pricing. Although there are various studies documenting predictability, these theories are not consistent in which important conditioning variables should enter the predictive regression. For example, the existing evidence on the size effect is not unambiguous. Researchers often differ substantially in results. De Moor & Sercu (2013) identify a post-1980s size effect which is persistent, however it has not been found by a Fama-French-style SMB. Besides the firm size, there are several similar cases, for instance key financial ratios (dividend-price ratio, dividend-yield, earnings-price ratio, dividend-payout ratio). Some scholars believe that these financial indicators have ability to predict stock return (e.g., Lettau and Ludvigson, 2001). On the other hand, Charles et al. (2016) have found that, for a wide range of international stock markets, a momentum indicator (price pressure) and short-term interest rate show a substantially higher degree of predictive ability for stock return than the financial ratios, both in-sample and out-of-sample.

Moreover, data analysis traditionally begins with model selection and then one makes inference under the selected model as if the specific model had been previously given. However, this traditional estimation procedure ignores the model uncertainty. More appropriate alternative may be using a model averaging technique which allows us to explicitly incorporate uncertainty about correct model into the estimation process by comparing all possible models simultaneously (Cremers, 2002). Hence it is robust to model misspecification, at least within the universe of linear forecasting models.

In this study, I will implement two model averaging techniques, specifically the Bayesian model averaging (BMA) and frequentist model averaging (FMA). In com-

parison with the FMA technique, there exists a vast literature implementing the BMA approach. The existing literature found the evidence that Bayesian model averaging approach gives a better result by comparing with other traditional strategies (for example, Cremers (2002), Avramov (2002), Chairakwattana & Nathaphan (2013) etc.)

Hypotheses Three principal hypotheses that I would like to test in the thesis are following:

> Hypothesis #1: A model averaging technique provides superior predictions to these of model selection criteria.
>
> Hypothesis #2: Frequentist model averaging approach outperforms Bayesian model averaging.
>
> Hypothesis #3: Within BMA specifications, the hyper-g prior specification performs the best.

Methodology In the thesis, I will use two model averaging techniques to analyze the sample evidence on return predictability in the presence of model uncertainty using freely available software in the statistical environment R. In particular, I will compare the results from traditional estimation procedure with estimates obtained using recently developed frequentist and alternative Bayesian model averaging techniques. The analysis will be based on data from Thomson Reuters database for selected stock market index.

The first model averaging technique will be the Bayesian model averaging approach where the uncertainty on model is acknowledged by setting a prior probability to each individual model. However, Hjort and Claeskens (2003) argue that there exist several difficulties with using BMA approach, such as how to set prior probabilities. In contrast with BMA, the FMA approach requires no priors and the corresponding estimators are fully determined by data. For that reason, the FMA approach has received much more attention in recent years (Wang et al. (2009) or Hjort and Claeskens (2003)). I will follow the approach presented in Amini & Parameter (2012) who suggest to orthogonalize the covariate space in order to make the analysis computationally feasible by reducing the total number of estimated models.

Moreover, within the BMA alternatives, I will compare the baseline BMA specification using the unit information g-prior to the hyper-g prior specification proposed by Feldkircher & Zeugner (2012). The main advantage of the hyper-g specification is that it should be more robust to outliers in the data. Predictors for stock return

predictability contain financial information and will be chosen according to the evidence from the previous literature. Various different regressors have been identified as having predictive power (see, for example, Fama, 1991).

Expected Contribution The application of the frequentist model averaging approach can be dated back to the work of Bates and Granger (1969) on forecast combination. Recently, it has been used in many economic fields (see Wang et al., 2009). I will contribute to the topic by extending the research into the context of stock return predictability where it has not been employed yet. Moreover, I will compare both model averaging methods with the traditional model selection criteria in new markets and using new predictive variables. Therefore, I will try to account for the uncertainty which is present in the financial literature not only about the correct model for excess stock returns, but also about which predictors to look at initially.

Outline The structure of the thesis will be as follows:

1. **Introduction** will provide the motivation behind the importance of accounting for model uncertainty in the estimation procedure while predicting excess stock returns.

2. **Literature Review:** This section will review the relevant literature on stock return predictability, Bayesian model averaging as well as frequentist model averaging.

3. **Data:** This section will describe the sample data used in the analysis with an overview of the variables.

4. **Methodology** will present econometrical framework for both model averaging techniques used in the study for testing the hypotheses.

5. **Results:** The main empirical results will be provided and discussed in detail.

6. **Conclusion:** I will summarize the most important findings of the thesis and suggest future improvements for similar researches.

Core bibliography

Amini, S. M. (2012). Model Uncertainty & Model Averaging Techniques (Doctoral dissertation, Virginia Tech).

Amini, S. M., & Parmeter, C. F. (2011). Bayesian model averaging in R. Journal of Economic and Social Measurement, 36(4), 253-287.

Amini, S. M., & Parmeter, C. F. (2012). Comparison of model averaging techniques: Assessing growth determinants. Journal of Applied Econometrics, 27(5), 870-876.

Avramov, D. (2002). Stock return predictability and model uncertainty. Journal of Financial Economics, 64(3), 423-458.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. Or, 451-468.

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

De Moor, L., & Sercu, P. (2013). The smallest firm effect: An international study. Journal of International Money and Finance, 32, 129-155.

Fama, E. F., 1991, Efficient Capital Markets: II, Journal of Finance, 46, 1575-1618.

Feldkircher, M., & Zeugner, S. (2012). The impact of data revisions on the robustness of growth determinants-a note on 'determinants of economic growth: Will data tell?'. Journal of applied econometrics, 27(4), 686-694.

Hjort, N. L., & Claeskens, G. (2003). Frequentist model average estimators. Journal of the American Statistical Association, 98(464), 879-899.

Chairakwattana, K., & Nathaphan, S. (2013). Stock Return Predictability by Bayesian Model Averaging: Evidence from Stock Exchange of Thailand. Faculty of Commerce and Accountancy, Thammasat University.

Charles, A., Darné, O., & Kim, J. H. (2016). International stock return predictability: Evidence from new statistical tests. International Review of Financial Analysis.

Lettau, M., & Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. the Journal of Finance, 56(3), 815-849.

Wang, H., Zhang, X., & Zou, G. (2009). Frequentist model averaging estimation: a review. Journal of Systems Science and Complexity, 22(4), 732.

Author

Supervisor

Introduction

In recent years, the predictability of stock returns has become one of the most discussed issue of empirical asset pricing. From an economic perspective, examining the question of whether returns on stock market are predictable is important for several reasons. The ability to predict stock returns would be beneficial for an individual investor since it may improve trading and asset allocation decisions. In fact, the evidence in the financial literature suggests that an investor may suffer considerable welfare losses if the asset allocation decisions are not made conditional upon predictions of future returns (Schrimpf 2010). Besides, examining stock market return predictability is of considerable importance for academia as well. The predictive equation explaining stock market returns facilitates the understanding of the dynamics of asset price movements as well as their causes (McMillan *et al.* 2015).

Although there are various studies documenting predictability, these theories are not consistent in which important conditioning variables should enter the predictive regression. For example, many previous empirical works suggest that lagged excess returns have the predictive ability to explain excess returns. However, there exist as many other studies which indicate that this variable has no power to predict excess returns at all (Zhou 2011). Moreover, the existing evidence on the size effect is not unambiguous as well. Researchers often differ substantially in results. De Moor & Sercu (2013) identify a post-1980s size effect which is persistent, however it has not been found by a Fama-French-style SMB. Besides the firm size and lagged returns, there are several similar cases, for instance key financial ratios (dividend-price ratio, dividend-yield, earnings-price ratio, dividend-payout ratio). Some scholars believe that these financial indicators have ability to predict stock return (e.g. Lettau & Ludvigson (2001)). On the other hand, Charles *et al.* (2016) have found that, for a wide range of international stock markets, a momentum indicator (price pressure) and short-term interest rate show a substantially higher degree of predictive ability for

stock return than the financial ratios, both in-sample and out-of-sample.

Moreover, data analysis traditionally begins with model selection and then one makes inference under the selected model as if the specific model had been previously given. However, such traditional estimation procedure ignores the issue of model uncertainty. To avoid the overfitting problem, Bossaerts & Hillion (1999) suggest using model selection criteria. Nevertheless, most of those methods only choose models according to particular criterion, which overlook the model uncertainty as well and may yield worse out-of-sample predictions. A suitable alternative is using a model averaging technique. The introduction of this methodology allows us to explicitly incorporate model uncertainty into the estimation process by comparing all possible models simultaneously. It is therefore robust to model misspecification, at least within the universe of linear forecasting models.

The objective of the thesis is to contribute to the finance literature not only about the existence of stock return predictability but also about the source of the predictability under the model averaging framework. That means, we will focus on two main questions. Which determinants are the most important in predicting excess stock returns? And more importantly, how do model averaging techniques perform in explaining stock returns? The dataset used for the analysis covers the period from June 1998 to December 2016. The dependent variable is monthly excess return on S&P 500 index. We will examine the effect of 20 potential predictors which are believed to contain financial information and will be chosen according to the evidence from the previous literature.

In this study, we will implement two model averaging techniques, specifically the Bayesian model averaging (BMA) and frequentist model averaging (FMA). In comparison with the FMA approach, the Bayesian model averaging has already been widely used by many scholars. The existing literature found the evidence that Bayesian model averaging approach gives the better result by comparing with other traditional strategies. In the BMA approach, one incorporates the model uncertainty by setting a prior probability to each individual model. However, Hjort & Claeskens (2003) claim that there are several potential issues while using this approach, for instance how to set prior probabilities. Although the intuition is similar, in contrast with BMA, the FMA approach requires no priors and the corresponding estimators are fully determined by

data. For that reason, the FMA approach has received much attention in recent years. To our best knowledge, frequentist model averaging has not been studied in the context of stock return predictability yet. Therefore, we will contribute to the literature by extending the FMA research into the return predictability frame.

Furthermore, we will compare the baseline BMA specification employing the unit information prior in combination with a uniform model prior suggested by Eicher *et al.* (2011) with two BMA alternative specifications. For the first alternative BMA specification we will use the BRIC g-prior (following Fernandez *et al.* (2001a)) in combination with a beta-binomial prior that assigns the same probability on each model size. Secondly, we will employ the more informative hyper-g prior proposed by Feldkircher & Zeugner (2012) which should be more robust to outliers, and we will accompany it with the random model prior.

The thesis is organized as follows. Following the Introduction, Chapter 1 reviews the relevant literature, both within the universe of model averaging techniques as well as the determinants used for stock return predictability. Chapter 2 describes data and their source, and defines the variables of interest. Chapter 3 reports the methodology for analysis used in the thesis and states the hypotheses to be tested. Chapter 4 presents the regression results and compares them with findings from previous studies. Within the results, chosen model averaging techniques are compared, and the occurring differences are explained. The last chapter concludes and gives ideas for future research. The Appendix includes additional information, figures and complete tables of estimates.

Chapter 1

Literature Review

In this chapter, several key studies regarding the topic are presented. It is divided into two individual sections. In the first section we review the empirical works examining determinants of stock returns regardless the method they use. In the second section, the application of model averaging techniques is surveyed.

1.1 Determinants of stock returns

Recently, the stock return predictability has become one of the most discussed issue of empirical asset pricing. Fama (1991) provides a great summary of studies documenting stock return predictability until 1990s. In addition, Rapach & Zhou (2013) survey more recent literature, focusing on U.S equity data. They argue that recent studies use improved methods for stock return forecasting and thus provide significant out-of-sample results. Moreover, they survey the evidence of international predictability of stock returns, which is considered to have important implications not only for the development of investment management strategies, but for the development of asset pricing models as well.

Besides the numerous studies in favor of time-varying expected returns[1], there are several scholars with rather skeptical views on stock return predictability. For example, Lo & MacKinlay (1990) investigate the extent to which tests of financial asset pricing models may yield misleading inferences by using properties of the data to construct the test statistics. They show that the effect of some kind of data-snooping can be substantial using two empirical examples

[1]See, for example, Campbell (1987), Ferson & Harvey (1991), Richardson (1993), or more recent studies by Welch & Goyal (2007) and Gupta *et al.* (2014).

as well as Monte Carlo simulations and analytical calculations. Furthermore, Foster *et al.* (1997) provide theoretical analyses of data mining in the context of return predictability as well. Another study, written by Bossaerts & Hillion (1999), has attempted to protect against "data snooping" by using traditional selection criteria[2]. They implement different statistical model selection criteria, in particular adjusted R^2, AIC, BIC, PIC, and FIC, in order to confirm the stock return predictability and examine significant determinants of stock returns. Using international stock market data they have succeed to confirm in-sample stock return predictability, however, even the "best" model has lacked the out-of-sample predictability power.

As mentioned earlier, although there are various studies documenting predictability, these theories are not consistent in which important conditioning variables should enter the predictive regression. According to Cremers (2002), past returns and dividend yields are possibly the most commonly used predictors, however there exist fairly the same number of empirical studies ignoring these predictors and instead including different ones based on their own economic intuition and motivation. In addition, the existing evidence on the size effect is not unambiguous. Researchers often differ substantially in results. Probably the first systematic evidence of a size effect was provided by Banz (1981) using U.S. stock returns. The empirical work by De Moor & Sercu (2013) identifies a persistent post-1980s size effect based on a wide-coverage international dataset, carefully filtered and screened for errors. However such a result has not been found by a Fama-French-style SMB.

Besides the firm size, there are several similar cases, for instance, key financial ratios (dividend-price ratio, dividend-yield, earnings-price ratio, dividend-payout ratio). Some scholars believe that these financial indicators have ability to predict stock return (e.g. Lettau & Ludvigson (2001)), whereas there are also several of these that have found conflicting results (see Welch & Goyal (2007)). On the other hand, using monthly data from January 2000 to June 2014, Charles *et al.* (2016) have found that, for a large number of stock markets around the world, the technical indicator (particularly price pressure) and short-term interest rate are shown to be substantially stronger predictors of

[2]To protect against data snooping, one should consider using out-of-sample tests where statistical models are tested using out-of-sample observations which are not used in the estimation itself (Rapach & Wohar 2006).

stock returns than the financial ratios, both in-sample and out-of-sample.

Chen *et al.* (1986) in their study examine macroeconomic variables that are suggested by the financial theory to have the effect on stock market returns. Their analysis is, in fact, an imitation of the the two-step procedure introduced by Fama & MacBeth (1973) and is implemented to New York Stock Exchange returns. They found out that several variables significantly affect U.S. equity returns, most notably, industrial production, changes in the risk premium, twist in the yield curve, and, somewhat more weakly, measures of unanticipated inflation and changes in expected inflation during periods when these variables were highly volatile. Many other scholars have been influenced by the work of Chen *et al.* (1986) in using these macroeconomic variables in the arbitrage pricing theory (APT), for example Rjoub *et al.* (2009).

Campbell (1987) documents the fact that the term structure of interest rates predicts excess stock returns, along with excess returns on bills and bonds, using monthly U.S. data for the period of years 1959-1979 and 1979-1983. Furthermore, these observations are used to test asset pricing models. In a more recent study, Campbell & Yogo (2006) argue that conventional tests used for analyzing the stock return predictability may lead to invalid inferences. In particular, this might be the case when the variables used for predictability are persistent as their innovations are highly correlated with stock returns. Thus, they came with an efficient statistical test that might solve the problem regardless of the degree of predictors' persistence. Using this test, the general results find the evidence for reliable predictor variables including the dividend-price and smoothed earnings-price ratios as well as the short rate and the long-short yield spread.

Whitelaw (1994) examines the predictability of four variables, namely dividend yield, the one-year Treasury yield, the Baa-Aaa spread and the commercial paper-Treasury yield. The combination of the financial variables provides an empirical evidence of predictable variation in stock market returns.

The empirical work of Pontiff & Schall (1998) examines an aggregate measure of the book-to-market and its ability to reliably predict future stock returns. They find evidence that the book-to-market ratio of the Dow Jones Industrial Average predicts market returns and small firm excess returns during the

period 1926-1994. Additionally, their analysis includes some other predictive variables that have been previously shown to predict market returns, such as default spreads, interest rates, term structure slopes, and dividend yields. The DJIA book-to-market ratio has generally a stronger power to predict market returns than other variables such as interest yield spreads and dividend yields. However, their findings are sample specific in the way that the DJIA book-to-market ratio's predictive ability is present only in the pre-1960 period and after 1960, there is no significant relation. Contrary to the DJIA finding, the book-to-market ratio of the Standard and Poor's Industrial Index (S&P) seems to perform slightly better at forecasting market returns after 1960. In contrast, the S&P book-to-market ratio provides some predictive ability in the post-1960 period, although this relation is dramatically weaker than the Dow Jones pre-1960 findings.

Including the book-to-market ratio into the predictive regression is motivated by the findings of Fama & French (1992), who show that the book-to-market ratio of individual stocks has the ability to explain cross-sectional variation in stock returns. Another study by Kothari & Shanken (1997) independently documents the fact that the book-to-market ratio of the Dow Jones Industrial Index (DJIA) predicts market returns during the period of years 1926-1991 using a Bayesian framework. Furthermore, they found evidence about predictive ability of dividend yield, and that the dividend-yield relation is even stronger than book-to-market ratio relation in the subperiod 1941-1991.

Oil prices are generally considered to play an important role in affecting the U.S. economic activity (as suggested by Hamilton (1983) in a seminal work). Regarding the relationship between oil prices and stock returns, there does not exist any consensus in the literature. In earlier studies, for instance Ferson & Harvey (1995); Jones & Kaul (1996) and Sadorsky (1999) found the evidence that changes in oil prices have negative effect on U.S. stock market returns. On the contrary, the results of Huang *et al.* (1996) show no such relation between oil future returns and stock returns using daily data in the period of the 1980s and thus they confirm the findings of Chen *et al.* (1986).

Since then, many scholars have attempted to further examine the effect of oil price changes on stock market returns. Although the studies differ in the way the oil price variable is measured, majority of them is in favour of sig-

nificant effect of changes in oil prices on the stock market returns (see for example Chen (2010); Kang *et al.* (2015) for the U.S. evidence or Park & Ratti (2008) and Cunado & de Gracia (2014) who uses European stock market data).

The existing evidence for the stock market return predictability generally relies on in-sample tests. Interestingly, few empirical works implementing out-of-sample tests have obtained negative results. Rapach & Wohar (2006) stress the importance of using tests with good power properties. They test for in-sample and out-of-sample predictive ability of several financial variables with respect to stock returns. These include the dividend-price ratio, price-earnings ratio, book-to-market ratio, dividend-payout ratio, "Fed q", term and default spreads on bonds, short-term interest rate, equity share in total new equity and debt issues, and consumption-wealth ratio. Using annual data on the S&P 500 and CRSP equal-weighted portfolios, they conclude that there is very little discrepancy between the in-sample and out-of-sample results once one uses powerful test statistics.

As far as the issue of robustness of the U.S. stock return predictability is concerned, plots by Pesaran & Timmermann (1995) of the inclusion frequency of the different factors in the forecasting models clearly show the importance of allowing for changes in the underlying process of excess returns in the U.S. stock market that seem to have taken place during the 1960 to 1992 period.

The Table 1.1 on the following page summarizes the variables predicting stock returns used by the studies which have been reviewed earlier in this section. The idea comes from the paper by Cremers (2002). This well-arranged way helps us to find a set of determinants we want to examine in our analysis. Since we are implementing model averaging techniques, we can include as many predictors as we want. Building on the table reported by Cremers (2002), we will display several more studies, mostly from the last decade and a half, and we will also enlarge the set of examined determinants. All variables will be characterized in more detail in Chapter 2.

Table 1.1: Determinants examined in published articles

Authors (Year)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Chen *et al.* (1986)	1	0	0	0	1	1	1	1	0	1	0	1	1	1	0	1	0	0	0
Campbell (1987)	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
Ferson & Harvey (1991)	1	1	0	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0
Whitelaw (1994)	0	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
Pesaran & Timmermann (1995)	0	1	1	0	0	1	1	0	0	1	0	1	0	1	0	0	0	0	0
Kothari & Shanken (1997)	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pontiff & Schall (1998)	0	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0
Bossaerts & Hillion (1999)	1	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0
Sadorsky (1999)	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0
Lettau & Ludvigson (2001)	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Cremers (2002)	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	0	0	0
Avramov (2002)	0	1	1	1	1	1	0	1	1	0	1	1	0	0	0	0	1	1	0
Rapach & Wohar (2006)	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1
Campbell & Yogo (2006)	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Welch & Goyal (2007)	0	1	1	1	1	1	0	1	0	0	0	1	0	0	0	0	0	1	1
Chairakwattana *et al.* (2013)	0	1	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0
Cunado & de Gracia (2014)	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0
Gupta *et al.* (2014)	0	1	1	1	1	1	0	1	0	0	0	1	0	0	0	0	0	1	1
Jordan *et al.* (2014)	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Charles *et al.* (2016)	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
This study	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

We consider following variables:
1. Lagged returns, 2. Dividend yield, 3. Earnings yield, 4. Book-to-market ratio, 5. Credit spread, 6. Yield on a short-term Treasury bill, 7. Change in the yield on a short-term Treasury bill, 8. Term premium, 9. Term spread, 10. Yield spread, 11. January dummy, 12. Inflation Rate, 13. Change in inflation, 14. Change in industrial production, 15. Volume of shares traded over price level, 16. Oil prices, 17. Size premium, 18. Default Risk Premium, 19. Payout ratio

Note: The numbers in the first row of the table refer to the numbering of variables above. Furthermore, a "1" indicates that particular variable was included in the analysis, whereas a "0" means that it was not included. Although several studies examine more determinants than we reported in the table (e.g. growth in consumption), we decided to focus only on the main and most frequent ones.

However, these studies generally use traditional models for stock return predictability and thus ignore the issue of model uncertainty. Several scholars have attempted to implement model averaging techniques to examine determinants of stock returns, for example Cremers (2002), Avramov (2002) and Chairakwattana *et al.* (2013). These will be reviewed in more detail in the following section.

1.2 Model averaging techniques

The extensive econometric environment offers large number of possible models that can examine the effect of our variables of interest on the outcome. The traditional analysis consists in selecting "the best model" among them and this model is then assumed to be the true data generating process. This is known as model selection technique.

However, selecting a single model does not acknowledge for model uncertainty. An alternative approach might be employing a model averaging technique. Model averaging techniques including Bayesian model averaging and frequentist model averaging are considered to be the main statistical tools which incorporate model uncertainty into the analysis. This approach estimates all the competing models and computes a weighted average of the estimates. In the presence of model uncertainty, recent studies have shown that model averaging techniques provide better predictive performances than any single (reasonably) selected model (Raftery *et al.* 1997).

More extensively, we propose a book written by Claeskens *et al.* (2008) as an excellent reference. It presents the statistical approach to the topic of the model selection as well as the model averaging techniques. Several model selection criteria on choosing the best model are compared, the model uncertainty is tackled, and Bayesian and frequentist methods are discussed. Real-data illustrations are also included to provide better insight into the methodology.

In addition, Moral-Benito (2015) in his paper presents an overview of model averaging methods with emphasis on recent developments. He discusses the advantages and drawbacks of both approaches. Two following subsections of this thesis review the relevant literature in the context of both mentioned model averaging techniques.

1.2.1 Bayesian Model Averaging

The Bayesian model averaging technique was initially proposed by Leamer (1978). The approach was basically ignored until the 1990s and 2000s, mostly because of the computational difficulties. After computer and theoretical developments which allow researchers to overcome such troubles, incorporating model uncertainty based on Bayesian approach became increasingly popular in the last two decades (for instance Mitchell & Beauchamp (1988), Madigan & Raftery (1994), Draper (1995), George & McCulloch (1997), Chipman *et al.* (2001), Fernandez *et al.* (2001a), Brown *et al.* (2002), or Clyde & George (2004)). For a survey of the early development of BMA, see Hoeting *et al.* (1999).

In the literature, Bayesian model averaging is commonly applied in economic growth regressions in order to incorporate the issue of the specification uncertainty[3] (e.g. Fernandez *et al.* (2001b), Feldkircher & Zeugner (2012), Moral-Benito (2012) or Eicher *et al.* (2011)). Majority of studies are cross-sectional analyses, except the study by Moral-Benito (2012) who extends the economic growth determinants analysis employing BMA methodology to panel data framework in order to address model uncertainty as well as the endogeneity issues.

Fernandez *et al.* (2001b) include 41 possible growth predictors which leads to a remarkably huge set of potential models. Their empirical results show that the posterior probability is very spread among many models and thus indicate the superior performance of the BMA approach compared to choosing any individual model. Out-of-sample predictive results support this claim. They conclude that the BMA offers the predictive advantage over any single model and therefore they recommend the use of the BMA technique when dealing with growth regression.

An analogous study written by Eicher *et al.* (2011) uses Bayesian framework to identify the determinants of economic growth. Here they attempt to compare 12 parameter priors and two model priors, motivated by the previous studies. They use the same growth dataset as Fernandez *et al.* (2001b) and also simulated data to examine the cross-validated predictive performance of

[3] It has been also used in the meta-regression analysis including a large set of determinants to resolve the issue of model uncertainty (Zigraiova & Havranek 2016).

potential default parameter priors. The general results show that the uniform information prior in combination with uniform model prior work better that other candidate priors. It also identified the largest number of economic growth determinants and thus they recommend to use these as default priors. Contrastingly, Feldkircher & Zeugner (2012) suggest the use of hyper g-prior[4] since it should have several advantages, for instance it should be more robust to outliers.

Regarding the stock return predictability, there exist few studies employing the Bayesian model averaging (Avramov (2002), and Cremers (2002) in the U.S. stock market, or Chairakwattana *et al.* (2013) in the emerging market). Avramov (2002) in his study employs BMA technique in order to incorporate model uncertainty into analyzing the evidence on predictability of stock returns. He shows robust properties of model averaging approach, in particular he finds out that it produces zero mean out-of-sample forecast errors that are serially uncorrelated over time. Furthermore, the results suggest both in-sample and out-of-sample stock return predictability, and that the out-of-sample performance of the Bayesian model averaging is superior to the approach based on model selection criteria (chosen in accordance with Bossaerts & Hillion (1999)). Regarding the predictive power of examined variables, term and market premia seem to be the most important ones as their predictive ability is superior to other predictors. However, the analysis reveals that the forecasting power of economic variables can be considerably weaker when accounting for model uncertainty.

Similar study has been conducted by Cremers (2002). As in the previous case, he implements Bayesian model averaging in order to study the predictability of S&P 500 index excess stock returns. Besides the uncertainty whether there actually is the predictability, he also attributes the uncertainty about the inclusion of particular conditioning variables. He argues that by using model averaging technique he minimizes the data snooping given the initial information set and the priors. He includes 14 different explanatory variables based on the evidence from the previous literature which leads to a total of 2^{14} individual models. The results obtained from the analysis are generally in favor of the use of BMA approach. Specifically, the posterior probability weighted average model outperforms traditional statistical methods. Lastly, the data

[4]It was initially proposed by Liang *et al.* (2008).

imply posterior probabilities that are in general more supportive of stock return predictability than the priors.

Although most studies are based on the U.S. market data, we mention also one study in the context of emerging market where the BMA approach works better as well. Chairakwattana *et al.* (2013) in their paper investigate the predictability power of determinants of stock returns. They include the determinants of stock returns found in the previous studies as well as those prevalent business cycle variables including dividend yield, book-to-market, earning yield, interest rate on Treasury bill, and also inflation rate. Moreover, they apply BMA approach and compare it with other traditional regression model (Fama and Macbeth two steps procedure). They focus on Thailand stock market in the period of years 2000-2011 and extend the research by exploring the forecasting power over financial crisis 2008-2009. Finally, they conclude that BMA approach outperforms the traditional regression model and also that it has out-of-sample predictability power.

However, there still remain several problems associated with using the Bayesian model averaging approach, as pointed out by Hjort & Claeskens (2003). In particular, they mention difficulties with setting up prior probabilities and with the way of dealing with the conflicting prior opinions when mixing together.

1.2.2 Frequentist Model Averaging

In comparison with the BMA technique, there has not been such a vast literature on the use of the FMA approach. Although the intuition is similar, in contrast with BMA, the FMA alternative requires no priors and the corresponding estimators are fully determined by data. For that reason, model averaging from the frequentist perspective has received much more attention over the last decade and a half, and some important progress has been made.

The forecasting combination articles in the 1970s can be considered the predecessors of the current frequentist model averaging literature. The study written by Bates & Granger (1969) has been one of the first papers examining the application of the FMA approach (Wang *et al.* 2009). They combine two sets of forecasts of airline passenger data and conclude that this composite set of forecasts may yield some improvements. Recently, it has expanded to many other

fields. For example, the paper by Wan & Zhang (2009) on tourism research uses WALS estimator [5], mostly because of its advantage of being computationally simple. Moreover, Kapetanios *et al.* (2008) apply frequentist information-theoretic approach to forecast UK inflation, and for this they use large UK dataset. They show that it can be a powerful alternative to Bayesian averaging technique and it may work even better in some cases.

In contrast to BMA, the FMA approach has started to receive attention only over the last decade; see, for example, Hjort & Claeskens (2003) and Hansen (2007). Hjort & Claeskens (2003) proposes various natural FMA strategies including frequentist view on Bayesian approach as well as illustrates "the FMA machinery" for some real data applications. The paper by Hansen (2007) acknowledges the issue of weights selection used for averaging across least squares estimates obtained from a set of models. So far, model averaging techniques have been based on information criteria weights, either AIC or BIC. In this paper, a new model averaging estimator, in particular Mallows model average (MMA) estimator, is developed. Furthermore, it is shown that the MMA estimator, with weights selected by minimizing a Mallows criterion, is asymptotically optimal in the sense that the fitted estimates asymptotically achieve the minimum squared error in a category of model averaging estimators.

Wang *et al.* (2009) provide a well-documented review on the theory and specific methods of frequentist model averaging approach. In particular, they present the definition of the estimators from the frequentist model averaging perspective, summary of the asymptotic theories as well as the weight choice methods in the FMA estimators. Besides, they include also some ideas for further research and development as the FMA technique is fairly new topic.

Amini & Parmeter (2012) in their study attempt to replicate the results of three economic growth studies which use BMA approach and then compare them with the results from alternative model averaging techniques. Specifically, they focus on two methods - Mallows model averaging (MMA) and weighted-least squares (WALS). Finally, they conclude that the sign and magnitude of these new estimates are nearly identical to those obtained using Bayesian model

[5] The WALS or weighted-average least square estimator was introduced by Magnus *et al.* (2010). It is a Bayesian combination of frequentist estimators, and it takes the required prior probabilities from the Laplace distribution and generates bounded risk (Amini & Parmeter 2012)

averaging method.

To conclude, there exists an extensive evidence in the previous literature that stock returns are predictable. Nevertheless, there has been only a little consensus about which determinants should be included in the predictive regression. Besides, the extant literature primarily relies on in-sample tests of stock return predictability. This raises concerns of overfitting, also referes as "data snooping". Several studies have implemented tests for out-of-sample predictability and obtained negative results. This fact may suggest that the in-sample evidence of return predictability is spurious. Therefore, in recent years there is a rising interest in using model averaging techniques to overcome such problems and account for model uncertainty. Bayesian model averaging is quite well-documented in the literature and has also been used to document stock return predictability. On the other hand, FMA is a more recent approach, and is believed to outperform BMA. However, frequentist model averaging technique has not been studied in the context of the predictability of stock returns yet. Therefore, we will contribute to the topic by extending the FMA research into the stock return predictability frame. Moreover, we will compare both model averaging methods with the traditional methods based on model selection criteria using new predictive variables. Accordingly, we will try to account for the considerable uncertainty which is present in the finance literature not only about the correct model for excess stock returns, but also about which predictors to look at initially.

Chapter 2

Data

2.1 Data construction

The dataset used for analysis covers the period from June 1998 to December 2016 and gives a total of 223 observations as all variables are measured at monthly frequencies. The length of the examined time period is chosen simply because of data availability at Thomson Reuters database. However, we believe that the number of observations is sufficient enough to conduct the estimation.

The dependent variable, the monthly excess return on the S&P 500 index, is computed as follows:

$$ExcessReturn_t = \frac{P_t - P_{t-1}}{P_{t-1}} - R_{f,t} = R_t - R_{f,t} \tag{2.1}$$

where P_t is the stock price at time t. In other words, it is the difference between a simple return on S&P 500 index R_t and the risk-free rate $R_{f,t}$. As the risk-free rate we use the three-month Treasury bill rate. We have chosen the Standard & Poor's 500 index for the analysis because it is generally considered as a good representation of the U.S. stock market.

The construction of regressors is explained in Table 2.1. The choice of particular variables is based on the findings from previous literature. We include the predictors that have been previously shown important by majority of studies and also the ones with ambiguous effect on stock returns (see Section 1.1 and Table 1.1 for motivation). The predictors in the dataset include various macroeconomic, technical, interest rate, liquidity and price level variables.

Previous studies examining stock return predictability are not always clear on what they consider to be the appropriate time lags between the explanatory variables and stock returns. Here we follow Cremers (2002) who claim that all predictors are instantly available except the three macroeconomic variables which we include with a one-month lag relative to other predictors to ensure their actual inclusion in the information set. These are, in particular, change in industrial production, inflation rate and its first difference. Moreover, Pesaran & Timmermann (1995) suggest that changes in interest rates rather than their absolute levels affect stock returns, hence we also include a change in a yield on a three-month Treasury bill as a separate variable. We add the change in inflation rate as a measure of unexpected inflation as well.

2.2 Data sources

Most of data are retrieved from Thomson Reuters Eikon database, namely month-end values of Standard & Poor's 500 index close prices, a monthly rate of a 3-month Treasury bill, year-on-year rate of inflation measured as change in producer price index for finished goods, year-on-year change in industrial production, Federal Funds rate, yield on a ten-year Treasury bond, yields on BAA and AAA Moody's corporate bonds, and producer price index for fuels and petroleum products.

Other data sources are as follows: the S&P 500 index volume and the S&P SmallCap 600 close prices are downloaded from Yahoo finance, one-year Treasury rate from Federal Reserve Board[1], 12-month moving sums of dividends and earnings on the S&P 500 index are from Robert Shiller's website[2] and book-to-market ratio for the Dow Jones Industrial Average index from Amit Goyal's updated dataset[3] used in the study "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction". We use the book-to-market ratio for the DJIA index (as in Welch & Goyal (2007)) because of lack of data availability for the S&P 500 index.

[1] Available at: `https://www.federalreserve.gov/datadownload/Choose.aspx?rel=H15`

[2] Monthly data of dividends and earnings of S&P 500 index are linearly interpolated from quarterly data since 1926 and are available online for download at `http://www.econ.yale.edu/~shiller/data.htm`

[3] Available at: `http://www.hec.unil.ch/agoyal/`

Table 2.1: Description of predictive variables

Explanatory variable	*Description*
1. Lagged returns	the S&P 500 index excess returns lagged once and twice
2. Dividend yield	total S&P 500 dividends accruing over the previous 12 months divided by the current level of the index
3. Earnings yield	trailing S&P 500 12-month earnings divided by index price
4. Book-to-market ratio	the ratio of book value to the market value for the Dow Jones Industrial Average
5. Credit spread	the difference between yields on BAA and AAA Moody's rated corporate bonds
6. 3-month T-bill	the yield on a three-month maturity Treasury bill
7. Change in 3-month T-bill	
8. Term premium	the difference between the yield on a 10-year maturity Treasury bond and a 3-month maturity Treasury bill
9. Term spread	the difference in yield of ten-year and one-year Treasuries
10. Yield spread	the difference between the Federal Funds rate and the yield on a 3-month maturity Treasury bill
11. January dummy	takes value '1' for the month of January and '0' otherwise
12. Inflation rate	the monthly, seasonally adjusted year-by-year rate of change of inflation as measured by the producer price index for finished goods
13. Change in inflation	measure of unexpected inflation
14. Change in industrial production	the monthly year-by-year rate of change in industrial production
15. Volume traded over price level	the S&P 500 volume of shares divided by price level (measure of general liquidity of the market)
16. Oil prices	month-by-month change in oil prices as measured by the producer price index for fuels
17. Size premium	the return of firms with small capitalization (S&P SmallCap 600) minus the return of firms with big capitalization (S&P 500)
18. Default risk premium	the difference between return on long-term corporate bonds and return on long-term government bond
19. Payout ratio	12-month dividends divided by 12-month earnings

2.3 Descriptive statistics

Table 2.2 presents descriptive statistics, in particular mean, standard deviation ("StDev"), correlation with dependent variable *excess returns* ("Cor"), first-order autocorrelation $\gamma[1]$ and second-order autocorrelation $\gamma[2]$.

We can observe that most of the series are highly persistent, both in first- and second-order autocorrelation (more than 90%). The exceptions are the lagged returns, the variables in first differences (change in inflation and change in T-bill), January dummy, oil prices and size premium. The correlations with dependent variable in absolute terms differ from higher values of -15.8%, -10.7% and -10.1% for inflation rate, yield spread and dividend yield, respectively, to the lowest value of 2% for earnings yield. The dependent variable, excess returns on S&P 500 index, is on average positive during the sample period, with mean around 0.003 and standard deviation 0.044.

Table 2.2: Descriptive statistics

	Mean	StDev	Cor	$\gamma[1]$	$\gamma[2]$
Excess Returns	0.003	0.044	1.000	0.089	-0.071
ExcessReturns_lag1	0.002	0.044	0.087	0.086	-0.069
ExcessReturns_lag2	0.002	0.044	-0.069	0.088	-0.068
Dividend yield	0.018	0.004	-0.101	0.971	0.932
Earnings yield	0.045	0.014	0.020	0.982	0.952
Book-to-market ratio	0.277	0.074	-0.034	0.954	0.912
Credit spread	0.011	0.004	-0.098	0.955	0.872
3-month T-bill	0.019	0.020	-0.096	0.989	0.974
Change in T-bill	-0.0002	0.002	0.055	0.458	0.259
Term premium	0.019	0.011	0.024	0.973	0.933
Term spread	0.016	0.011	0.047	0.980	0.949
Yield spread	0.002	0.003	-0.107	0.873	0.773
January dummy	0.081	0.273	-0.092	-0.088	-0.089
Inflation rate	0.021	0.031	-0.158	0.939	0.853
Change in inflation	0.0001	0.011	-0.032	0.216	0.051
Industrial production	0.011	0.043	0.055	0.969	0.927
Volume over price	45.482	31.526	-0.091	0.942	0.902
Oil prices	0.004	0.046	0.084	0.232	0.060
Size premium	0.004	0.031	0.057	-0.195	0.099
Payout ratio	0.513	0.547	0.066	0.966	0.892
Default risk premium	0.021	0.006	-0.082	0.957	0.884

The BMA approach assumes low linear dependences among explanatory variables. Otherwise, the choice of explanatory variables across all models might be affected by high correlation coefficients which may lead to offsetting the variables one against another in an individual model and thus causing zero overall effect. Therefore we need to check correlations between each pair of explanatory variables to investigate possible multicollinearity (see Table A.2 in Appendix A). None of the correlation coefficients exceeds 0.85, the only exception is the correlation between two variables constructed in similar nature, particularly *term spread* and *term premium*. The correlation coefficient between these two predictors equals 0.98. However, we include both term variables in the model (as in Avramov 2002) and see the results.

Chapter 3

Methodology

The aim of this chapter is to shed more light on the issue of specification uncertainty and to present methodology which will be used in our analysis. We begin with the brief description of both model averaging techniques that explicitly incorporate uncertainty on model into the estimation procedure. In particular, the Bayesian model averaging is outlined in the first section including the different priors we employ, whereas the frequentist model averaging is explained in the second section. Additionally, the method which is traditionally used in predictability of stock returns and generally ignores the issue of model uncertainty is presented in the third section together with formal definitions of several model selection criteria frequently used in the literature. In the last section, we present several criteria we use for assessing the predictive accuracy of different models.

To conduct the analysis we use statistical software R. The main reason is the availability of BMS package, which relieves employing Bayesian model averaging method. Considering 20 possible predictor variables, BMA compares $2^{20} = 1\ 048\ 576$ models and uses a Markov Chain Monte Carlo (MCMC) algorithm for approximation.

3.1 Bayesian Model Averaging

Regarding Bayesian perspective, Hoeting *et al.* (1999)[1] present an extensive review of BMA, including the implementation as well as its limitations. We

[1]It has been later implemented and discussed in, for example, Amini & Parmeter (2012) or Kapetanios *et al.* (2008)

follow their approach and briefly outline the theory behind it.

To start with, we consider a linear model in the following structure:

$$y = \alpha_k + X_k\beta_k + \epsilon \qquad \epsilon \sim N(0, \sigma^2 I) \tag{3.1}$$

where y is a dependent variable, α_k a constant, β_k are the coefficients, and ϵ stands for an error term which is normally i.i.d and has a variance σ^2.

Furthermore, assume the model space M and denote K as the number of possible explanatory variables. Hence there are 2^K different combinations of regressor variables, thus 2^K potential models indexed by M_k for $k = 1, 2, ...2^K$. In cases where many models are being entertained, it is important to be explicit about which model is under consideration.

Given data D, the posterior distribution for any coefficient of interest β is:

$$Pr(\beta \mid D) = \sum_{k=1}^{2^K} Pr(\beta \mid M_k, D)Pr(M_k \mid D) \tag{3.2}$$

This is a weighted average of the posterior distributions under each of the models considered. BMA uses their posterior model probabilities $Pr(M_k \mid D)$ as weights for averaging and thus PMP is fundamental in the Bayesian framework.

We can derive the posterior model probability (PMP) for an individual model M_k using Bayes theorem, given the prior model probability $P(M_k)$, as:

$$Pr(M_k \mid D) = \frac{P(D \mid M_k)Pr(M_k)}{\sum_{j=1}^{2^K} Pr(D \mid M_j)Pr(M_j)} \tag{3.3}$$

In other words, it is the ratio of its marginal likelihood to the sum of marginal likelihoods over the entire model space where

$$Pr(D \mid M_k) = \int Pr(D \mid \beta^{(k)}, M_k)Pr(\beta^{(k)}, M_k)d\beta^{(k)} \tag{3.4}$$

is the marginal (or integrated) likelihood of model and $\beta^{(k)}$ is the vector of parameters from an individual model M_k. $P(M_k)$ is the prior probability that model M_k is the true model (given that one of the models considered is true).

The posterior means and standard deviations are then constructed as follows:

$$E[\hat{\beta} \mid D] = \sum_{k=1}^{2^K} \hat{\beta} Pr(M_k \mid D) \tag{3.5}$$

$$Var[\hat{\beta} \mid D] = \sum_{k=1}^{2^K} (Var[\hat{\beta} \mid D, M_k] + \hat{\beta}^2) Pr(M_k \mid D) - E[\beta \mid D]^2 \tag{3.6}$$

where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_{2^K})$ and it is equal to $\hat{\beta} = E[\beta \mid D, M_k]$. In other words, it is the estimate of coefficients β from the model M_k.

Standard Bayesian framework also reports the posterior inclusion probability (PIP). It reflects the importance of a predictor and can be interpreted as the probability that a particular variable is included in the "true" model. In our case, regressors with high posterior probabilities of being included are considered as robust determinants of stock returns. For variable h, it is computed as the sum of PMPs of all the models that include the particular variable h:

$$PIP = Pr(\beta_h \neq 0 \mid D) = \sum_{\beta_h \neq 0} Pr(M_k \mid D) \tag{3.7}$$

This is a general textbook derivation of Bayesian framework. However, maximum likelihood and PMP, and thus the posterior mean and variance, depend on priors[2]. The prior framework is described in the following subsection. We also present our choices for parameter and model priors.

3.1.1 Priors

There are two types of priors we need to specify in the Bayesian model averaging approach. The literature emphasizes that priors are essential in determining the posterior probabilities (Ciccone & Jarociński 2010; Feldkircher & Zeugner 2012; Liang *et al.* 2008).

The first type of prior is g, the prior on model parameters. We follow a common approach in the literature which considers a specific prior structure known as "Zellner's prior". It places the non-informative priors on the constant term and

[2]Zeugner (2011) provides the explanation how the choice of g affects the posterior distribution of coefficients and posterior variance.

error variance from Equation 3.1:

$$p(\alpha_k) \propto 1$$

$$p(\sigma) \propto \sigma^{-1}$$

which means they are evenly distributed.

The most important prior is the one on the coefficients β_k. Prior to looking at the data, we formulate our beliefs about the coefficients' mean and variance and we assume that these coefficients follow the normal distribution. Researchers generally assume a prior mean of zero which indicates very little knowledge about coefficients. In addition, they define the variance according to Zellner's g as: $\sigma^2(g(X_k'X_k)^{-1}$. Therefore, the distribution of coefficient β_k conditional on g is as follows:

$$\beta_k \mid g \sim N(0, \sigma^2(g(X_k'X_k)^{-1}) \tag{3.8}$$

The hyperparameter g indicates whether a researcher is very conservative or not, in other words how certain she is that the coefficients are truly zero. The smaller g, the more conservative the researcher is. A small g signifies very little prior variance of coefficients, i.e. she is quite certain that the coefficients are indeed zero. On the contrary, the larger values of g, the more uncertain a researcher is about the coefficients. As g goes to infinity, the estimator of β_k approaches the standard OLS estimator β_k^{OLS}. Therefore the form of the hyperparameter g is very important.

Three most popular parameter prior settings include:

- Unit-information prior (UIP): $g = N$
- Benchmark prior (BRIC): $g = max\{N, K^2\}$
- Hyper-g prior: $\frac{g}{1+g} \sim Beta(1, \frac{a}{2} - 1)$; where the parameter $a \in$ (2,4]

where N is the number of observations and K stands for the total number of explanatory variables.

The first two priors listed above (UIP and BRIC) are so-called 'fixed-g' priors which set the parameter prior for all the models under consideration. Contrastingly, hyper-g prior allows us to update the prior for an individual model in a

Bayesian manner. Selecting $a = 4$ corresponds to the unit-information prior which is a common 'default' choice of parameter prior. On the other hand, setting $a = 2$ concentrates the prior mass close to unity, which corresponds to $g \to \infty$. Liang *et al.* (2008) provides more details on hyper-g prior setting.

The second prior $p(M_i)$ is specified on the model space. According to Moral-Benito (2012), the most popular setting is the binomial distribution. The prior on model M_i can be expressed as

$$p(M_i) = \theta^{k_i}(1-\theta)^{K-k_i} \tag{3.9}$$

where K stands for the total number of explanatory variables, k_i for the subset of regressors chosen, and θ is the probability that specific regressor is included in the model. One of the common settings is to consider $\theta = 1/2$ which results in uniform model prior that assumes equal probability $p(M_i) = 2^{-K}$ for all models in consideration. Uniform model prior tends to put more weight on moderate model sizes. Alternatively, Steel & Ley (2007) propose the use of the so-called beta-binomial model prior which is not as tight around the expected value of the model size. It treats θ as random rather than fixed.

In the thesis, we follow Eicher *et al.* (2011) who suggest UIP together with uniform model priors as a default prior for BMA in linear regression as this combination has been shown to produce good predictive performance. For the first alternative BMA specification we use the BRIC g-prior (following Fernandez *et al.* (2001a)) in combination with a beta-binomial prior that assigns the same probability on each model size. Secondly, we employ a more informative g-prior, the hyper-g prior proposed by Feldkircher & Zeugner (2012) which should be more robust to outliers, and we accompany it with the random model prior.

3.1.2 MCMC sampler

Although the use of Bayesian model averaging appears to be an intuitively attractive method to solve the issue of uncertainty about the correct model in empirical research, there are computational difficulties when conducting the BMA with a large number of potential predictors K. The BMA method uses a sampling algorithm which goes through only part of possible combinations of

variables, referred as model space. It allows to get the results in a reasonable time frame.

The BMA method relies on the Metropolis-Hastings algorithm which is outlined by Zeugner (2011) in the following way: At any step i, a random model M_i (current model) having PMP $p(M_i \mid y, X)$ is drawn. In the following step $i+1$, a candidate model M_j is proposed. The probability that the sampler switches to the new model M_j is as follows:

$$p_{i,j} = \min\left(1, \frac{p(M_j \mid y, X)}{p(M_i \mid y, X)}\right) \tag{3.10}$$

If the model M_j is accepted, it is then considered by the sampler as the current model. If M_j is rejected, another model M_k is drawn and compared with the current model M_i.

We use 'bms' package in R, which was developed by Zeugner and Feldkircher, and allows us to specify the number of initial burn-ins, the number of iterations and one of the following MC^3 sampler settings:

- *Birth-death algorithm:* A randomly chosen explanatory variable is added if it is not already included in the current model M_i or removed if it is already included.
- *Reversible-jump algorithm:* With 50% probability, it behaves as birth-death samples. With 50% probability, it randomly drops one of the variables in M_i and adds one which was previously excluded from M_i.
- *Enumeration:* It walks through all possible covariate combinations.

In this thesis, we use birth-death sampler in all of the BMA specifications.

The correlation between analytical PMP and the PMP from the MCMC sampling algorithm measures the approximation quality. According to Zeugner (2011), the correlation around 0.99 indicates "a good degree of convergence". In case of the level of correlation is smaller, the number of iterations can be increased.

3.2 Frequentist Model Averaging

As a result of recent theoretical advances, frequentist model averaging has become easier to implement and received more attention among researchers. Practical implementation of the FMA requires no priors on model space and parameters, contrary to the Bayesian approach. Hence, the estimated coefficients are determined entirely by data. Contrastingly with the BMA approach which is much more complex, there are no computational difficulties while employing FMA.

There are several possible estimators from the frequentist perspective. We follow Amini & Parmeter (2012) on the use of Mallows model averaging (MMA) estimator for linear regression model which is proposed by Hansen (2007). He shows that MMA estimator is asymptotically optimal in the sense of achieving the lowest possible squared error in a class of discrete model average estimators.

We consider the standard linear regression model structure:

$$y = X\beta + \epsilon = \sum_{i=1}^{\infty} \beta_i x_i + \epsilon \qquad \epsilon \sim i.i.d.D(0, \sigma^2) \tag{3.11}$$

where y is the $n \times 1$ vector of dependent observations and β is the vector of explanatory variables. We denote X_{k_j} the sub-matrix of X including first k_j columns of X, where j indicates the sequence of models and might be of value $j = 1, ..., K$. Hence the l-th model is in the following form:

$$y = X_{k_l}\beta^{(l)} + \epsilon_l \tag{3.12}$$

and is estimated by least squares. Then $\hat{\beta}^{(l)}$ is the vector of estimated coefficients under model l and $\hat{\epsilon}_l$ the vector of estimated residuals. This procedure is repeated for all K models. The final estimator of β is

$$\hat{\beta} = \sum_{j=1}^{K} \omega_j \hat{\beta}^{(j)} \qquad \text{where} \quad j = 1, 2..., K \tag{3.13}$$

In the previous equation, ω_j are normalized weights that are computed by minimizing the Mallows criterion:

$$C_p(\Omega) = \Omega' \hat{\epsilon}' \hat{\epsilon} \Omega + 2\sigma^2 L' \Omega \tag{3.14}$$

with:

- $\Omega_{(K\times 1)} = (\omega_1, \omega_2, ..., \omega_K)'$ representing the vector of weights
- $\hat{\epsilon}_{(n\times K)} = (\hat{\epsilon}_1, \hat{\epsilon}_2, ..., \hat{\epsilon}_K)$ representing the matrix of all residual vectors
- $L_{(K\times 1)} = (k_1, k_2, ..., k_K)'$ representing the vector of the number of regressors used for each of the K models.

We can write the optimization problem mathematically as:

$$\begin{aligned} &\min_{\omega_j} C_p(\Omega) = \Omega'\hat{\epsilon}'\hat{\epsilon}\Omega + 2\sigma^2 L'\Omega \\ &\text{s.t.} \sum_{j=1}^{K} \omega_j = 1, \quad 0 \leq \omega_j \leq 1, \forall j = 1, 2, ..., K \end{aligned} \tag{3.15}$$

It is possible to replace σ^2 which is unknown with the estimator

$$\hat{\sigma}_K^2 = (n-K)^{-1}\hat{\epsilon}_K'\hat{\epsilon}_K \tag{3.16}$$

Buckland *et al.* (1997) propose estimating the variance of a frequentist model averaging estimator as:

$$Var\left(\hat{\beta}\right) = \sum_{j=1}^{K} \omega_j \sqrt{Var\left(\hat{\beta}^{(j)} \mid \text{bias}_j^2\right) + \text{bias}_j^2} \tag{3.17}$$

where $\text{bias}_j = \hat{\beta}^{(j)} - \hat{\beta}$ is the misspecification bias that arises in estimating β under model j. A well-documented review of the application of MMA estimator is presented in Wang *et al.* (2009).

3.3 Traditional model selection

The number of explanatory variables which could have effect on the dependent variable in the traditional linear regression framework is often very large. As the form of true model is unknown, its construction often begins by including all the variables into the model. However, this may yield imprecise results since the higher number of regressors increases standard errors on coefficients and results in less accurate estimation. Empirical research usually approaches this issue by sequentially eliminating least significant variables on the basis of statistical tests to arrive at the single best model with all the irrelevant regressors

omitted.

This section outlines a formal definition of each statistical criterion considered in the thesis for estimating traditional regressions based on model selection criteria. We refer to Kadane & Lazar (2004) for theoretical definition and discussion of appropriate criteria for model selection.

One of the most popular criteria for model selection is R-squared (or R^2). It is defined as one minus the ratio of sum of square errors (abbreviated as SSE) to the total sum of squares (SST):

$$R^2 = 1 - \frac{SSE}{SST} \tag{3.18}$$

However, the issue with using R^2 as a model selection criterion is that SSE, and therefore R-squared itself, increases with the number of explanatory variables included in a particular model. An adjusted version of R^2 as an alternative measure is conventionally used instead to avoid such problem. Adjusted R^2 takes into consideration the number of variables included in the model and is defined (in a relation to simple R^2) as:

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-p} \tag{3.19}$$

where n is number of observations in the sample, and p indicates the number of variables in the model including the intercept. Hence, the use of adjusted R^2 over regular R^2 is advantageous since R^2_{adj} is able to compare models of different sizes.

A closely related criterion (to that of adjusted R-squared) is Mallows criterion C_p:

$$C_p = \frac{SSE_p}{\sigma^2} - (2p - n) \tag{3.20}$$

where SSE_p is the sum of square residuals in the model which includes p variables and σ^2 is the estimated error variance for the full model.

Moreover, there are two other common criteria used for choosing the best subset of regressors, namely the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Both AIC and BIC are based on the maximum

likelihood estimates of the model parameters. For the sake of consistency, we explain both AIC and BIC in terms of SSE as the previous criteria. Accordingly, n stands for the number of observations in the sample, and p indicates the number of variables in the model including the intercept. Hence the formal definitions are as follows:

$$AIC = n\ ln\left(\frac{SSE}{n}\right) + 2p \tag{3.21}$$

$$BIC = n\ ln\left(\frac{SSE}{n}\right) + p\ ln(n) \tag{3.22}$$

They are very similar in nature, the only practical difference is that the Bayesian information criterion penalizes an inclusion of additional variables more heavily which usually results in a sparser model.

The aim of model selection is to choose the simplest model that explains the data sufficiently well and thus to avoid over-fitting problems. To implement model selection analysis based on above-mentioned criteria, we use the function *regsubsets* from package leaps (for subset selection according to adjusted R^2, C_p and BIC) and function *stepAIC* from package MASS in R.

3.4 Predictive performance

To compare the predictive performance of models, we consider three criteria which are conventionally used for assessing the quality of prediction. The first one is Root Mean Square Error (RMSE) that normally takes precedence over the others. It is constructed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2} \tag{3.23}$$

Note that mean squared error (MSE) consists of two components, namely the error variance and the square of the mean error. Hence, by minimizing MSE, we are implicitly minimizing the bias as well as the variance of the errors. Minimizing MSE is equivalent to minimizing RMSE which is the square root of the mean squared error.

Several scholars criticize the use of root mean square error and propose an

alternative criterion instead, in particular the mean absolute error (MAE). It is less sensitive to the large errors than root mean square error which squares the errors in the calculation. Mean absolute error is defined as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|\epsilon_t| \tag{3.24}$$

In addition, following Eicher *et al.* (2011) we also include another scoring rule, in particular log-predictive score (LPS)[3]. It measures the quality of the predictive distribution as a whole. We use LPS to compare the performance of three BMA alternative models as suggested by Fernandez *et al.* (2001b) or Madigan & Raftery (1994).

We define the probability distribution of BMA as

$$pr_{BMA}(y_{new}) = \sum_{k=1}^{K} pr(y_{new} \mid x_{new}, D^{Train}, M_k)\ pr(M_k \mid D^{Train}) \tag{3.25}$$

Then LPS equals

$$LPS = -\sum_{y_{new} \in D^{Test}} log\ pr_{BMA}(y_{new}) \tag{3.26}$$

where D^{Train} and D^{Test} stand for training and testing set, respectively, and y_{new} and predictors x_{new} are from testing set. However, the log-predictive score may be very sensitive to outliers since it penalizes poor probabilistic forecasts more.

All of the three above-mentioned criteria are negatively oriented, i.e. the lower value of a particular scoring rule, the better performance of predictions. Based on these criteria, we want to test following hypotheses:

- Hypothesis #1: A model averaging technique provides superior predictions to these of model selection criteria.
- Hypothesis #2: Frequentist model averaging approach outperforms Bayesian model averaging.
- Hypothesis #3: Within BMA specifications, the hyper-g prior specification performs the best.

[3]The *lps.bma* function is available in BMS package in R.

Chapter 4

Results

This chapter presents the empirical results of our analysis. It consists of several sections. Firstly, we describe results from baseline specification using UIP g-prior combined with uniform model prior. We find which variables get most support and compare our estimates with findings from previous literature, both in terms of sign of coefficients and inclusion in the model. In the second section we provide robustness check with two alternative BMA specifications, in particular BRIC with random model prior and data-dependent hyper-g prior combined with random model prior as well. Within all three BMA alternatives, we decide which one performs the best. Thirdly, we run FMA and compare the results with these obtained from BMA method. Lastly, we estimate traditional model based on model selection criteria, namely adjusted R^2, Mallows criterion, AIC and BIC, and assess the quality of its predictive performance in comparison with model averaging techniques.

4.1 Baseline results

Although various empirical works examining stock return predictability have found several determinants as having the ability to predict excess stock returns, other papers suggest that these findings may not be correct due to overfitting (or so-called data snooping). To address model overfitting, the finance literature have proposed employing Bayesian model averaging (BMA). We examine the ability of 20 possible variables (presented and described in Chapter 2) to predict excess returns on S&P 500 stock index. All estimations are performed in R. Moreover, the intercept is included in all regressions.

Firstly, we follow Eicher *et al.* (2011) using UIP g-prior (in our case g = N = 223) combined with uniform model prior in the baseline specification. The unit information prior contains an equal amount of information as a single observation, while the uniform model prior assigns the same prior probability to each one of individual models, i.e. it favours intermediate model sizes. We conduct the analysis using 'bms' package in R developed by Martin Feldkircher and Stefan Zeugner. For approximation, birth-death MC^3 sampler (described in Subsection 3.1.2) is used. It is the most popular form of the MCMC sampling algorithm used under BMA framework. Moreover, we specify 1 000 000 burn-in draws and 3 000 000 subsequent iterations that guarantee sufficient convergence of the MCMC sampling algorithm.

We begin with the description of basic model diagnostics that are depicted in the Table 4.1. Besides several facts that we already know, such as the number of draws and burn-ins or the number of observations, it includes also some important additional information that offers deeper insight into the model. For instance, mean number of regressors is 10.1986. This suggests that the true model would most likely include ten variables. Furthermore, "% Topmodels" indicates the sum of posterior model probabilities of best 5000 models and equals 100. "% visited" demonstrates a share of evaluated models with respect to the total model space and the percentage equals 88. The quality of approximation is assessed by the correlation between the best models' MCMC frequencies and their marginal likelihoods ("Corr PMP") which equals 0.9998 for the best 5000 models. This number is sufficiently high and indicates very good degree of convergence (Zeugner 2011). It may be partly explained by the fact that we have specified the number of iterations three times as high as the model space.

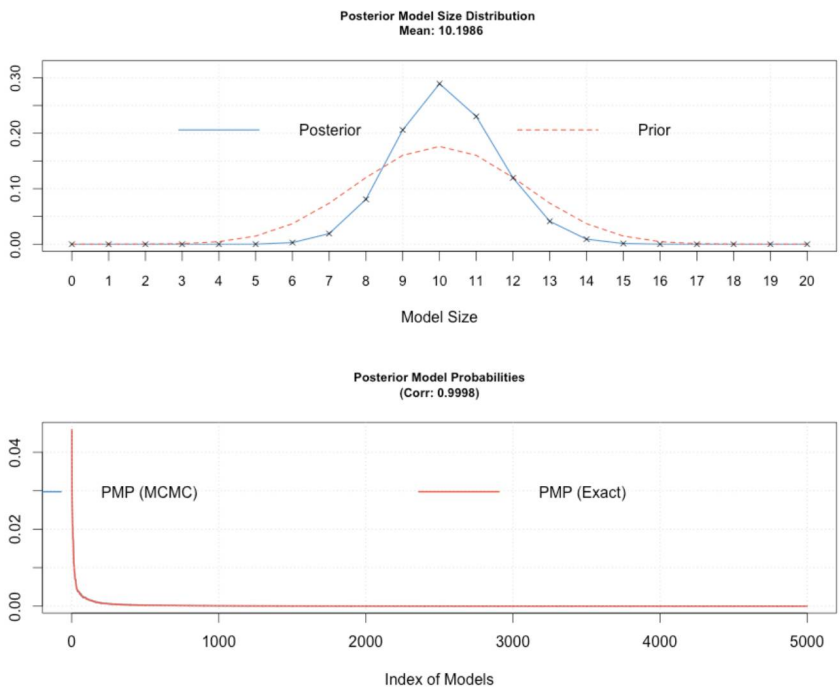
Table 4.1: Model diagnostics, UIP + uniform model prior

Mean no. regressors	**Draws**	**Burnins**	**Time**
10.1986	3e+06	1e+06	14.75298 mins
No. models visited	**Model space 2^K**	**% visited**	**% Topmodels**
917629	1e+06	88	100
Corr PMP	**No. Obs.**	**Model Prior**	**g-Prior**
0.9998	223	uniform / 10	UIP
Shrinkage-Stats			
Av=0.9955			

The Figure 4.1 combines two plots. In particular, the upper plot compares the prior model size (red line) with the posterior distribution of model size (blue line). As already mentioned, for the prior on model space we use the uniform model prior which implies a symmetric distribution around one half of the total number of the predictors (in our case K/2 = 10). The posterior model size equals 10.1986, therefore there is no much difference after updating it with the data.

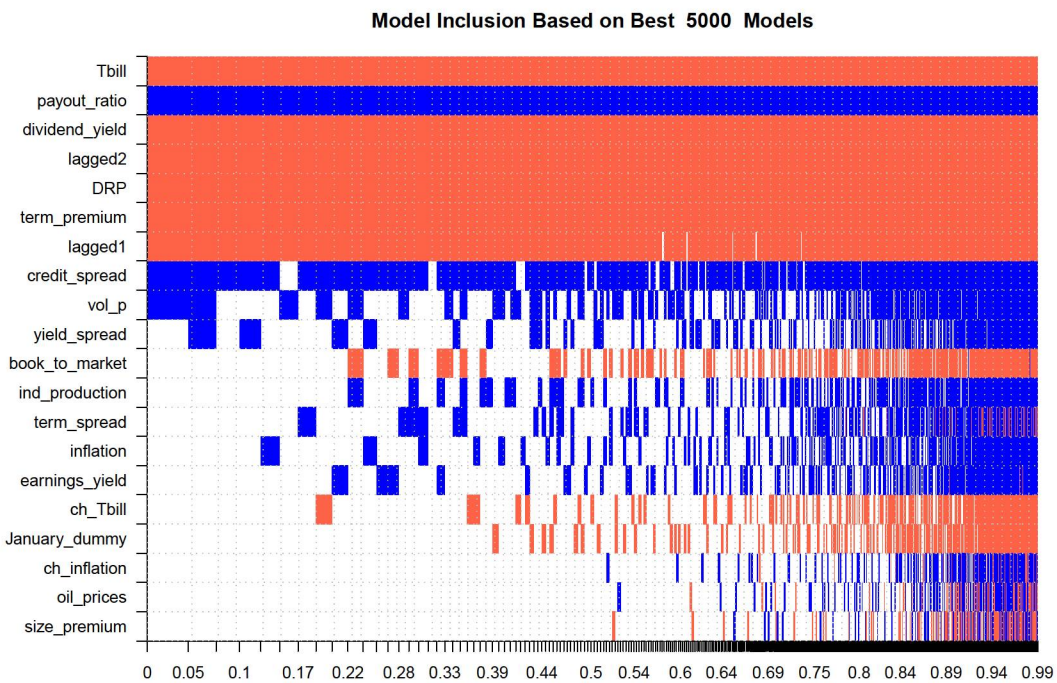
Additionally, the lower plot in Figure 4.1 indicates how well the MCMC sampler has performed. It provides a better visualization of the convergence between analytical and MCMC posterior model probabilities. Since the convergence coefficient equals 0.9998, the blue line representing PMP from the MCMC sampler cannot be indeed observed. It is based on 5000 best models, therefore we provide more detailed version based on 30 best models in Figure A.1 in Appendix. There we observe that the shape of either of curves is not as smooth as it seems in the lower plot of Figure 4.1. The curve is actually refracted as an individual model M_k is drawn. The reason might be the complicated marginal likelihoods of a single model M_k.

Figure 4.1: Model size and convergence, UIP + uniform model prior



Furthermore, Figure 4.2 depicts model inclusion in Bayesian model averaging based on best 5000 models and gives a more comprehensive overview over the models. The dependent variable is monthly return on S&P 500 index in excess of three-month Treasury bill rate. The columns denote individual models and the variables on the vertical axis are sorted by their posterior inclusion probability in descending order. Blue colour indicates that the variable is included in the particular model and the estimated sign is positive, while red colour corresponds to inclusion as well, but the estimated sign is negative. In addition, the blank space means that the variable is not included in the model. The width of a column represents the posterior model probability of an individual model M_k. It can be observed that most models have generally very small posterior model probabilities, thus only a few of them are considered as highly informative. The horizontal axis shows cumulative posterior model probabilities.

Figure 4.2: Model inclusion in BMA, UIP + uniform model prior



Here we see that the best model includes *Tbill*, *payout_ratio*, *dividend_yield*, *lagged*2, *DRP*, *term_premium*, *lagged*1, *credit_spread*, and *vol_p*[1]. More-

[1]We believe that the shortened names of predictors, used for the sake of simplicity in the R code, are self-explanatory and easy to understand. However, we explain them in Table A.1 in Appendix since they are used throughout the whole Chapter 4.

over, the figure shows that these variables are included in nearly all model mass unanimously with positive (*payout_ratio*, *credit_spread*, *vol_p*) or negative coefficient (*Tbill*, *dividend_yield*, *lagged*2, *DRP*, *term_premium*, and *lagged*1). Contrastingly, for instance *oil_prices* or *size_premium* are included very little, and their estimated signs change in accordance with particular model.

In Table 4.2 we focus on the models which perform the best. For the illustration purposes, only three models with highest PMP values are reported. The model with the highest posterior model probability accounts for only 4.5% of the total posterior probability and would include following variables: *Tbill*, *payout_ratio*, *dividend_yield*, *lagged*2, *DRP*, *term_premium*, *lagged*1, *vol_p* and *credit_spread* (as we have already observed from the first column in Figure 4.2). To put it differently, those nine predictors would be able to explain only 4.5% of excess returns in the model. However, such model cannot explain remaining 95.5% of model mass. BMA results thus indicate considerable model uncertainty. Hence, although it is considered to be the best model, it would be misleading to rely on the results from such individual model. In the similar fashion the second and third best models can be interpreted, with PMPs 3.1% and 2.6% correspondingly. Altogether, the top 10 models account for 24% which can be checked on the x-axis in Figure 4.2.

Finally, we proceed with the numerical results of the first BMA specification which are presented on the left-hand side of Table 4.3. The table includes posterior mean (Post Mean), posterior standard deviation (Post SD) and posterior inclusion probability (PIP). An estimated coefficient is the coefficient averaged over all models and takes posterior model probabilities as weights. It also includes the models where a certain regressor is not contained, i.e. the coefficient is zero. Posterior inclusion probability reports how likely the variable is to be included in the true model (analogous to statistical significance). It is the crucial statistics in Bayesian model averaging framework. Here we follow the classification proposed by Kass & Raftery (1995) and Havranek *et al.* (2017) according to which the significance of particular predictor is considered *weak* if its PIP lies between 0.5 and 0.75. Furthermore, PIP between 0.75 and 0.95 implies *positive* significance, while values of PIP lying between 0.95-0.99 and 0.99-1 imply *strong* and *decisive* significance, correspondingly.

Table 4.2: Top 3 models inclusion, UIP + uniform model prior

	#1	#2	#3
dividend_yield	1	1	1
earnings_yield	0	0	0
book_to_market	0	0	0
credit_spread	1	1	1
Tbill	1	1	1
ch_Tbill	0	0	0
term_premium	1	1	1
term_spread	0	0	0
yield_spread	0	1	0
January_dummy	0	0	0
inflation	0	0	0
ch_inflation	0	0	0
ind_production	0	0	0
vol_p	1	1	0
oil_prices	0	0	0
size_premium	0	0	0
payout_ratio	1	1	1
lagged1	1	1	1
lagged2	1	1	1
DRP	1	1	1
PMP (Exact)	0.046	0.031	0.026
PMP (MCMC)	0.045	0.031	0.026

The variables with PIP higher than 0.99 (i.e. with decisive significance) are following: *Tbill*, *dividend_yield*, *payout_ratio*, and *lagged*2. *Default risk premium (DRP)* and *term_premium* are considered strongly significant while *lagged*1 shows positive significance. Another variable with at least weak significance (PIP above 0.5) is *credit_spread*. The rest of the variables have lower posterior inclusion probabilities thus we do not consider their effect as specifically important in determining excess returns. The constant term yields posterior inclusion probability equal to 1. Note that the sum of PIPs is equal to the posterior model size of 10.1986 (see a posterior statistic "Mean no. regressors" in Table 4.1).

Additionally, we run a simple OLS as a frequentist check. The results including estimate, standard error and p-value are reported on the right-hand side of Table 4.3. We observe fairly similar results, both in magnitude and sign of coefficients. Regarding the sign of estimates, the only difference is *yield_spread*, which is positive under BMA estimation and negative in the OLS regression. However, it is shown to be insignificant. Besides, we observe that nearly all of the coefficients are larger in absolute values in the frequentist check. It might

be explained by the fact that choosing a conservative zero for each parameter generally drives the posterior means for coefficients in Bayesian model averaging towards zero. Those variables with high PIP under BMA framework have p-value lower than 5% which indicates strong significance. The only exceptions are variables that are found to be significant at 10% level in the OLS regression and have lower PIP in BMA, particularly *book_to_market* (PIP = 0.197), *January_dummy* (PIP = 0.115) and *ind_production* (PIP = 0.198). It is consistent with findings of Avramov (2002) who shows that many variables which appear to be significant in a simple OLS, lose their significance once model uncertainty is considered. The R-squared value for this regression is 0.3968 and the constant term shows statistical significance as well. Looking at the results, we can confirm in-sample stock return predictability since many of variables have p-value lower than 0.1 which indicates significance at 10% level (Avramov 2002).

Table 4.3: Results from BMA, baseline specification

	Bayesian Model Averaging			Frequentist check (OLS)		
	Post Mean	Post SD	PIP	Estimate	St. Error	P-value
dividend_yield	-13.716	2.956	1.000	-13.098	2.708	0.000
earnings_yield	0.126	0.339	0.173	0.149	0.565	0.792
book_to_market	-0.038	0.093	0.197	-0.278	0.111	0.013
credit_spread	3.170	2.374	0.720	5.935	1.590	0.000
Tbill	-3.549	0.654	1.000	-4.358	0.564	0.000
ch_Tbill	-0.319	1.006	0.135	-1.698	1.918	0.377
term_premium	-3.517	1.445	0.980	-6.275	1.774	0.001
term_spread	0.477	1.420	0.199	2.872	1.837	0.120
yield_spread	0.548	1.226	0.217	-0.133	1.733	0.939
January_dummy	-0.001	0.005	0.115	-0.016	0.009	0.080
inflation	0.032	0.085	0.174	0.094	0.130	0.471
ch_inflation	0.005	0.063	0.053	0.099	0.261	0.705
ind_production	0.048	0.118	0.196	0.323	0.142	0.024
vol_p	0.000	0.000	0.369	0.000	0.000	0.173
oil_prices	0.000	0.013	0.049	0.021	0.061	0.726
size_premium	-0.000	0.018	0.049	0.003	0.081	0.974
payout_ratio	0.066	0.011	1.000	0.071	0.012	0.000
lagged1	-0.191	0.093	0.882	-0.244	0.065	0.000
lagged2	-0.271	0.068	0.992	-0.300	0.063	0.000
DRP	-4.667	1.511	0.986	-6.955	1.155	0.000
(Intercept)	0.410		1.000	0.501	0.069	0.000

We also run simple OLS only with the variables having PIP higher than 0.5, i.e. those variables that are likely to be effective in determining excess stock returns (having at least weak effect). In this case, all variables are found to be

significant at least at 5% significance level and R^2 equals 0.3242 (see Table A.5 in Appendix for more detail).

To our best knowledge, there are only two dominant studies examining stock return predictability using BMA approach and thus we refer to those while comparing our results to the previous literature. Avramov (2002) attempts to predict the returns of the six size/book-to-market portfolios while Cremers (2002) examines excess returns on the S&P 500 index. Although the same model averaging technique is applied and there are also similarities regarding the dependent variable, surprisingly, controversial results are obtained. Table A.6 provides the brief summary of these two studies. For example, the variable inflation seems to be one of the most robust predictors in Avramov (2002) and he confirms significant stock return predictability, while inflation appears to be one of the weakest predictors in Cremers (2002) and his paper only provides marginal evidence for return predictability. These inconsistent results demand further investigation of BMA.

Interestingly, previous empirical works have examined mostly stock return predictability itself and have not paid much attention to direction of the parameters' effect. Therefore, it is quite difficult to compare our results to these from the previous studies[2]. Besides, researchers consider different time periods and different set of possible predictors, and they also use different techniques to explain the predictability of excess stock returns. Following paragraphs attempt to explain the importance of variables and the direction of their effect.

The most important variable seems to be *dividend yield* because of its comparatively large posterior mean -13.716. In contrast to our findings, Avramov (2002) and Cremers (2002) both found dividend yield to be one of the weakest predictors. Even though the theory suggests positive effect of dividend yield on stock returns, our posterior mean yields negative value. Avramov (2002) reports negative slope coefficient as well, but it is insignificant.[3]

[2]Neither Avramov (2002) nor Cremers (2002) discuss the direction of the effect. Avramov (2002) provides at least tables with slope coefficients. Hence, we also refer to other evidence in the financial literature.

[3]It is important to note, that the nature of our analysis is a bit different. Most studies use explanatory variables with a one-month time lag. For that reason, we re-run the OLS regression with all variables taking one lag. Then the dividend yield's estimated coefficient is positive and statistically significant which is in line with economic theory and intuition. The results are reported in Table A.12.

Another explanatory variable with substantial effect on stock returns is *three-month Treasury bill rate*. Its posterior inclusion probability equals 1 and posterior mean is -3.549. This predictor is included in the regression explaining excess stock returns most often. In fact, both Avramov (2002) and Cremers (2002) find 3-month Treasury bill rate as one of most robust predictors in their analyses. Its negative effect on U.S. excess returns is in line with economic theory which suggests that an increase in interest rate leads to a decrease in stock prices. Numerically, the posterior mean indicates that each additional percentage point increase in three-month Treasury bill rate leads to decrease of excess return on S&P 500 index by 3.55 percentage points, ceteris paribus. However, the posterior inclusion probability of change in Treasury bill is not very large (0.135). Therefore we conclude that changes in short-term interest rate do not seem to carry any additional information than the interest rate level itself.

The variable *payout ratio* has PIP equal to 1 and positive posterior mean 0.066. The positive effect of payout ratio on excess returns confirms our expectations. The finding also verifies results of Lamont (1998) examining the post-war U.S. market. His work suggests that higher payout ratio implies higher expected returns and that the usefulness of payout ratio rests in its ability to capture the temporary components of earnings. Furthermore, we have included *lagged excess returns* into the regression following the work of Cremers (2002). Both lagged returns negatively affect excess stock returns on S&P 500 index. Although Cremers (2002) considers both lagged returns as weakest predictors, we, on the other hand, observe decisive and positive significance for *lagged2* (PIP = 0.992) and *lagged1* (PIP = 0.882), respectively. It confirms findings of Bossaerts & Hillion (1999) who employ model selection criteria in their analysis in order to tackle the overfitting problem and find that lagged returns are included in virtually all models based on PIC (posterior inclusion criterion) in the U.S. (and in other examined countries as well).

Another variable *term premium* has been shown important in Avramov (2002) and in Cremers (2002) for confident investor as well. Besides, it has been found as significant determinant for stock returns in the previous literature (e.g. Fama & French (1993)). Its PIP equals 0.980 indicating strong significance and the effect seems to be negative, with posterior mean -3.517. As we mentioned in Chapter 2, the correlation between the variables *term premium*

and *term spread* is very high (0.98) due to the fact they are constructed in a very similar nature. Therefore, we run an estimation with *term spread* and *term premium* separately and we present the results in Table A.7 in Appendix. It does seem that both of these variables are significant separately (PIP =1), and that the sign and magnitude is also very similar. However, when both predictors are included in the regression, only *term premium* ends up being significant. Avramov (2002) reports the same results. Second predictor with PIP higher than 0.95 is *default risk premium*, more specifically it yields PIP of 0.986. The effect seems to be quite large in magnitude, with posterior mean -4.667. Such result is in line with the evidence presented in Rapach *et al.* (2016) who show that default risk premium has the largest estimate among the predictors which display predictive ability. Although economic theory suggests that default risk premium should have a positive impact on stock prices, our results indicate a negative sign of regression coefficients.

The last variable that is shown to have at least weak effect on U.S. excess returns (PIP higher than 0.5) is *credit spread*. In other words, it is the difference between BAA and AAA Moody's rated corporate bond yields. The reported posterior inclusion probability is 0.72, while the posterior mean is 3.17 which indicates the positive effect on stock returns as we have expected. Other variables have posterior inclusion probabilities lower than 0.5 and also their estimated coefficients are quite small which means that the results often include models where these coefficients are zero. Therefore, their effect on the excess stock returns does not appear to be of particular importance. Three weakest predictors according to their PIP are: *change in inflation*, *oil prices* and *size premium*. It is believed that changes in oil prices increase the uncertainty in the capital market and thus influence stock market returns. However, regression coefficient of the variable *oil_prices* has very small value (around zero) and its PIP is only 0.049, which means that oil prices do not seem to have a significant influence on the stock returns. On the other hand, insignificance of *January dummy* is not very surprising since the January effect is not that pronounced for large stock (Bossaerts & Hillion 1999).

Moreover, as the first robustness check we conduct the same BMA analysis using reverse-jump sampler. Switching to reverse-jump sampler has only slightly altered PIPs, and the only noteworthy difference is that PIP of the variable *lagged1* has increased to 0.955 and thus it has become strongly significant.

Mean number of regressors is 10.1856 which is almost the same, and the correlation between analytical PMP and the PMP from the MCMC sampler equals 0.9997 which indicates very good degree of convergence. Percentage of models visited is only 65%, in contrast to 88% in the estimation using birth-death sampler. The results are presented in Table A.4 in Appendix. To sum it up, an alternative reverse-jump sampler does not significantly change the results.

To conclude, our baseline analysis reveals some degree of model uncertainty, therefore the appropriate methods should be used while examining determinants of excess stock returns. Besides, we agree with general conclusion reported in Lettau & Ludvigson (2001) that returns are predictable. Among the most important predictors of S&P 500 stock returns are: three-month Treasury bill rate, payout ratio, dividend yield, default risk premium, lagged returns, term premium and credit spread. Using the most recent data, we conclude that determinants of stock returns are not consistent in time and need to be revise regularly.

4.2 Alternative BMA specifications

However, researchers suggest that BMA results may be sensitive to the data revisions under certain prior structures (Ciccone & Jarociński (2010) or Eicher *et al.* (2011)). In order to check robustness of the results in baseline specification, we perform also the estimation using alternative priors. First one is BRIC parameter prior together with random model prior and second alternative is data-dependent hyper-g prior combined with random model prior. For g, BRIC uses $g = max\{n, K^2\}$ (see Subsection 3.1.1). Regarding the dimension of our dataset, we have 223 observations and 20 possible predictors ($K^2 = 400$). Since 400 is bigger than 223, it is not the same as UIP (where $g = N$) and it does make sense to employ BRIC parameter prior as well. Besides, we employ hyper-g prior which is considered by economists as superior.

For the sake of simplicity, since both BMA alternatives employ random model prior, they will be hereinafter referred to as "BRIC" specification and "hyper-g" specification. Random model prior, or so-called beta-binomial model prior, gives each model size equal prior probability and was proposed by Steel & Ley (2007).

To start with, we present the general differences while looking at the model diagnostics in the second and third column in Table A.8 in Appendix for the BRIC and hyper-g specification correspondingly. We have specified the same number of iterations and burn-ins as in the case of baseline specification, 3 000 000 and 1 000 000 respectively, and also the birth-death sampler has been used. The number of models visited is the highest for the hyper-g specification and equals 1 058 791, in comparison to 844 266 for BRIC alternative and 917 629 for UIP with uniform model prior. Mean number of regressors attains the value of 9.4869 for BRIC estimation which is a little lower than in the baseline specification. On the other hand, mean number of regressors in hyper-g estimation is 17.2239 which is significantly higher comparing to both baseline or BRIC specification. Posterior model size distributions are presented in the upper plots of Figures A.2 and A.3 in Appendix. It is not surprising that posterior model size distribution is slightly shifted towards smaller number in the specification employing BRIC prior, whereas it is shifted positively (i.e. towards larger number of explanatory variables) in the hyper-g model. Furthermore, "Corr PMP" attains 0.9998 for the BRIC alternative and 1 for the hyper-g alternative, which signifies a very good convergence in both cases. The quality of approximation is depicted in the lower plot of Figures A.2 and A.3 in Appendix. Again, since the correlation is high enough, blue line representing the PMP from the MCMC sampler is not really observable.

However, the most striking difference can be seen in Figures 4.3 and 4.4 that represent model inclusion based on best 5000 models. The best models are shown on the left and the most useful variables at the top of the figures. Whereas the upper figure (Figure 4.3) looks similarly to that of baseline BMA specification, the BMA alternative employing hyper-g prior would include considerably more variables into the best model (Figure 4.4). We display top three models from both BMA alternatives in the Table A.10. Regarding the BRIC specification, the best model would include *dividend_yield*, *credit_spread*, *Tbill*, *term_premium*, *payout_ratio*, *lagged*1, *lagged*2, and *DRP* with posterior model probability 5.2%. To that, the second best model would add one more variable, namely vol_p and its PMP equals 5.1%. The sum of ten best models equals 32.26% which is more than the specification implementing the "default" UIP g-prior (only 24.11%). On the other hand, hyper-g prior alternative would include all of the explanatory variables in the best model with posterior model probability roughly 17%. The second best model would ex-

clude only *size_premium* and the PMP attains the value of 2.3%. Similarly, the third best model would exclude only the variable *yield_spread* and its PMP is 2.3% as well. Sum of PMPs of ten best models totals in 33.04%.

Figure 4.3: Model inclusion in BMA, BRIC + random model prior

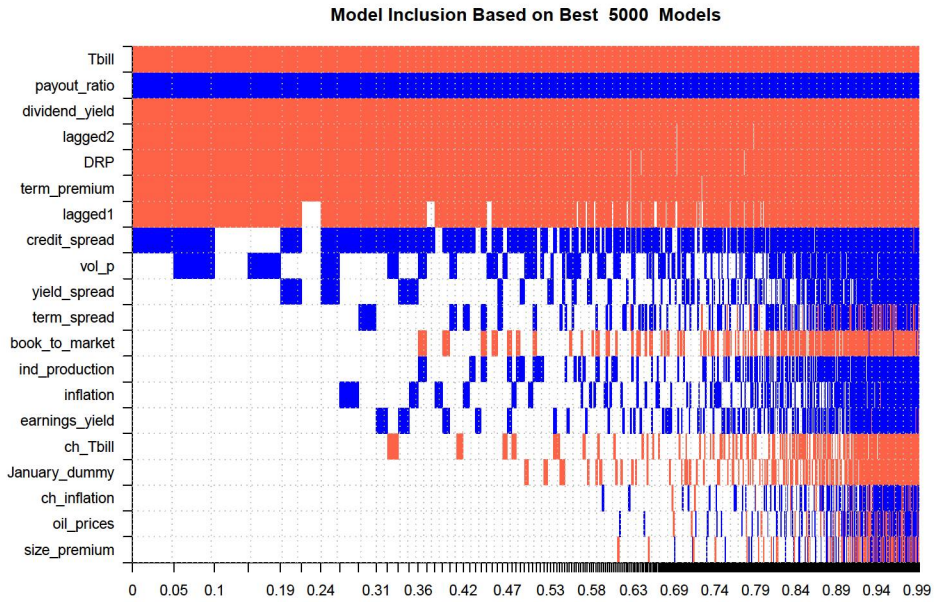
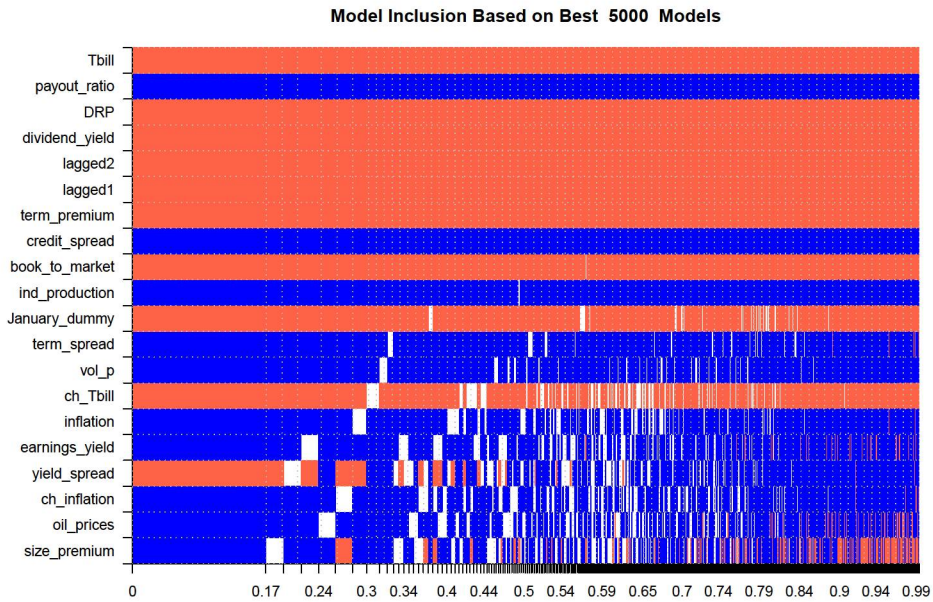


Figure 4.4: Model inclusion in BMA, hyper-g + random model prior



In Table 4.4, we report numerical results including posterior mean, standard deviation and posterior inclusion probability for both BMA alternative specifications. Results from BRIC estimation are on the left, while those from hyper-g alternative are reported on the right-hand side of the table. Although there are not considerable differences in the coefficient signs or magnitudes of estimates, posterior inclusion probabilities differ substantially. In the BMA alternative implementing the benchmark prior, the variables showing decisive significance (PIP higher than 0.99) are following: *Tbill*, *dividend_yield*, *payout_ratio*, and *lagged*2. *DRP* and *term_premium* are considered strongly significant while the predictor *lagged*1 yields positive significance. The last variable with at least weak significance (PIP above 0.5) is *credit_spread* and all remaining regressors have very low posterior inclusion probabilities. These results are exactly the same as we have observed in the specification with UIP and uniform model prior. Contrastingly, regarding the hyper-g specification, following predictors have obtained PIP of more than 0.95: *dividend_yield*, *credit_spread*, *Tbill*, *term_premium*, *payout_ratio*, *lagged*1, *lagged*2 and *DRP*. Moreover, all of the other variables show PIP higher than 0.5 (even more than 0.6) so all of the twenty regressors have at least weak effect on excess stock returns. The posterior model probabilities are very high, on average, which results in posterior mean of more than 17 variables in the model. The constant term shows statistical significance in both BMA alternatives.

To examine the signs of variables in more detail, we present Table A.9 in Appendix which shows sign certainty of 20 potential predictor variables across all of three BMA specifications. In other words, it displays probability that the coefficient is positive conditional on inclusion (1 being 100% probability). The sign of each variable is more or less robust across models, the only exception is *size premium* whose estimated sign is not certain. Two specifications, particularly those employing the UIP g-prior and the benchmark prior, show only around 40% probability that the sign is positive, whereas the number is higher than 80% in the hyper-g model. Moreover, the table confirms the negative effect of dividend yield since the conditional posterior sign is positive in 0% of the cases in all of three BMA alternatives. To conclude, the regression signs of the variables are stable regardless of the prior structure in the BMA specification.

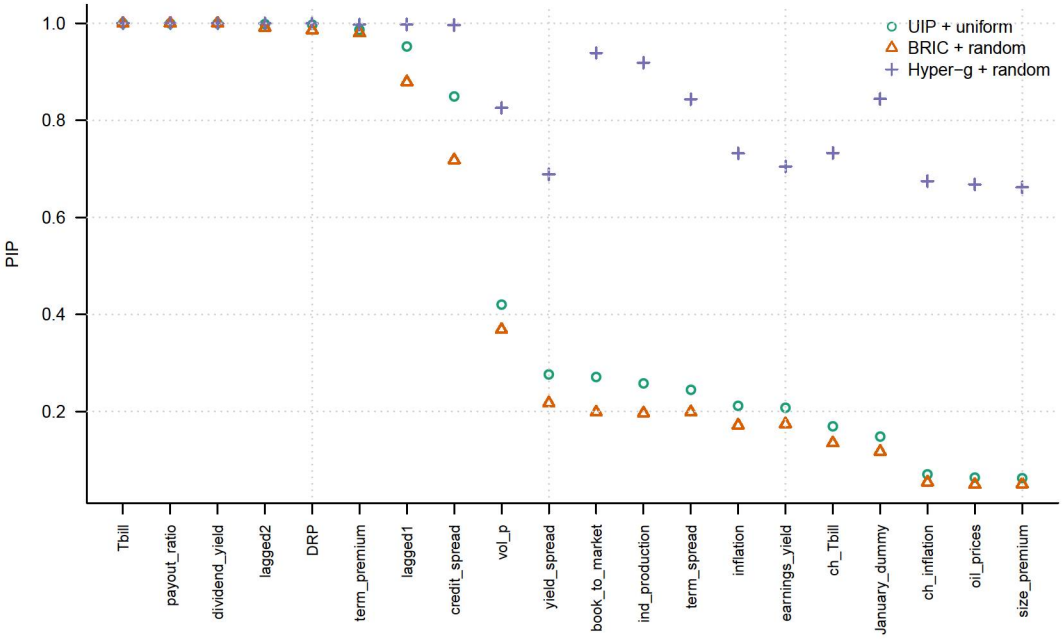
Table 4.4: Results from BMA, alternative specifications

	BRIC + random			hyper-g + random		
	Post Mean	Post SD	PIP	Post Mean	Post SD	PIP
dividend_yield	-13.716	2.956	1.000	-11.506	2.620	1.000
earnings_yield	0.126	0.339	0.173	0.173	0.462	0.709
book_to_market	-0.038	0.093	0.197	-0.214	0.114	0.940
credit_spread	3.170	2.374	0.720	4.937	1.511	0.998
Tbill	-3.549	0.654	1.000	-3.698	0.514	1.000
ch_Tbill	-0.319	1.006	0.135	-1.066	1.611	0.734
term_premium	-3.517	1.445	0.980	-5.006	1.700	0.999
term_spread	0.477	1.420	0.199	2.125	1.811	0.845
yield_spread	0.548	1.226	0.217	0.228	1.376	0.691
January_dummy	-0.001	0.005	0.115	-0.011	0.009	0.846
inflation	0.032	0.085	0.174	0.067	0.109	0.731
ch_inflation	0.005	0.063	0.053	0.063	0.203	0.678
ind_production	0.048	0.118	0.196	0.248	0.147	0.921
vol_p	0.000	0.000	0.369	0.000	0.000	0.831
oil_prices	0.000	0.013	0.049	0.010	0.046	0.669
size_premium	-0.000	0.018	0.049	0.002	0.061	0.663
payout_ratio	0.066	0.011	1.000	0.062	0.011	1.000
lagged1	-0.191	0.093	0.882	-0.207	0.061	0.999
lagged2	-0.271	0.068	0.992	-0.256	0.059	1.000
DRP	-4.667	1.511	0.986	-5.777	1.086	1.000
(Intercept)	0.410		1.000	0.422		1.000

To visually summarize the results of all BMA specifications, we present model comparison in Figure 4.5. The first model shows the PIPs under baseline specification (uniform information prior together with uniform model prior). The second model employs the benchmark prior (BRIC) combined with random model prior and the third one portrays results when we use hyper-g with random model prior. We see that "the UIP + uniform" model and "the BRIC + random" model illustrate the same order of variables according to their PIPs. Besides, two major regularities are observed. Firstly, switching to hyper-g prior leads to significantly higher posterior inclusion probabilities and model size than in the baseline case, and secondly BRIC prior alternative yields slightly lower PIPs. However, there are also several differences in ordering. The most outstanding distinction is *January dummy* which appears to be more important in hyper-g estimation. Specifically, its PIP has risen from 0.11 to 0.85. On the other hand, *vol_p* and *yield_spread* seem to be less important in hyper-g specification than in those employing fixed-g priors. Thus these findings reveal sensitivity to the prior selection which confirms results of Eicher *et al.* (2011). Nevertheless, the most important predictors according to their posterior inclusion probabilities remain practically the same, namely Treasury bill rate,

payout ratio, dividend yield, two-month lagged excess returns, default risk premium, and term premium. All of these variables have PIP higher than 0.95 and thus are strongly significant. Three least important variables are identical across models as well, in particular change in inflation, oil prices and size premium.

Figure 4.5: Model comparison of different BMA alternatives



The benchmark prior (BRIC) is prior suggested by Fernandez et al. (2001) who argue that higher g minimizes prior impact on the results. We confirm the evidence of Feldkircher & Zeugner (2012) that the benchmark prior favours parsimonious models (which is one of the expected consequences due to value of g). On the other hand, they explain the problem with BRIC prior and argue that hyper-g should be preferred because its predictive performance is superior to the other priors. Regarding the accuracy of predictions, we compare all three BMA alternatives in Section 4.5 using popular criteria as root mean square error and mean absolute error. Besides, we also compute the logarithmic predictive score (LPS), which has been used in the BMA context in Madigan & Raftery (1994), Fernandez *et al.* (2001b), Steel & Ley (2007) or Eicher *et al.* (2011).

4.3 Frequentist Model Averaging

To our best knowledge, frequentist model averaging has not been used in the literature to examine the determinants of excess stock returns yet. Here we use Mallows' model averaging estimator (Hansen 2007) and follow the approach by Amini & Parmeter (2012) who suggest orthogonalization of covariate space to narrow the number of estimated models. In effect, by virtue of orthogonalization we are able to select a single variable and continue to add one variable linearly at a time to the model space. Then we orthogonalize the vector space constructed by those variables, and compute the weights by minimizing the Mallows' criterion provided in Section 3.2. The advantage over BMA is that a frequentist method is generally less difficult to implement.

The results obtained from frequentist model averaging method are presented in Table 4.5, including estimated coefficient, standard error and p-value for all 20 regressors and the intercept term. Eight out of twenty variables yield p-value lower than 0.01. These include, in particular, dividend yield, credit spread, three-month Treasury bill rate (*Tbill*), term premium, payout ratio, excess returns lagged once and twice (*lagged1* and *lagged2*) and default risk premium (*DRP*), hence these are significant at 1% level. Highest magnitude in absolute terms of a regressor's effect is ascribed to dividend yield, followed by term premium, credit spread, and 3-month T-bill rate. Moreover, other two predictors are significant at 5% level. These are book-to-market ratio and industrial production. The coefficient sign of the variable *industrial production*, as a measure of economic activity, is positive, in accordance with our expectations. The potential explanation is that an increase in industrial production leads to an increase in economic activity and thus to increase in earnings for companies. Therefore higher earnings recognized by investors should result in an increase in stock valuations resulting in higher returns. Besides, *January dummy* has p-value 0.08 which indicates its significance at 10% level. However, negative January effect contradicts the economic theory which suggests a seasonal anomaly that stock returns are above normal in January. We obtain opposite results, probably due to financial crisis 2008-2009 present in the observed period. Nevertheless, given the dataset, such result is not very surprising. The average excess return in January is approximately -0.011 in the sample period, while the average return in other months equals 0.00345. Furthermore, the constant term shows statistical significance as in all of our estimations so

far. The rest of explanatory variables do not show statistical significance in the observed period hence they do not appear to have an important effect on S&P 500 excess returns.

Table 4.5: Results from frequentist model averaging

	Coefficient	Sd.Err	p-value
dividend_yield	-13.095	2.710	0.000
earnings_yield	0.149	0.565	0.792
book_to_market	-0.278	0.112	0.013
credit_spread	5.934	1.591	0.000
Tbill	-4.357	0.565	0.000
ch_Tbill	-1.698	1.917	0.376
term_premium	-6.274	1.775	0.000
term_spread	2.872	1.838	0.118
yield_spread	-0.133	1.733	0.939
January_dummy	-0.016	0.009	0.080
inflation	0.094	0.130	0.470
ch_inflation	0.099	0.261	0.705
ind_production	0.323	0.142	0.023
vol_p	0.000	0.000	0.134
oil_prices	0.021	0.061	0.725
size_premium	0.003	0.081	0.974
payout_ratio	0.071	0.012	0.000
lagged1	-0.244	0.065	0.000
lagged2	-0.300	0.063	0.000
DRP	-6.954	1.156	0.000
constant	0.501	0.069	0.000

In addition, complete summary of results from both model averaging techniques including estimated coefficients and standard deviations are reported in Table A.13 in Appendix. Except the variable *dividend yield*, we observe that the estimated coefficients are higher in magnitude in Mallows model averaging. Overall, we do not observe any remarkable differences across models which is in accordance with the evidence of Amini & Parmeter (2012).

Recently, researchers favour frequentist model averaging over the Bayesian approach because of two main reasons. Firstly, BMA requires selection of priors that are subjectively chosen by the researcher and secondly Bayesian approach may be computationally very demanding when estimating extremely large number of possible models (Moral-Benito 2015). Besides, frequentist model aver-

aging is believed to have better predictive performance. We assess its quality of predictions in Section 4.5.

4.4 Model Selection Criteria

To compare the results from model averaging techniques with more traditional approach, we run the estimation based on model selection criteria. In the previous literature, the use of model selection criteria is a popular procedure in order to choose the "best" model from a set of candidate sub-models. Bossaerts & Hillion (1999) propose to employ such procedure in the context of examining excess stock returns. They claim that including irrelevant predictors may lead to unnecessary complexity.

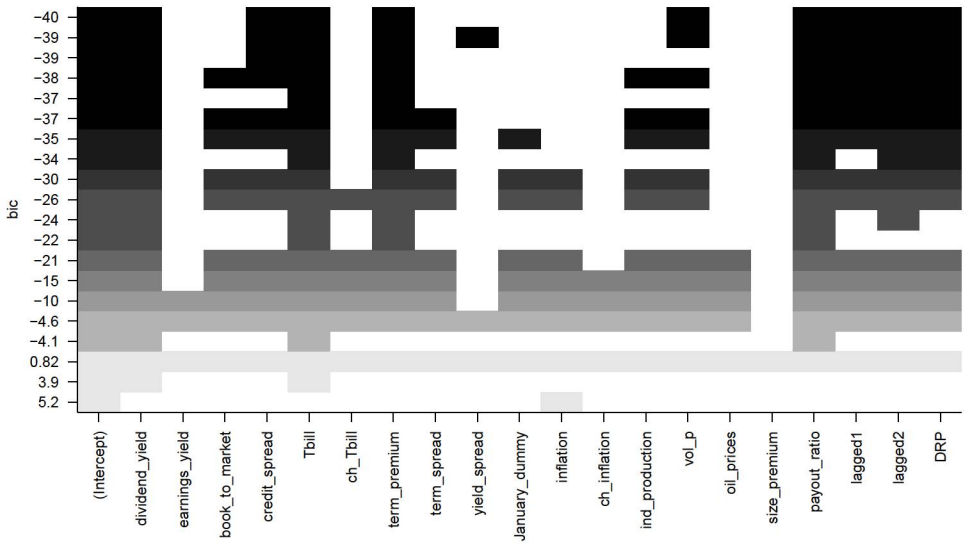
One way of finding the best model is to fit a full linear model, then eliminate variables one-at-a-time and compare the statistics from the model summary. However, this procedure may be exhaustive and very time consuming considering large number of possible predictors. We chose more attractive way that is offered by package *leaps* in R. We decided to select the best model according to four statistical criteria - the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the Mallows' criterion (abbreviated as C_p) and the adjusted R^2. The choice of these specific criteria depends simply on their popularity in the literature and the fact that the model selection is easily conducted in R. We will not discuss the advantages and disadvantages of individual criteria, since it is not the main interest of this thesis. However, there exist reasons in favour of each of them. Formal definitions of all four model selection criteria used in this thesis are outlined in Section 3.3.

For selecting the best model, we use the function *regsubsets* from package leaps in statistical software R. We specify only one best model in each number of variables included in the model. Then it identifies the best predictor submodel that contains a given number of predictors, where "best" is quantified using sum of square residuals. The summary of this process is depicted in Table A.11 in Appendix. The output reports the best set of variables for each model size (i.e. for 1 to 20 predictor models). For example, the best subset model with only one variable includes *inflation*. As already mentioned, we specify several criteria, namely Mallows' C_p, BIC and adjusted R^2. In case of first two criteria, we aim to find the best sub-model with the lowest value. On the other hand,

while using adjusted R-squared, we want the reduced model with the highest value. In Figure A.6 in Appendix, the graphs plotting values of these criteria for the submodel of each model size are presented and the highest (or lowest) point is marked by red colour.

Following Figure 4.6 shows graphical representation of best submodels ordered by a specific criterion. For illustration purposes, we present only best submodels ordered according to BIC. The first row represents the best overall model based on BIC criterion, therefore the model with nine variables has the lowest value of BIC. On the x-axis, all explanatory variables including the intercept are depicted, while the y-axis shows values of BIC of each submodel. Similarly, plots of best submodels ordered according to adjusted R^2 and Mallows' C_p are presented in Figure A.4 and Figure A.5 in Appendix, respectively.

Figure 4.6: Plot of best submodels ordered according to BIC



Besides, we use the *stepAIC* function from package *MASS* in R to include one more statistical criterion. This function performs stepwise model selection by AIC. In other words, it selects a single model, based on AIC criterion for ending the process, rather than generating sets of possible models of different sizes. We summarize the best submodels chosen according to a specific model selection criterion in Table 4.6. The best submodels include 9, 13 and 14 variables selected by BIC, C_p and adjusted R^2, respectively. The Akaike information cri-

terion has selected 13 variables which is the same number as Mallows C_p has. It confirms the finding of Boisbunon *et al.* (2014) who claim that Mallows' Cp is equivalent to AIC in the case of (Gaussian) linear regression. As we expected, BIC has selected less variables than R^2_{adj} or AIC. The reason is that the Bayesian information criterion imposes much heavier penalty for adding more variables than other criteria which usually results in a sparser model. On the other hand, the number of predictors included by AIC and adjusted R^2 is very similar, in accordance with our expectations. The best submodel based on minimizing the BIC criterion includes these nine variables: *dividend yield, credit spread, Tbill, term premium, volume over price, payout ratio, DRP, lagged1* and *lagged2*. Both Mallows criterion and the Akaike criterion would add these four predictors to the nine variables already mentioned: *book-to-market ratio, industrial production, January dummy* and *term spread*. Adjusted R^2 would include one more variable, namely *inflation*.

Table 4.6: Inclusion of predictors according to model selection criteria

	Adj R^2	C_p	BIC	AIC
dividend_yield	1	1	1	1
earnings_yield	0	0	0	0
book_to_market	1	1	0	1
credit_spread	1	1	1	1
Tbill	1	1	1	1
ch_Tbill	0	0	0	0
term_premium	1	1	1	1
term_spread	1	1	0	1
yield_spread	0	0	0	0
January_dummy	1	1	0	1
inflation	1	0	0	0
ch_inflation	0	0	0	0
ind_production	1	1	0	1
vol_p	1	1	1	1
oil_prices	0	0	0	0
size_premium	0	0	0	0
payout_ratio	1	1	1	1
lagged1	1	1	1	1
lagged2	1	1	1	1
DRP	1	1	1	1

Note: 1 = inclusion, 0 = exclusion

However, there exist various different criteria that could be used for model selection in order to guard against data snooping. For instance, Bossaerts & Hillion (1999) include FIC, PIC, PLS and their new PLS-MDC[4] criterion (alongside adjusted R^2, AIC and BIC). Besides, they find out that the best model changes over time. It is interesting to compare model selection technique with model averaging technique, both based on minimizing Mallows' criterion. Here we have presented the best submodels chosen according to four statistical criteria and we assess their predictive performance in the following section.

4.5 Predictive Performance

Finally, this section aims to assess the predictive performance of model averaging techniques. Prediction is considered as a neutral criterion to compare different methods, hence we compare predictions from model averaging techniques to the ones from traditional model selection methods as well as from simple OLS regression. Although majority of previous studies consider only in-sample performance of predictions, what is more important is the out-of-sample predictive performance. Therefore, here we provide comparison of predictions among all methods employed in the analysis in-sample as well as out-of-sample. We base evaluating of predictive performance on three different criteria, namely root mean square error (RMSE), mean absolute error (MAE) and log-predictive score (LPS). All of these three scoring rules are negatively oriented, i.e. the lower value, the better performance of the model.

We follow the approach by Eicher *et al.* (2011) and divide the dataset randomly into a training set which is used for estimation and contains 80% of the data, and a testing set which is used to assess the quality of the resulting predictions (remaining 20% of the data). Then we compute root mean squared error of prediction which is the most popular measure used in the economics to evaluate the predictive performance. Because we divide the dataset randomly and each split generates a different fit, we use one hundred random splits rather than only one to obtain reliable results. Then the mean of a hundred different RMSEs is calculated[5]. Besides, to assess in-sample accuracy of predictions,

[4]It is the Markov dimension criterion (MDC) in conjunction with predictive least squares criterion (PLS).

[5]As we mentioned earlier, BMA estimation is quite time demanding, hence we specify only 100 000 burn-ins and 300 000 subsequent iterations for all BMA alternatives since we

we also compute RMSE of prediction based on full dataset. The results are summarized in Table 4.7.

Table 4.7: Root Mean Square Error - different models

	RMSE	
	In-sample	Out-of-sample
MS (Adjusted R^2)	0.03431981	0.03828878
MS (Mallow's C_p)	0.03440959	0.03864535
MS (BIC)	0.03568757	0.03958411
MS (AIC)	0.03440959	0.03840069
OLS (full model)	0.03422893	0.03847728
BMA (UIP + uniform model prior)	0.03511589	0.03874673
BMA (BRIC + random model prior)	0.03533821	0.04080510
BMA (hyper-g + random model prior)	0.03449299	0.03797870
FMA	0.03421904	0.03851625

Note: MS = Model Selection

To decide which method performs the best, we compare the values of RMSE of different models. Generally, lower RMSE indicates better model. Firstly, we aim to test whether model averaging techniques perform better than traditional models selected by classical criteria. We observe that FMA performs better than a model selection method. It attains the lowest root mean square error of in-sample predictions, specifically around 0.0342. However, BMA fails to provide superior predictions to the model selection criteria based on in-sample predictions. Secondly, we can conclude that FMA outperforms BMA in-sample. And thirdly, among BMA alternatives, hyper-g specification performs the best in-sample. To evaluate the predictions out-of-sample, we look at the second column of Table 4.7. We obtain following general results. The Bayesian model averaging performs a lot better out-of-sample confirming the evidence of Cremers (2002), contrary to predictive performance of frequentist model averaging approach. However, the performance of BMA depends on the selection of priors. The specification employing hyper-g prior performs better than any other BMA alternative and also outperforms all model selection techniques and frequentist model averaging. Hence FMA cannot provide better predictions than BMA out-of-sample.

repeat the procedure one hundred times. However, the correlations between analytical and MCMC PMPs are still very high, in particular higher than 0.99 for all prior structures.

However, several scholars suggest using mean absolute error as a more appropriate measure for assessing the quality of prediction. MAE gives the same weight to all errors, while RMSE penalizes large errors more. The steps for computing in-sample and out-of-sample predictions are the same as in the case of RMSE described earlier. Values are reported in the following Table 4.8:

Table 4.8: Mean Absolute Error - different models

	MAE	
	In-sample	Out-of-sample
MS (Adjusted R^2)	0.02638113	0.02976701
MS (Mallow's C_p)	0.0266116	0.03018427
MS (BIC)	0.02782025	0.03113038
MS (AIC)	0.02666197	0.02995081
OLS (full model)	0.02626750	0.02990394
BMA (UIP + uniform model prior)	0.02743226	0.03049002
BMA (BRIC + random model prior)	0.02909199	0.03197930
BMA (hyper-g + random model prior)	0.02671568	0.02971921
FMA	0.02625994	0.02992750

Note: MS = Model Selection

We do not observe almost any difference, compared to the concluding comments based on RMSE. As expected, values of MAE are slightly smaller than the the root mean squared errors in Table 4.7. Therefore, using MAE we arrive to very similar conclusion. Mallows model averaging (FMA) performs the best in-sample, while the BMA specification with hyper-g prior provides the best out-of-sample predictive performance. Surprisingly, other two BMA alternatives do not perform very well in terms of RMSE or MAE. They cannot beat simple OLS neither in-sample nor out-of-sample.

Steel (2017) argues that the use of simple point predictions is very common in the literature because of its simplicity, but they do not allow us to address the uncertainty about the prediction. Hence, we compute the log-predictive score for all of three BMA alternatives and also for OLS regression including all explanatory variables. The log-predictive score as a scoring rule has been used in the BMA context in Madigan & Raftery (1994), Fernandez *et al.* (2001b), Steel & Ley (2007) or Eicher *et al.* (2011). Again, lower LPS indicates better performance of a model. We report the results in Table 4.9.

Table 4.9: Logarithmic predictive score - different models

	LPS	
	In-sample	Out-of-sample
OLS (full model)	-1.95533	-1.814856
BMA (UIP + uniform model prior)	-1.928661	-1.810539
BMA (BRIC + random model prior)	-1.922384	-1.75913
BMA (hyper-g + random model prior)	-1.943672	-1.832951

The results show that the hyper-g specification has the lowest value of LPS and thus performs the best among BMA alternative specifications in-sample and out-of-sample. Such finding confirms the evidence of Feldkircher & Zeugner (2012). However, it cannot outperform a simple OLS regression in-sample.

Overall, it seems that the predictive performance of BMA analysis is sensitive to the choice of priors. To conclude, no other combination of priors beats hyper-g specification in terms of RMSE, MAE or LPS. Interestingly, FMA works better than BMA and than model selection methods in-sample as the literature suggests, however it cannot provide superior predictions out-of-sample. Contrary to evidence reported in Eicher *et al.* (2011), we cannot confirm the best predictive performance of UIP g-prior together with uniform model prior. Additionally, the specification employing BRIC prior on parameters together with beta-binomial model prior performs the worst compared to all other estimations.

Conclusion

Although a lot of effort has been devoted to examining the predictability of stock excess returns in the past, there is no consensus about which predictors should be included in the predictive regression. In this thesis, we have examined twenty potential determinants of S&P 500 stock returns in excess of three-month Treasury bill rate in the time period from June 1998 to December 2016. However, having a relatively large number of predictors, we need to incorporate the uncertainty about the model into the estimation procedure. For that reason we have employed two model averaging techniques, in particular Bayesian model averaging and more recently developed frequentist model averaging. In addition, we have examined robustness of results by different combinations of priors under BMA framework. Particularly, we have followed Eicher *et al.* (2011), who recommend using the unit information prior for the parameters and the uniform model prior for the model in our baseline specification. Additionally, we have performed the analysis implementing two alternative combinations of priors. Specifically, we used BRIC g-prior in combination with a beta-binomial model prior and the more informative hyper-g prior together with random model prior. The use of hyper-g prior is strongly suggested by scholars because it should be less sensitive to outliers. We have compared the results of both model averaging techniques with the traditional methods based on model selection criteria (adjusted R^2, AIC, BIC, and Mallows' criterion C_p). Lastly, we have assessed the quality of the predictions according to three scoring rules, namely root mean square error, mean absolute error and log-predictive score and decided which of the methods performs the best in-sample and out-of-sample. First two scoring rules, RMSE and MAE, assess the quality of point estimation, while the log-predictive score focuses on the entire predictive distribution.

First of all, the analysis confirms the general evidence in the financial literature that stock returns are predictable. We find only very little diversity in

the set of predictors identified as having an effect on the stock returns using different methods. Our findings suggest that among the most important variables to explain excess returns on S&P 500 stock index are: three-month Treasury bill rate, dividend yield, payout ratio, excess returns lagged twice, and default risk premium. Under the BMA framework, all of these predictors yield posterior inclusion probability higher than 0.95 indicating strong significance. On the other hand, three variables which appear to be the weakest predictors are: change in inflation, oil price and size premium. The dividend yield seems to have the largest effect in magnitude, followed by default risk premium and term premium. The constant term shows statistical significance in all models we have employed. It is believed that the choice of priors is the main weakness of Bayesian model averaging approach and may have impact on the results. However, we conclude that the set of most important variables is fairly robust to prior selection. There are only small differences in the magnitude of coefficients, but the posterior inclusion probabilities are, on average, considerably higher in the hyper-g specification compared to models with fixed-g priors.

In addition, our empirical analysis reveals some degree of model uncertainty therefore we recommend employing model averaging techniques to explain stock returns. Regarding the quality of predictions, frequentist model averaging outperforms BMA and model selection methods in-sample as the literature suggests, however it cannot provide superior predictions out-of-sample. Additionally, the predictive performance of BMA appears to be sensitive to the choice of priors. Among the BMA alternative specifications, no other combination of priors beats hyper-g specification in terms of root mean square error, mean absolute error or log-predictive score. In fact, the hyper-g prior specification provides superior predictions out-of-sample comparing to models selected by classical criteria and to frequentist model averaging. On the other hand, the specification employing the benchmark prior on parameters together with beta-binomial model prior performs the worst compared to all other estimations.

However, the approach implemented in this thesis has also some limitations. We have made assumptions such as normality and linearity. Besides, we subjectively chose the combinations of priors and we also preselected the set of 20 possible predictors according to previous literature which may lead to some kind of data snooping. Moreover, the time span of the dataset is limited because of the data availability at Thomson Reuters database and the presence

of crisis may have an impact on the results as well. Additionally, Amini & Parmeter (2012) argue that the comparison of a nonuniform prior BMA approach with Mallows model averaging may not necessarily be valid. Lastly, we treat the explanatory variables as fixed, while in reality most are lagged stochastic variables.

The simplicity of our analysis leaves a lot of room for future research. Current BMA approach may be extended to the dynamic Bayesian framework which allows models to change over time. Besides, since the frequentist model averaging is a fairly new topic, it would need further research as well. In the thesis we focus only on the U.S stock market. It may be interesting to examine possible determinants of international stock market returns which may differ from those in the United States. Moreover, there is a rising number of recently introduced variables which may possibly have an effect on excess stock returns. Therefore, our set of potential predictors should be expanded by variables such as consumption-wealth ratio or volatility of returns.

Despite the limitations, we believe that our analysis valuably contribute to the finance literature about the existence of stock return predictability. Based on the most recent data, we conclude that determinants of stock returns are not consistent in time and need to be examine regularly. Even though the OLS estimator is the most common estimator used in the context of return predictability, we recommend implementing model averaging techniques instead.

Bibliography

Amini, S. M. & C. F. Parmeter (2012): "Comparison of model averaging techniques: Assessing growth determinants." *Journal of Applied Econometrics* **27(5)**: pp. 870–876.

Avramov, D. (2002): "Stock return predictability and model uncertainty." *Journal of Financial Economics* **64(3)**: pp. 423–458.

Banz, R. W. (1981): "The relationship between return and market value of common stocks." *Journal of financial economics* **9(1)**: pp. 3–18.

Bates, J. M. & C. W. Granger (1969): "The combination of forecasts." *Or* pp. 451–468.

Boisbunon, A., S. Canu, D. Fourdrinier, W. Strawderman, & M. T. Wells (2014): "Akaike's information criterion, cp and estimators of loss for elliptically symmetric distributions." *International Statistical Review* **82(3)**: pp. 422–439.

Bossaerts, P. & P. Hillion (1999): "Implementing statistical criteria to select return forecasting models: what do we learn?" *The Review of Financial Studies* **12(2)**: pp. 405–428.

Brown, P. J., M. Vannucci, & T. Fearn (2002): "Bayes model averaging with selection of regressors." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64(3)**: pp. 519–536.

Buckland, S. T., K. P. Burnham, & N. H. Augustin (1997): "Model selection: an integral part of inference." *Biometrics* pp. 603–618.

Campbell, J. Y. (1987): "Stock returns and the term structure." *Journal of financial economics* **18(2)**: pp. 373–399.

Campbell, J. Y. & M. Yogo (2006): "Efficient tests of stock return predictability." *Journal of financial economics* **81(1)**: pp. 27–60.

Chairakwattana, K., S. Nathaphan *et al.* (2013): *Stock Return Predictability by Bayesian Model Averaging: Evidence from Stock Exchange of Thailand.* Faculty of Commerce and Accountancy, Thammasat University.

Charles, A., O. Darné, & J. H. Kim (2016): "International stock return predictability: Evidence from new statistical tests." *International Review of Financial Analysis* .

Chen, N.-F., R. Roll, & S. A. Ross (1986): "Economic forces and the stock market." *Journal of business* pp. 383–403.

Chen, S.-S. (2010): "Do higher oil prices push the stock market into bear territory?" *Energy Economics* **32(2)**: pp. 490–495.

Chipman, H., E. I. George, R. E. McCulloch, M. Clyde, D. P. Foster, & R. A. Stine (2001): "The practical implementation of bayesian model selection." *Lecture Notes-Monograph Series* pp. 65–134.

Ciccone, A. & M. Jarociński (2010): "Determinants of economic growth: will data tell?" *American Economic Journal: Macroeconomics* **2(4)**: pp. 222–46.

Claeskens, G., N. L. Hjort *et al.* (2008): *Model selection and model averaging*, volume 330. Cambridge University Press Cambridge.

Clyde, M. & E. I. George (2004): "Model uncertainty." *Statistical science* pp. 81–94.

Cremers, K. M. (2002): "Stock return predictability: A bayesian model selection perspective." *The Review of Financial Studies* **15(4)**: pp. 1223–1249.

Cunado, J. & F. P. de Gracia (2014): "Oil price shocks and stock market returns: Evidence for some european countries." *Energy Economics* **42**: pp. 365–377.

De Moor, L. & P. Sercu (2013): "The smallest firm effect: An international study." *Journal of International Money and Finance* **32**: pp. 129–155.

Draper, D. (1995): "Assessment and propagation of model uncertainty." *Journal of the Royal Statistical Society* **57(1)**: pp. 45–97.

EICHER, T. S., C. PAPAGEORGIOU, & A. E. RAFTERY (2011): "Default priors and predictive performance in bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

FAMA, E. F. (1991): "Efficient capital markets: Ii." *The journal of finance* **46(5)**: pp. 1575–1617.

FAMA, E. F. & K. R. FRENCH (1992): "The cross-section of expected stock returns." *the Journal of Finance* **47(2)**: pp. 427–465.

FAMA, E. F. & K. R. FRENCH (1993): "Common risk factors in the returns on stocks and bonds." *Journal of financial economics* **33(1)**: pp. 3–56.

FAMA, E. F. & J. D. MACBETH (1973): "Risk, return, and equilibrium: Empirical tests." *Journal of political economy* **81(3)**: pp. 607–636.

FELDKIRCHER, M. & S. ZEUGNER (2012): *Benchmark priors revisited: on adaptive shrinkage and the supermodel effect in Bayesian model averaging. Revised version of 2009.* 9-202. International Monetary Fund.

FERNANDEZ, C., E. LEY, & M. F. STEEL (2001a): "Benchmark priors for bayesian model averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

FERNANDEZ, C., E. LEY, & M. F. STEEL (2001b): "Model uncertainty in cross-country growth regressions." *Journal of applied Econometrics* **16(5)**: pp. 563–576.

FERSON, W. E. & C. R. HARVEY (1991): "The variation of economic risk premiums." *Journal of Political Economy* **99(2)**: pp. 385–415.

FERSON, W. E. & C. R. HARVEY (1995): "Predictability and time-varying risk in world equity markets." *Research in Finance* **13**: pp. 25–88.

FOSTER, F. D., T. SMITH, & R. E. WHALEY (1997): "Assessing goodness-of-fit of asset pricing models: The distribution of the maximal r2." *The Journal of Finance* **52(2)**: pp. 591–607.

GEORGE, E. I. & R. E. MCCULLOCH (1997): "Approaches for bayesian variable selection." *Statistica sinica* pp. 339–373.

GUPTA, R., S. HAMMOUDEH, M. P. MODISE, & D. K. NGUYEN (2014): "Can economic uncertainty, financial stress and consumer sentiments predict us

equity premium?" *Journal of International Financial Markets, Institutions and Money* **33**: pp. 367–378.

Hamilton, J. D. (1983): "Oil and the macroeconomy since world war ii." *Journal of political economy* **91(2)**: pp. 228–248.

Hansen, B. E. (2007): "Least squares model averaging." *Econometrica* **75(4)**: pp. 1175–1189.

Havranek, T., M. Rusnak, & A. Sokolova (2017): "Habit formation in consumption: A meta-analysis." *European Economic Review* **95**: pp. 142–167.

Hjort, N. L. & G. Claeskens (2003): "Frequentist model average estimators." *Journal of the American Statistical Association* **98(464)**: pp. 879–899.

Hoeting, J. A., D. Madigan, A. E. Raftery, & C. T. Volinsky (1999): "Bayesian model averaging: a tutorial." *Statistical science* pp. 382–401.

Huang, R. D., R. W. Masulis, & H. R. Stoll (1996): "Energy shocks and financial markets." *Journal of Futures Markets* **16(1)**: pp. 1–27.

Jones, C. M. & G. Kaul (1996): "Oil and the stock markets." *The Journal of Finance* **51(2)**: pp. 463–491.

Jordan, S. J., A. J. Vivian, & M. E. Wohar (2014): "Forecasting returns: new european evidence." *Journal of Empirical Finance* **26**: pp. 76–95.

Kadane, J. B. & N. A. Lazar (2004): "Methods and criteria for model selection." *Journal of the American statistical Association* **99(465)**: pp. 279–290.

Kang, W., R. A. Ratti, & K. H. Yoon (2015): "The impact of oil price shocks on the stock market return and volatility relationship." *Journal of International Financial Markets, Institutions and Money* **34**: pp. 41–54.

Kapetanios, G., V. Labhard, & S. Price (2008): "Forecasting using bayesian and information-theoretic model averaging: an application to uk inflation." *Journal of Business & Economic Statistics* **26(1)**: pp. 33–41.

Kass, R. E. & A. E. Raftery (1995): "Bayes factors." *Journal of the american statistical association* **90(430)**: pp. 773–795.

Kothari, S. P. & J. Shanken (1997): "Book-to-market, dividend yield, and expected market returns: A time-series analysis." *Journal of Financial Economics* **44(2)**: pp. 169–203.

Lamont, O. (1998): "Earnings and expected returns." *The journal of Finance* **53(5)**: pp. 1563–1587.

Leamer, E. E. (1978): "Specification searches."

Lettau, M. & S. Ludvigson (2001): "Consumption, aggregate wealth, and expected stock returns." *the Journal of Finance* **56(3)**: pp. 815–849.

Liang, F., R. Paulo, G. Molina, M. A. Clyde, & J. O. Berger (2008): "Mixtures of g priors for bayesian variable selection." *Journal of the American Statistical Association* **103(481)**: pp. 410–423.

Lo, A. W. & A. C. MacKinlay (1990): "Data-snooping biases in tests of financial asset pricing models." *The Review of Financial Studies* **3(3)**: pp. 431–467.

Madigan, D. & A. E. Raftery (1994): "Model selection and accounting for model uncertainty in graphical models using occam's window." *Journal of the American Statistical Association* **89(428)**: pp. 1535–1546.

Magnus, J. R., O. Powell, & P. Prüfer (2010): "A comparison of two model averaging techniques with an application to growth empirics." *Journal of econometrics* **154(2)**: pp. 139–153.

McMillan, D., A. Black, O. Klinkowska, & F. McMillan (2015): "The existence and source of stock return predictability: Evidence from dividend, output and consumption ratios." *Journal of Asset Management* **16(3)**: pp. 186–208.

Mitchell, T. J. & J. J. Beauchamp (1988): "Bayesian variable selection in linear regression." *Journal of the American Statistical Association* **83(404)**: pp. 1023–1032.

Moral-Benito, E. (2012): "Determinants of economic growth: a bayesian panel data approach." *Review of Economics and Statistics* **94(2)**: pp. 566–579.

Moral-Benito, E. (2015): "Model averaging in economics: An overview." *Journal of Economic Surveys* **29(1)**: pp. 46–75.

Park, J. & R. A. Ratti (2008): "Oil price shocks and stock markets in the us and 13 european countries." *Energy economics* **30(5)**: pp. 2587–2608.

Pesaran, M. H. & A. Timmermann (1995): "Predictability of stock returns: Robustness and economic significance." *The Journal of Finance* **50(4)**: pp. 1201–1228.

Pontiff, J. & L. D. Schall (1998): "Book-to-market ratios as predictors of market returns." *Journal of Financial Economics* **49(2)**: pp. 141–160.

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

Rapach, D. & G. Zhou (2013): "Forecasting stock returns." In "Handbook of economic forecasting," volume 2, pp. 328–383. Elsevier.

Rapach, D. E., M. C. Ringgenberg, & G. Zhou (2016): "Short interest and aggregate stock returns." *Journal of Financial Economics* **121(1)**: pp. 46–65.

Rapach, D. E. & M. E. Wohar (2006): "In-sample vs. out-of-sample tests of stock return predictability in the context of data mining." *Journal of Empirical Finance* **13(2)**: pp. 231–247.

Richardson, M. (1993): "Temporary components of stock prices: A skeptic's view." *Journal of Business & Economic Statistics* **11(2)**: pp. 199–207.

Rjoub, H., T. Türsoy, & N. Günsel (2009): "The effects of macroeconomic factors on stock returns: Istanbul stock market." *Studies in Economics and Finance* **26(1)**: pp. 36–45.

Sadorsky, P. (1999): "Oil price shocks and stock market activity." *Energy economics* **21(5)**: pp. 449–469.

Schrimpf, A. (2010): "International stock return predictability under model uncertainty." *Journal of International Money and Finance* **29(7)**: pp. 1256–1282.

STEEL, M. F. (2017): "Model averaging and its use in economics." *arXiv preprint arXiv:1709.08221* .

STEEL, M. F. & E. LEY (2007): *On the effect of prior assumptions in Bayesian model averaging with applications to growth regression.* The World Bank.

WAN, A. T. & X. ZHANG (2009): "On the use of model averaging in tourism research." *Annals of Tourism Research* **36(3)**: pp. 525–532.

WANG, H., X. ZHANG, & G. ZOU (2009): "Frequentist model averaging estimation: a review." *Journal of Systems Science and Complexity* **22(4)**: p. 732.

WELCH, I. & A. GOYAL (2007): "A comprehensive look at the empirical performance of equity premium prediction." *The Review of Financial Studies* **21(4)**: pp. 1455–1508.

WHITELAW, R. F. (1994): "Time variations and covariations in the expectation and volatility of stock market returns." *The Journal of Finance* **49(2)**: pp. 515–541.

ZEUGNER, S. (2011): "Bayesian model averaging with bms." *Tutorial to the R-package BMS 1e30* .

ZHOU, H. D. (2011): "Incorporating model uncertainty and model instability in forecasting bond risk premia and term structure of government bond yield–a bayesian model averaging approach." .

ZIGRAIOVA, D. & T. HAVRANEK (2016): "Bank competition and financial stability: Much ado about nothing?" *Journal of Economic Surveys* **30(5)**: pp. 944–981.

Appendix A

Additional tables and figures

Table A.1: List of predictive variables and their abbreviations

	Explanation
dividend_yield	S&P 500 dividend yield
earnings_yield	S&P 500 earnings yield
book_to_market	book-to-market ratio of DJIA
credit_spread	credit spread
Tbill	3-month Treasury bill rate
ch_Tbill	change in 3-month Treasury bill rate
term_premium	term premium
term_spread	term spread
yield_spread	yield spread
January_dummy	January dummy
inflation	inflation rate
ch_inflation	change in inflation rate
ind_production	change in industrial production
vol_p	volume of shares traded over price level
oil_prices	change in oil prices
size_premium	size premium
payout_ratio	payout ratio
lagged1	excess returns on S&P 500 lagged once
lagged2	excess returns on S&P 500 lagged twice
DRP	default risk premium

Note: For detailed description of predictors see Table 2.1.

Table A.2: Correlation matrix of examined variables

	ExRet	DY	EY	B/M	cr_spread	Tbill	ch_Tbill	term_p	term_s	yield_spread	January	infl
ExRet	1	-0.101	0.020	-0.034	-0.098	-0.096	0.055	0.024	0.047	-0.107	-0.092	-0.158
DY	-0.101	1	0.010	0.677	0.754	-0.638	-0.067	0.368	0.379	-0.296	0.010	-0.159
EY	0.020	0.010	1	0.601	-0.385	-0.191	0.198	-0.133	-0.071	-0.136	0.002	0.470
B/M	-0.034	0.677	0.601	1	0.264	-0.681	0.193	0.305	0.345	-0.496	-0.021	0.072
cred_spread	-0.098	0.754	-0.385	0.264	1	-0.332	-0.219	0.300	0.274	-0.123	0.043	-0.178
Tbill	-0.096	-0.638	-0.191	-0.681	-0.332	1	0.004	-0.791	-0.826	0.616	-0.005	0.217
ch_Tbill	0.055	-0.067	0.198	0.193	-0.219	0.004	1	-0.081	-0.119	-0.474	-0.041	-0.041
term_p	0.024	0.368	-0.133	0.305	0.300	-0.791	-0.081	1	0.983	-0.520	-0.001	-0.113
term_s	0.047	0.379	-0.071	0.345	0.274	-0.826	-0.119	0.983	1	-0.505	0.012	-0.119
yield_spread	-0.107	-0.296	-0.136	-0.496	-0.123	0.616	-0.474	-0.520	-0.505	1	-0.014	0.255
January	-0.092	0.010	0.002	-0.021	0.043	-0.005	-0.041	-0.001	0.012	-0.014	1	0.003
infl	-0.158	-0.159	0.470	0.072	-0.178	0.217	-0.041	-0.113	-0.119	0.255	0.003	1
ch_infl	-0.032	-0.177	0.074	-0.092	-0.294	0.049	0.153	0.015	0.016	0.018	-0.011	0.167
ind_prod	0.055	-0.577	0.631	0.071	-0.665	0.283	0.197	-0.347	-0.322	0.145	-0.004	0.428
vol_p	-0.091	0.830	-0.068	0.487	0.768	-0.471	-0.097	0.397	0.401	-0.181	0.015	0.027
oil	0.084	-0.172	0.008	-0.067	-0.201	0.097	0.198	-0.032	-0.035	0.067	0.072	-0.019
SMB	0.057	-0.074	-0.053	-0.077	-0.013	-0.020	0.009	0.062	0.057	-0.067	-0.025	0.050
payout	0.066	0.613	-0.622	0.025	0.726	-0.211	-0.066	0.265	0.244	-0.147	-0.008	-0.405
ExRet1	0.087	-0.129	0.066	-0.012	-0.201	-0.088	0.119	0.038	0.051	-0.159	0.070	-0.109
ExRet2	-0.069	-0.115	0.116	0.0002	-0.245	-0.063	0.234	0.045	0.053	-0.197	0.054	-0.041
DRP	-0.082	0.599	-0.381	0.191	0.770	-0.515	-0.352	0.465	0.493	-0.124	0.0001	-0.229

The dependent variable is excess returns on S&P 500 (*ExRet*).
The explanatory variables are following: dividend yield (*DY*), earnings yield (*EY*), book-to-market ratio (*B/M*), credit spread (*cred_spread*), three-month Treasury bill rate (*Tbill*) and its first difference (*ch_Tbill*), term premium (*term_p*), term spread (*term_s*), yield spread (*yield_spread*), January dummy (*January*), inflation (*infl*) and its first difference (*ch_infl*), industrial production (*ind_prod*), volume over price (*vol_p*), oil prices (*oil*), size premium (*SMB*), payout ratio (*payout*), excess returns lagged once and twice (*ExRet1* and *ExRet2*), default risk premium (*DRP*).

Table A.3: Correlation matrix of examined variables - continued

	ch_infl	ind_prod	vol_p	oil	SMB	payout_ratio	ExRet1	ExRet2	DRP
ExRet	-0.032	0.055	-0.091	0.084	0.057	0.066	0.087	-0.069	-0.082
DY	-0.177	-0.577	0.830	-0.172	-0.074	0.613	-0.129	-0.115	0.599
EY	0.074	0.631	-0.068	0.008	-0.053	-0.622	0.066	0.116	-0.381
B/M	-0.092	0.071	0.487	-0.067	-0.077	0.025	-0.012	0.0002	0.191
cred_spread	-0.294	-0.665	0.768	-0.201	-0.013	0.726	-0.201	-0.245	0.770
Tbill	0.049	0.283	-0.471	0.097	-0.020	-0.211	-0.088	-0.063	-0.515
ch_Tbill	0.153	0.197	-0.097	0.198	0.009	-0.066	0.119	0.234	-0.352
term_p	0.015	-0.347	0.397	-0.032	0.062	0.265	0.038	0.045	0.465
term_s	0.016	-0.322	0.401	-0.035	0.057	0.244	0.051	0.053	0.493
yield_spread	0.018	0.145	-0.181	0.067	-0.067	-0.147	-0.159	-0.197	-0.124
January	-0.011	-0.004	0.015	0.072	-0.025	-0.008	0.070	0.054	0.0001
infl	0.167	0.428	0.027	-0.019	0.050	-0.405	-0.109	-0.041	-0.229
ch_infl	1	0.114	-0.130	0.236	-0.055	-0.182	0.133	0.195	-0.223
ind_prod	0.114	1	-0.573	0.128	-0.076	-0.802	0.051	0.086	-0.564
vol_p	-0.130	-0.573	1	-0.116	-0.031	0.671	-0.125	-0.099	0.597
oil	0.236	0.128	-0.116	1	0.048	-0.015	0.196	0.060	-0.205
SMB	-0.055	-0.076	-0.031	0.048	1	0.017	0.100	0.005	0.010
payout	-0.182	-0.802	0.671	-0.015	0.017	1	-0.015	-0.089	0.563
ExRet1	0.133	0.051	-0.125	0.196	0.100	-0.015	1	0.086	-0.215
ExRet2	0.195	0.086	-0.099	0.060	0.005	-0.089	0.086	1	-0.272
DRP	-0.223	-0.564	0.597	-0.205	0.010	0.563	-0.215	-0.272	1

The dependent variable is excess returns on S&P 500 (*ExRet*).
The explanatory variables are following: dividend yield (*DY*), earnings yield (*EY*), book-to-market ratio (*B/M*), credit spread (*cred_spread*), three-month Treasury bill rate (*Tbill*) and its first difference (*ch_Tbill*), term premium (*term_p*), term spread (*term_s*), yield spread (*yield_spread*), January dummy (*January*), inflation (*infl*) and its first difference (*ch_infl*), industrial production (*ind_prod*), volume over price (*vol_p*), oil prices (*oil*), size premium (*SMB*), payout ratio (*payout*), excess returns lagged once and twice (*ExRet1* and *ExRet2*), default risk premium (*DRP*).

Figure A.1: PMP convergence based on 20 best models

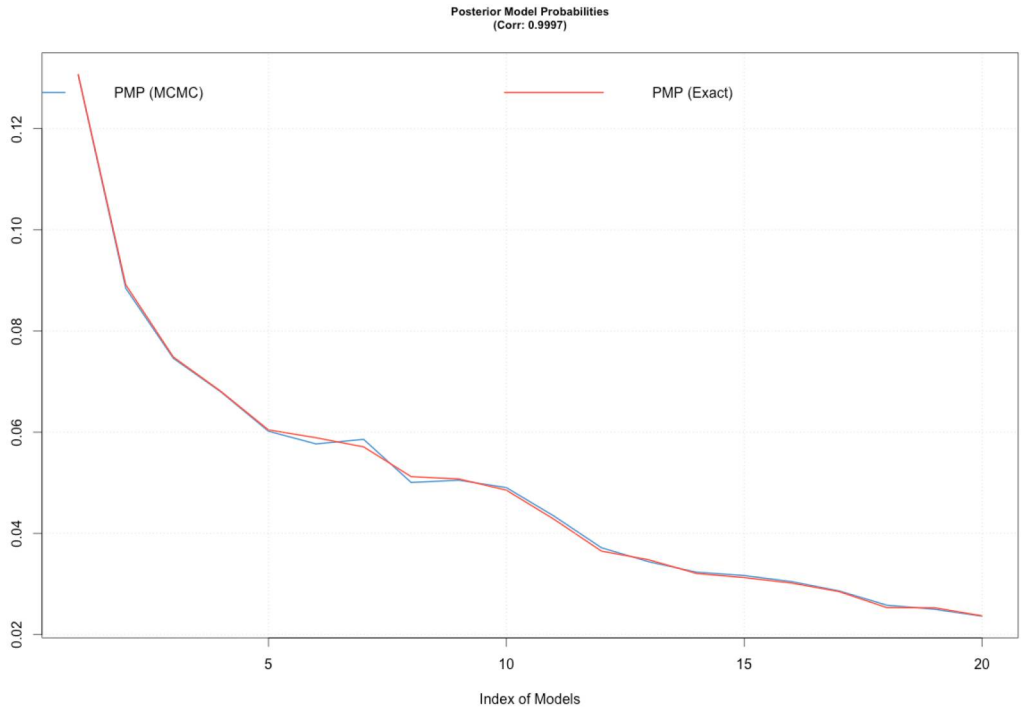


Table A.4: BMA baseline specification using reverse-jump sampler

	PIP	Post Mean	Post SD	Cond.Pos.Sign	Idx
dividend_yield	1.000	-14.289	2.858	0	1
earnings_yield	0.204	0.147	0.363	0.995	2
book_to_market	0.271	-0.053	0.105	0.003	3
credit_spread	0.851	3.819	2.139	1.000	4
Tbill	1	-3.722	0.591	0	5
ch_Tbill	0.171	-0.406	1.122	0	6
term_premium	0.989	-3.794	1.480	0	7
term_spread	0.242	0.654	1.544	0.954	8
yield_spread	0.276	0.703	1.350	1.000	9
January_dummy	0.148	-0.002	0.006	0	10
inflation	0.211	0.039	0.091	1.000	11
ch_inflation	0.069	0.007	0.071	0.874	12
ind_production	0.257	0.064	0.134	1	13
vol_p	0.423	0.0002	0.0002	1	14
oil_prices	0.063	0.0004	0.015	0.698	15
size_premium	0.062	-0.0002	0.020	0.423	16
payout_ratio	1	0.067	0.011	1	17
lagged1	0.955	-0.210	0.078	0	18
lagged2	0.999	-0.276	0.063	0	19
DRP	0.999	-5.096	1.304	0	20
(Intercept)	1	0.429			0

Table A.5: Frequentist check (OLS with variables having PIP>0.5)

	Dependent variable:
	excess_returns
dividend_yield	−13.372***
	(1.661)
Tbill	−3.437***
	(0.389)
payout_ratio	0.061***
	(0.008)
DRP	−4.720***
	(0.880)
credit_spread	3.822***
	(1.420)
term_premium	−3.010***
	(0.468)
lagged1	−0.206***
	(0.065)
lagged2	−0.269***
	(0.062)
Constant	0.401***
	(0.043)
Observations	223
R^2	0.324
Adjusted R^2	0.299
Residual Std. Error	0.037 (df = 214)
F Statistic	12.833*** (df = 8; 214)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.6: Comparison of two studies - Avramov (2002) and Cremers (2002)

	Avramov (2002)	Cremers (2002)
Time period	April 1953 - December 1998	January 1954 - January 1998
Frequency of obs.	monthly and quarterly	monthly
Dependent variable	Equity portfolio return (six size/book-to-market portfolios)	S&P 500 excess return
Predictive variables	dividend yield on the NYSE	dividend yield
	earnings yield	earnings yield
	book-to-market	excess returns lagged once
	winners-minus-losers	excess returns lagged twice
	default risk spread	credit spread
	3-month T-bill rate	3-month T-bill rate
	excess return on value-weighted	change in T-bill
	default risk premium	yield spread
	term structure premium	term spread
	January dummy	January dummy
	inflation rate	inflation rate
	size premium	change in inflation rate
	value premium	volume of shares over price
	term structure slope	industrial production
	all variables lagged once	*only three macroeconomic variables lagged once*
Model prior	equal	skeptic/confident investor
Most robust predictors	EY	volume over price
	inflation rate	credit spread
	term structure premium	ch_inflation / Term
	January dummy	January dummy
	Tbill	Tbill
		ch_Tbill
		yield spread
Weakest predictors	DY	DY
	WML	ExRet-1
	HML	ExRet-2
		inflation

Table A.7: BMA with term spread and term premium separately

	only term_premium			only term_spread		
	Post Mean	Post SD	PIP	Post Mean	Post SD	PIP
dividend_yield	-14.49	2.85	1.00	-12.75	2.66	1.00
earnings_yield	0.16	0.38	0.22	0.62	0.59	0.61
book_to_market	-0.04	0.10	0.24	-0.03	0.08	0.18
credit_spread	3.50	2.08	0.83	1.09	1.72	0.36
Tbill	-3.75	0.60	1.00	-3.13	0.63	1.00
ch_Tbill	-0.46	1.19	0.19	-0.72	1.52	0.24
term_spread	-	-	-	-2.76	0.69	1.00
term_premium	-3.24	0.62	1.00	-	-	-
yield_spread	0.78	1.41	0.30	0.87	1.53	0.31
January_dummy	-0.00	0.01	0.14	-0.00	0.00	0.10
inflation	0.04	0.09	0.21	0.01	0.05	0.09
ch_inflation	0.01	0.07	0.07	0.00	0.07	0.06
ind_production	0.06	0.12	0.24	0.04	0.09	0.18
vol_p	0.00	0.00	0.46	0.00	0.00	0.28
oil_prices	0.00	0.01	0.06	-0.00	0.02	0.06
size_premium	-0.00	0.02	0.06	-0.00	0.02	0.06
payout_ratio	0.07	0.01	1.00	0.07	0.01	1.00
lagged1	-0.21	0.08	0.95	-0.15	0.10	0.79
lagged2	-0.27	0.06	1.00	-0.25	0.07	0.99
DRP	-4.89	1.22	1.00	-2.94	1.36	0.92
(Intercept)	0.43		1.00	0.33		1.00

Table A.8: Model diagnostics, all three BMA specifications

	BMA1	BMA2	BMA3
Mean no. regressors	10.1986	9.4869	17.2239
Draws	3e+06	3e+06	3e+06
Burnins	1e+06	1e+06	1e+06
Time	14.75298 mins	13.35791 mins	16.51003 mins
No. models visited	917629	844266	1058791
Model space 2^K	1e+06	1e+06	1e+06
% visited	88	81	101
% Topmodels	100	100	100
Corr PMP	0.9998	0.9998	1.0000
No. Obs.	223	223	223
Model Prior	uniform / 10	random /10	random /10
g-Prior	UIP	BRIC	hyper (a=2.008969)
Shrinkage-Stats	Av=0.9955	Av=0.9975	Av=0.8667, Stdev=0.05

Note: BMA1 = UIP + uniform model prior , BMA2 = BRIC + random, BMA3 = hyper-g + random

Figure A.2: Model size and convergence, BRIC + random model prior

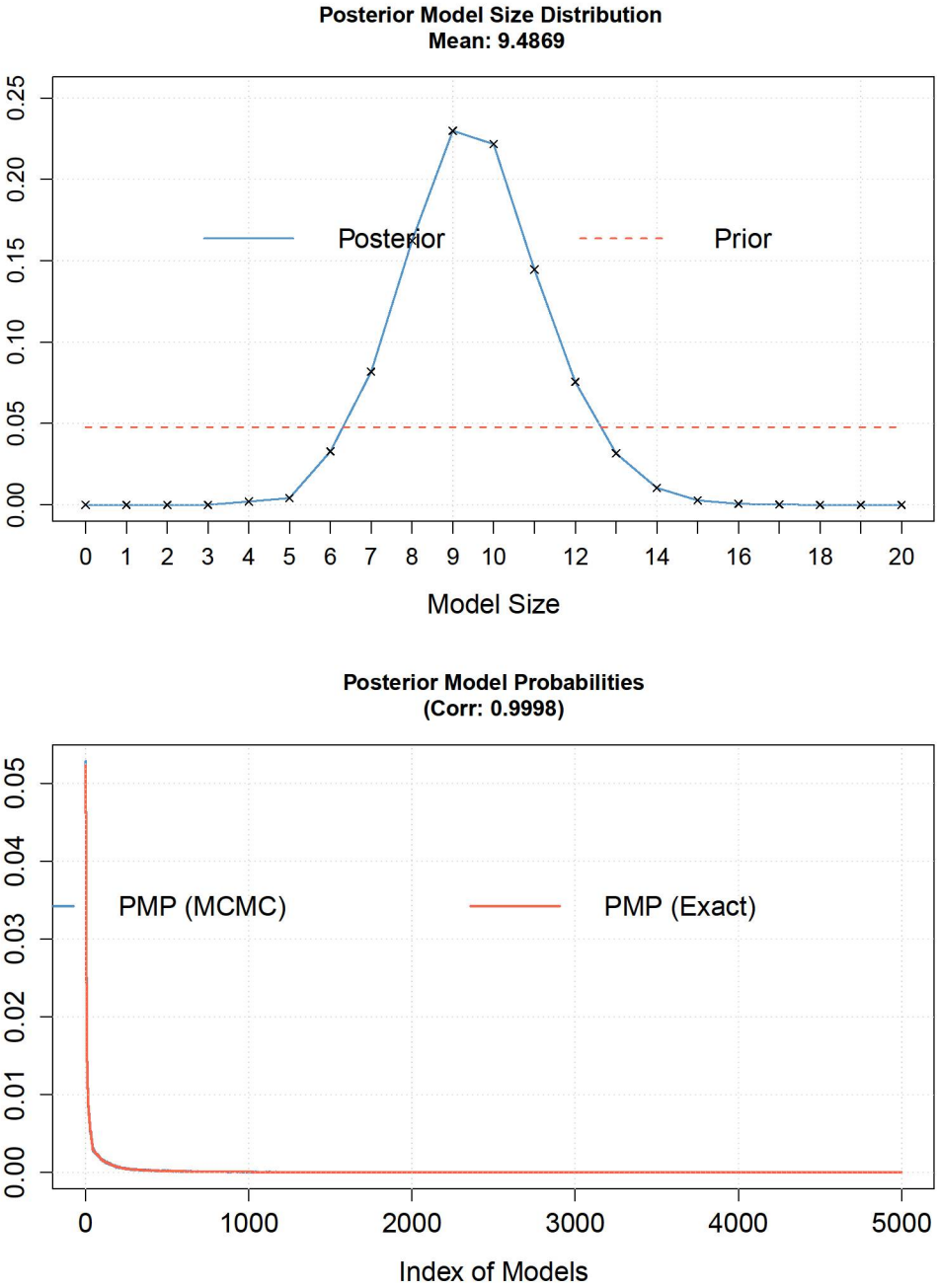


Figure A.3: Model size and convergence, hyper-g + random model prior

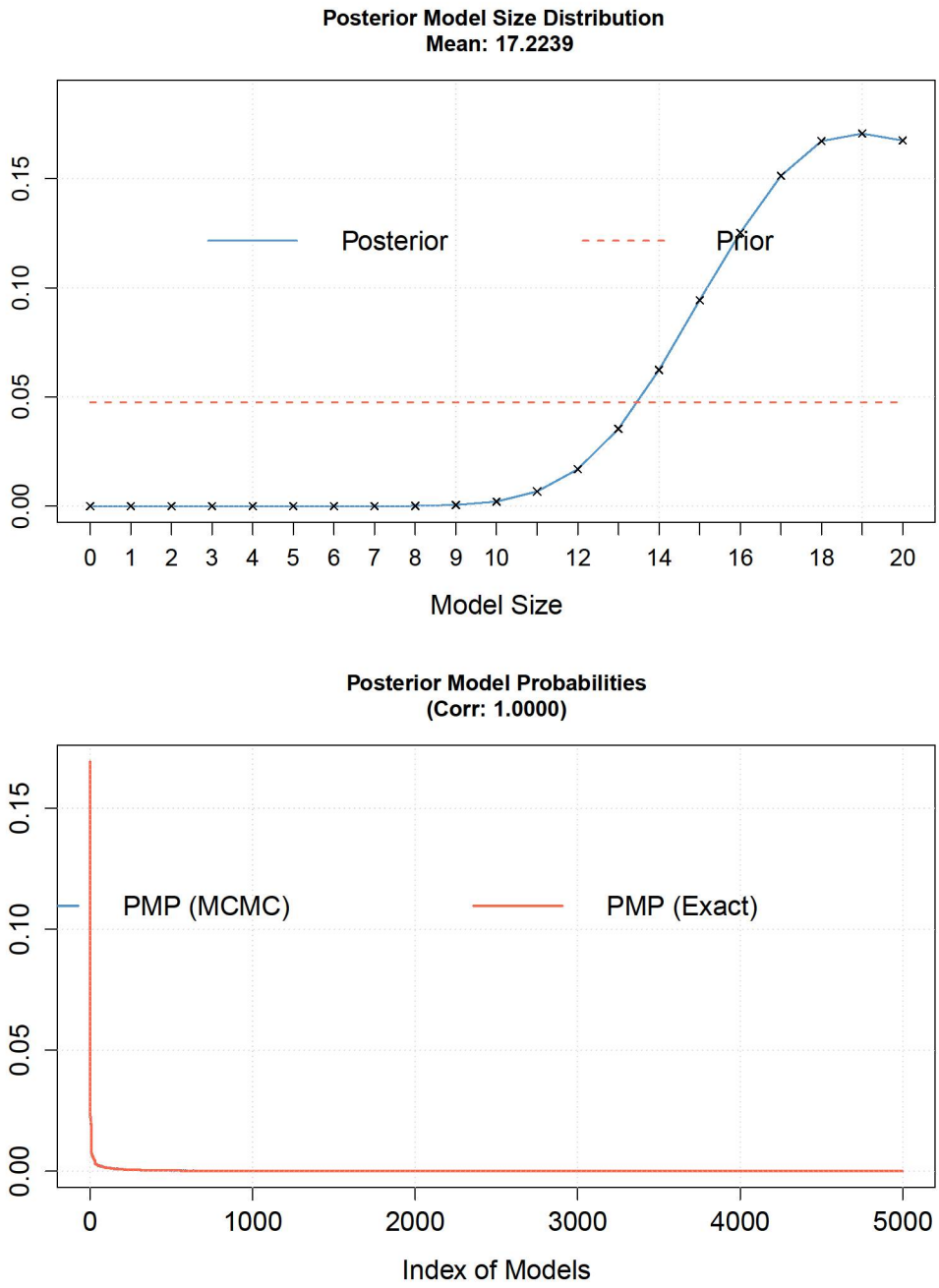


Figure A.4: Plot of best submodels ordered according to Adjusted R2

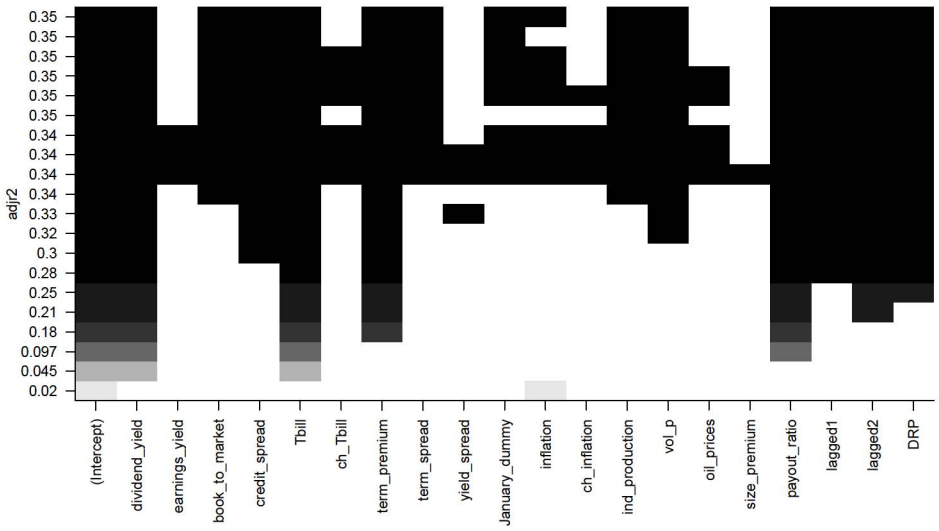


Figure A.5: Plot of best submodels ordered according to Cp

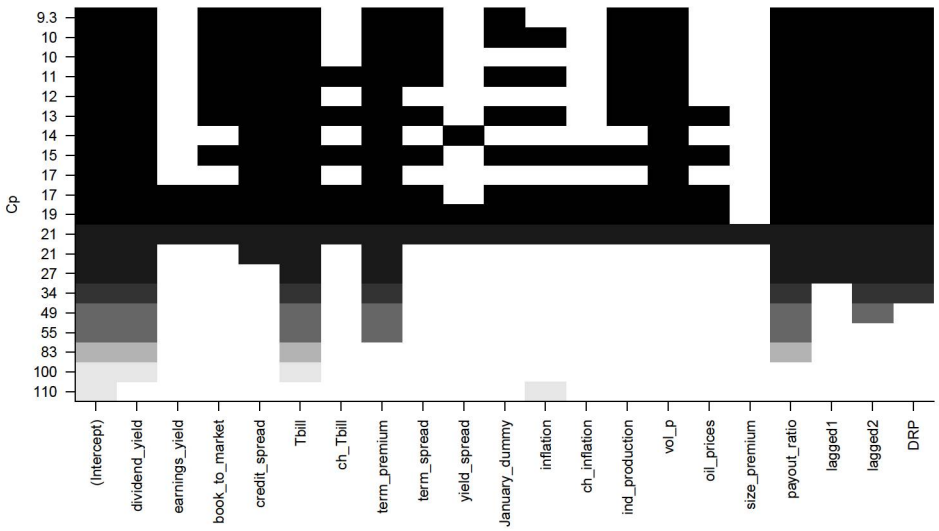


Figure A.6: Choosing best submodel according to statistical criteria (full sample)

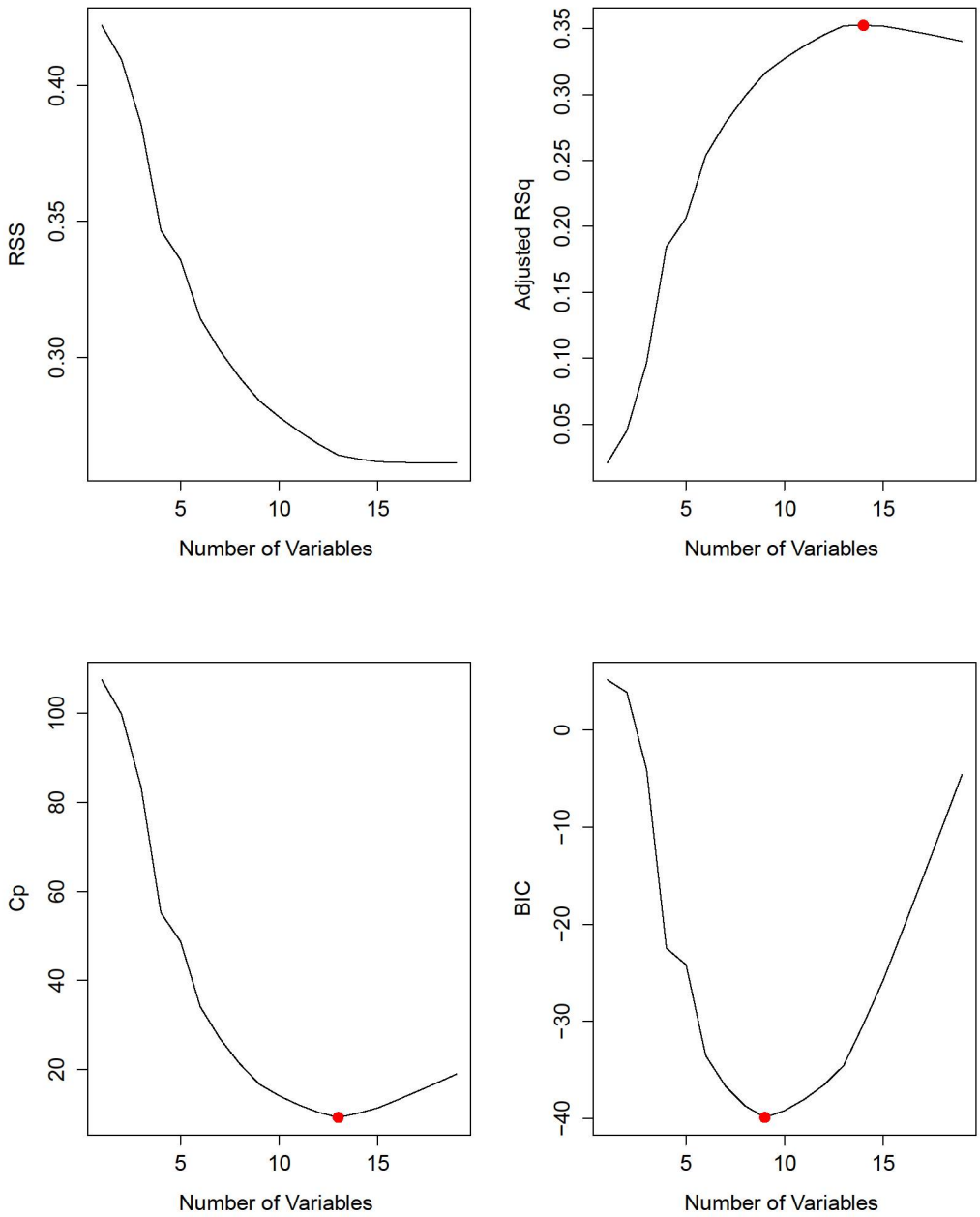


Table A.9: Sign certainty of predictors

	BMA_1	BMA_2	BMA_3
dividend_yield	0.000	0.000	0.000
earnings_yield	0.995	0.996	0.982
book_to_market	0.003	0.004	0.000
credit_spread	1.000	1.000	1.000
Tbill	0.000	0.000	0.000
ch_Tbill	0.000	0.000	0.000
term_premium	0.000	0.000	0.000
term_spread	0.954	0.898	0.998
yield_spread	1.000	1.000	0.557
January_dummy	0.000	0.000	0.000
inflation	1.000	0.999	0.999
ch_inflation	0.874	0.826	0.998
ind_production	1.000	1.000	1.000
vol_p	1.000	1.000	1.000
oil_prices	0.698	0.687	0.970
size_premium	0.424	0.444	0.823
payout_ratio	1.000	1.000	1.000
lagged1	0.000	0.000	0.000
lagged2	0.000	0.000	0.000
DRP	0.000	0.000	0.000
(Intercept)			

Note: 1 = posterior probability of a positive coefficient expected value conditional on inclusion

Table A.10: Top 3 models inclusion - BMA alternative specifications

	BRIC + random			hyper-g + random		
	#1	#2	#3	#1	#2	#3
dividend_yield	1	1	1	1	1	1
earnings_yield	0	0	0	1	1	1
book_to_market	0	0	0	1	1	1
credit_spread	1	1	0	1	1	1
Tbill	1	1	1	1	1	1
ch_Tbill	0	0	0	1	1	1
term_premium	1	1	1	1	1	1
term_spread	0	0	0	1	1	1
yield_spread	0	0	0	1	1	0
January_dummy	0	0	0	1	1	1
inflation	0	0	0	1	1	1
ch_inflation	0	0	0	1	1	1
ind_production	0	0	0	1	1	1
vol_p	0	1	0	1	1	1
oil_prices	0	0	0	1	1	1
size_premium	0	0	0	1	0	1
payout_ratio	1	1	1	1	1	1
lagged1	1	1	1	1	1	1
lagged2	1	1	1	1	1	1
DRP	1	1	1	1	1	1
PMP (Exact)	0.052	0.051	0.043	0.169	0.023	0.023
PMP (MCMC)	0.053	0.051	0.043	0.167	0.022	0.022

Table A.11: Summary of best submodels in model selection

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18	19	20
1 (1)											*									
2 (1)	*				*															
3 (1)	*				*												*			
4 (1)	*				*		*										*			
5 (1)	*				*		*										*		*	
6 (1)	*				*		*										*		*	*
7 (1)	*				*		*										*	*	*	*
8 (1)	*			*	*		*										*	*	*	*
9 (1)	*			*	*		*							*			*	*	*	*
10 (1)	*			*	*		*		*					*			*	*	*	*
11 (1)	*		*	*	*		*						*	*			*	*	*	*
12 (1)	*		*	*	*		*	*					*	*			*	*	*	*
13 (1)	*		*	*	*		*	*		*			*	*			*	*	*	*
14 (1)	*		*	*	*		*	*		*	*		*	*			*	*	*	*
15 (1)	*		*	*	*	*	*	*		*	*		*	*			*	*	*	*
16 (1)	*		*	*	*	*	*	*		*	*		*	*	*		*	*	*	*
17 (1)	*		*	*	*	*	*	*		*	*	*	*	*	*		*	*	*	*
18 (1)	*	*	*	*	*	*	*	*		*	*	*	*	*	*		*	*	*	*
19 (1)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*
20 (1)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

The numbers in the first row of the table stand for the predictors as follows:
1. dividend yield, 2. earnings yield, 3. book-to-market ratio, 4. Credit spread, 5. Yield on a short-term Treasury bill, 6. Change in the yield on a short-term Treasury bill, 7. Term premium, 8. Term spread, 9. Yield spread, 10. January dummy, 11. Inflation Rate, 12. Change in inflation, 13. Change in industrial production, 14. Volume of shares traded over price level, 15. Oil prices, 16. Size premium, 17. Payout ratio, 18. Default Risk Premium, 19. lagged1, 20. lagged2

Table A.12: OLS with all variables lagged

	Dependent variable:
	excess_returns
dividend_yield	**10.679*****
	(3.185)
earnings_yield	**2.136*****
	(0.633)
book_to_market	**−0.453*****
	(0.124)
credit_spread	**−2.492**
	(1.852)
Tbill	**0.688**
	(0.708)
ch_Tbill	**−0.837**
	(2.460)
term_premium	**−0.840**
	(1.983)
term_spread	**2.088**
	(2.011)
yield_spread	**−2.052**
	(2.142)
January_dummy	**−0.012**
	(0.010)
inflation	**−0.409*****
	(0.152)
ch_inflation	**−0.013**
	(0.353)
ind_production	**0.614*****
	(0.166)
vol_p	**−0.0005***
	(0.0002)
oil_prices	**0.094**
	(0.082)
size_premium	**0.136**
	(0.090)
payout_ratio	**0.047*****
	(0.016)
DRP	**0.875**
	(1.363)
lagged1	**−0.006**
	(0.075)
lagged2	**−0.146****
	(0.073)
Constant	**−0.184****
	(0.084)
Observations	223
R^2	0.251
Adjusted R^2	0.177
Residual Std. Error	0.040 (df = 202)
F Statistic	3.381*** (df = 20; 202)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A.13: Summary of model averaging techniques (Coefficient, StDev)

	BMA1 Post Mean (Post SD)	BMA2 Post Mean (Post SD)	BMA3 Post Mean (Post SD)	MMA Coefficient (Sd.Err)
dividend_yield	-14.289	-13.716	-11.506	-13.095
	(2.858)	(2.956)	(2.620)	(2.710)
earnings_yield	0.147	0.126	0.173	0.149
	(0.363)	(0.339)	(0.462)	(0.565)
book_to_market	-0.053	-0.038	-0.214	-0.278
	(0.105)	(0.093)	(0.114)	(0.112)
credit_spread	3.819	3.170	4.937	5.934
	(2.139)	(2.374)	(1.511)	(1.591)
Tbill	-3.722	-3.549	-3.698	-4.357
	(0.591)	(0.654)	(0.514)	(0.565)
ch_Tbill	-0.406	-0.319	-1.066	-1.698
	(1.122)	(1.006)	(1.611)	(1.917)
term_premium	-3.794	-3.517	-5.006	-6.274
	(1.480)	(1.445)	(1.700)	(1.775)
term_spread	0.654	0.477	2.125	2.872
	(1.544)	(1.420)	(1.811)	(1.838)
yield_spread	0.703	0.548	0.228	-0.133
	(1.350)	(1.226)	(1.376)	(1.733)
January_dummy	-0.002	-0.001	-0.011	-0.016
	(0.006)	(0.005)	(0.009)	(0.009)
inflation	0.039	0.032	0.067	0.094
	(0.091)	(0.085)	(0.109)	(0.130)
ch_inflation	0.007	0.005	0.063	0.099
	(0.071)	(0.063)	(0.203)	(0.261)
ind_production	0.064	0.048	0.248	0.323
	(0.134)	(0.118)	(0.147)	(0.142)
vol_p	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
oil_prices	0.000	0.000	0.010	0.021
	(0.015)	(0.013)	(0.046)	(0.061)
size_premium	-0.000	-0.000	0.002	0.003
	(0.020)	(0.018)	(0.061)	(0.081)
payout_ratio	0.067	0.066	0.062	0.071
	(0.011)	(0.011)	(0.011)	(0.012)
lagged1	-0.210	-0.191	-0.207	-0.244
	(0.078)	(0.093)	(0.061)	(0.065)
lagged2	-0.276	-0.271	-0.256	-0.300
	(0.063)	(0.068)	(0.059)	(0.063)
DRP	-5.096	-4.667	-5.777	-6.954
	(1.304)	(1.511)	(1.086)	(1.156)
(Intercept)	0.429	0.410	0.422	0.501
				(0.069)

BMA1 = UIP + uniform model prior
BMA2 = BRIC + random model prior
BMA3 = hyper-g + random model prior
MMA = Mallows model averaging (FMA)