Parameter & Variable-Selection Uncertainty in Asset Allocation. Should a decision-theorist investor take variable-selection uncertainty into account?

Erasmus University Rotterdam

Author:
Marios Lioutas (382859)

Supervisor:
Dr. Bart Diris

First Reader:
Dr. Bart Keijsers

July, 2016

Abstract

The present thesis focuses on the problem of finding the optimal allocation strategy in a financial portfolio, using an econometric point of view. Its main contribution is the investigation of a new Bayesian approach for the portfolio choice. Markov Chain Monte Carlo (MCMC) algorithm, recently proposed in the Bayesian literature, is introduced and applied for a decision-theoretic approach of the optimal weight asset allocation strategy. In particular, the Gibbs sampler proposed by Korobilis (2013) is used in order to estimate the parameters of econometric models for finding the optimal portfolio of an investor. The proposed approach, except for the parameter uncertainty, takes into account the variable-selection uncertainty. More precisely, a computationally efficient algorithm for variable selection is proposed and the approach is compared with the ones in the relevant econometric literature for a managed decision-theoretic portfolio construction.

Design is not making beauty, beauty emerges from selection, affinities, integration, love

Louis Kahn

Contents

1 Introduction

This paper aims to describe the problem of portfolio choice using an econometric view. The research is focused on the choice of the optimal portfolio weights and the decision-theoretic approach. The main contribution of the present thesis is the introduction of the most recent techniques for the Bayesian estimation of a model, applied in the problem of the optimal portfolio choice. The methods which are used are compared with existing approaches in the literature and improvements are deduced for finding the optimal portfolio allocation strategy. The model which serves as a basis is the vector autoregressive (VAR) model. The state vector in the VAR may include asset returns and predictor variables (financial and macroeconomics factors). Specifically, the managed portfolio approach is used, in combination with conditional information based on some financial and macroeconomics factors for future investment. Using conditional information based on some factors that have significant explanatory power, the robustness and reliability of this portfolio can be tested in several ways. In more detail, the parameters of the models used are estimated, taking into account the parameter uncertainty, while a significant contribution of the thesis is the additional use of computationally efficient algorithms for stochastic variable selection in the model.

As documented by Brandt (2009), the economic theory underlying an investor's optimal portfolio choice, pioneered by Markowitz (1952), Merton (1969) and Merton (1971), Samuelson (1969), and Fama (1970), is by now well understood. Based on recent empirical financial research, it is also well-known that there is substantial evidence for predictability of asset returns (see, for example Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988), Fama and French (1988), and Campbell (1990)). As a result, it is now argued by financial economists that asset return predictability can introduce a wedge between the asset allocation strategies of short and long-term investors. Welch and Goyal (2008) highlight the statement of Lettau and Ludvigson (2001):

> "It is now widely accepted that excess returns are predictable by variables such as dividend-price ratios, earnings-price ratios, dividend-earnings ratios, and an assortment of other financial indicators."

This statement summarizes the common findings in the finance literature that the dividend price ratio and dividend yield, the earnings price ratio and dividend-earnings (payout) ratio, various interest rates and spreads, the inflation rates, the book-to-market ratio, volatility, the investment-capital ratio, the consumption, wealth, and income ratio, and aggregate net or equity issuing activity, are only some of the variables that can be used as predictors of asset returns. Except for this evidence of predictability of asset returns the renewest interest, of academic researchers, in the topic of practical portfolio advice, has been also fuelled by realistic issues including model and parameter uncertainty, learning, background risks and frictions. In this paper we also assume that there is a large number of predictor variables, based on the aforementioned literature.

There are two common approaches in the finance literature of portfolio choice problems: the plug-in estimation and the use of decision theory for calculating optimal portfolio weights. However, it is well-documented in the finance literature (see for example Barberis (2000) and Brandt (2009)), that the decision theoretic approach has many advantages compared with plug-in estimation for the calculation optimal portfolio weights. The main research question of the present thesis lies on the direction of decision theory. More precisely, previous efforts of treating the portfolio choice problem are reviewed, using well-known results from the decision theory. Finally, an alternative approach of choosing the optimal weights is proposed, using recent results from the statistical literature of Bayesian analysis and is further compared with existing approaches (Barberis (2000), Diris (2014)).

A common finding in the financial and econometric literature is the poor finite-sample properties of plug-in estimates, as documented in Brandt (2009). One proposed solution by Brandt (2009) to that problem is the use of factor models, where a significant dimension reduction is achieved. An additional approach of treating the poor finite-sample properties of plug-in estimations is the use of constraints on the portfolio weights, as fully analysed in Frost and Savarino (1988). Some of the most popular constraints in the finance literature are the portfolio constraints that limit short and long selling in risky assets. As noted by Brandt (2009), constraints on the maximum position in a single security, on the maximum exposure to a given industry or economic sector, on the liquidity of a security, or on the risk characteristics of a security are constraints arising in realistic investment problems. Decision theory gives an alternative way of treating the problem of portfolio choice. Applying this approach, the investor chooses portfolio weights that are optimal with respect to her subjective belief about the true return distribution. Decision theory uses Bayesian rules in order to update the posterior distribution of the model parameters and, thus, takes into account the parameter uncertainty of the model. Consequently, a natural approach for the parameter uncertainty incorporation is the Bayesian approach, as pointed out in Barberis (2000). The main advantage of the Bayesian methods, as opposed to the frequentist approach, is the fact that the Bayesian analysis of a model summarizes the uncertainty for the estimation of the model parameters into the posterior distribution of the parameters given the data-returns. One of the most useful tools of the Bayesian analysis in the predictive distribution of future returns which is the distribution of the future returns given the current returns without conditioning on the fixed parameters (because they have been integrated out) of the model as is the case in the frequentist approach (see, for example Gelman, Carlin, Stern, and Rubin (2014) for a detailed introduction to Bayesian methods).

Considering estimation, one of the earliest attempts on treating the portfolio problem using the Bayesian methods is the approach of Barberis (2000), where the parameters uncertainty is accounted to construct optimal portfolios. In his paper, Barberis (2000) relies on the posterior distribution of the parameters of his VAR model, given the data, in order to construct the predictive distribution of the future returns. One of the main results in his paper is that in both a static buy-and-hold problem and a dynamic problem with optimal rebalancing, when parameter uncertainty is incorporated, the optimal allocation is significantly changed. In the case where parameter uncertainty is ignored, he points out that predictability in asset returns leads to strong horizon effects, something which is explained by the fact that time-variation in expected returns induces mean-reversion in returns, reducing the variance of cumulative returns over long horizons. This seems to be the reason why stocks appear less risky to long-horizon investors and leads them to allocate more to equities than investors with shorter horizons would.

One of the most recent approaches in the decision-theoretic literature of incorporating the parameter uncertainty in finding the optimal portfolio allocation strategy appears in the paper of Diris (2014). In his paper, Diris (2014), apart from the inclusion of the parameter uncertainty, undertakes the problem of incorporating model uncertainty in the optimal allocation of the assets. In general, the terminology of "model uncertainty" regards the problem of predictor variables which are to be included in a model. As noted by Aït-Sahalia and Brandt (2001), looking beyond expected returns, it is not clear which selection or combination of predictive variables the investor should focus on. Diris (2014) develops a method that includes model uncertainty w.r.t. asset returns and w.r.t. predictors. More precisely, he incorporates model uncertainty over long-term predictions using the Bayesian Model Averaging (BMA) and analyzes whether model uncertainty is present and relevant to investors with investment horizons up to 30 years. Most importantly, Diris (2014) concludes that the relevance of predictor variables depends on the horizon. For example, he notes that model uncertainty is important at short horizons, while he concludes that all variables are important at long horizons, either by predicting stock returns directly or the predictors of stock returns. Another interesting conclusion of his paper is that the incorporation of model uncertainty leads to an upward sloping term structure of risk (in crisis periods) and, finally, the incorporation of model uncertainty leads to a decreasing stock allocation when the investment horizon is sufficiently large. In the context of a comprehensive presentation of the relative literature, it should be noted that some early attempts of incorporating model uncertainty in the optimal portfolio allocation problem, using the Bayesian methods, are presented in the papers of Avramov (2002) and Cremers (2002).

This thesis aims to test the performance of a financial portfolio, when both parameter and variable selection uncertainty is taken into account. In the direction of the parameter uncertainty, the Bayesian approach is used in order to infer the parameters of the models through the posterior distribution of the parameters given the data-returns and to compute the predictive distribution of future returns by integrating out the parameters from the likelihood of the model. We use Markov Chain Monte Carlo (MCMC) methods and, in particular, the Gibbs sampler (see, for example Gelfand and Smith (1990) and Dellaportas and Roberts (2003) for a detailed presentation of MCMC methods and the Gibbs sampler) for the Bayesian estimation of the model used. Furthermore, the model parameters are estimated by using standard techniques from the Bayesian literature for the estimation of VAR models. The crucial step in developing an efficient MCMC algorithm for the Bayesian estimation of a VAR model is the choice of the prior distribution which summarizes the researcher's uncertainty over the model parameters. A well-known pitfall of VAR models is the danger of overparametrization. The early response in this danger was the use of shrinkage methods, like the so-called Minnesota prior (Doan, Litterman, and Sims (1984)), while more recent approaches attempt to solve this problem by applying variable selection priors or model selection priors.

Several approaches have been used in the literature to incorporate variable selection and model uncertainty. Avramov (2002) notes, among others, that incorporating model uncertainty in asset allocation decisions, through Bayesian methods, makes the allocation of the portfolio weights robust at least within the class of linear forecasting models. More precisely, Avramov (2002), Cremers (2002), Garratt, Koop, Mise, and Vahey (2012) and Diris (2014) use the Bayesian Model Averaging (BMA) technique to assign posterior probabilities to a wide set of competing return-generating models and then they use the probabilities as weights on the individual models to obtain a composite weighted model, for short and long run predictions respectively. BMA, in finance, models the investor's uncertainty for including different predictor variables in a forecasting model. An alternative way to incorporate model uncertainty in forecasting models through factor models, where a set of observed or unobserved factors are used to model the correlation between the series of log-returns, see Stock and Watson (2002a) and Stock and Watson (2002b) for use of factor models in a classical setting and Aguilar and West (2000) for a Bayesian perspective.

In this thesis we propose to incorporate model uncertainty in the problem of portfolio analysis, using an approach which can be seen as complementary to the above approaches. In the thesis we do not use model the uncertainty for including different predictor variables, but in the spirit of the work of George, Sun, and Ni (2008) and Villani (2009) we propose the use of a stochastic search algorithm to automatically select the variables included in a Bayesian Vector Autoregressive (BVAR) model. In particular, we use the Gibbs sampler proposed by Korobilis (2013). This Gibbs sampler does not rely on the estimation of all the VAR model combinations to decide which variables have to be included in the model, but decides for the more suitable variables with goodness of fit technique. This is achieved with the γ Bernoulli random variables, with which the most probable models are visited in stochastic manner, without using the less efficient method of estimating and enumerating all the possible models. Finally, the problem of variable selection in VAR models is common in the econometrician literature (see Villani (2009) for example), but this application of variable selection in the Gibbs sampler proposed by Korobilis (2013), is the main contribution of our thesis.

Korobilis (2013) introduces random variables (indicators), which indicate the variables that are significant for the VAR model in an automatic way, during the MCMC algorithm. Using this method, all possible VAR model combinations are avoided to be estimated, while this method is independent of the prior assumptions about the coefficients of the VAR model. In this sense, the contribution of the present thesis not only concerns the introduction of this new application for the portfolio allocation problem with a decision-theoretic approach, but in contrast to model uncertainty literature, the final posterior distribution of returns is a product of continuous dynamic algorithmic selection, in the sense that in each simulation the model updates, evaluates and selects.

In the empirical application, predictor variables are used from two different data-sets in order to test differences in the predictability of the model. The first data-set that is used arises from the approach of Stock and Watson (2002a). The major idea in Stock and Watson (2002a) is to create predictors that can capture the shifts and the co-movements of global financial and macroeconomic time series via principal component analysis. These predictors could explain several assets classes. The lags of assets classes by themselves do not have such a predictability. The main reason is that the range of the cross-correlated global financial and macroeconomic time series is quite wide and regards both financial time series such as a) stocks, bonds, interest rates, real estate prices, exchange rates, commodities and b) macroeconomics times series such as Inflation, consumer price index, GDP growth and unemployment rate. As long as economies of several countries are highly interacted, this range of cross-correlations does not have region-specific restrictions. Thus, Stock and Watson also take into account various Financial and Macroeconomic time series of several countries. Following Stock and Watson (2002a), the predictors are summarized by a dynamic factor model with VAR dynamics. The second data-set of predictor variables is the one that uses financial series and is proposed by Diris (2014), which is based on Welch and Goyal (2008). These series include several asset specific characteristics such as the dividend-to-price ratio, the book-to-market ratio, the price-earnings ratio, the detrended nominal yield, the yield spread, the credit spread, the ratio of 12-month moving sums of net issues by NYSE stocks and the total end-of-year market capitalization of these stocks and the stock return variance.

The main result of the present thesis is that, taking into account both parameter and variable uncertainty when using the algorithm of Korobilis (2013), it is found that incorporating variable-selection uncertainty increases investor's uncertainty and invests less to risky assets such as stocks and bonds. This result is, also, comparable with the main result of Diris (2014), where model uncertainty is taken into account. To be more precise, this phenomenon mainly appears if we use the data-set with financial series (cross-sectional premiums, Treasury Bills, Long Term Yield, Corporate Bond Returns, Inflation, investment to Capital Ratio etc.) as predictor variables, which is proposed by Diris (2014), based on Welch and Goyal (2008). Finally, for each one of the data-sets with predictors that is used, the matrix with the mean posterior selection-coefficients for including a predictor-factor in the model is provided. This is possible due to the selection to use the method of Korobilis (2013) for the (Bayesian) estimations and predictions.

The structure of the present thesis is as follows. In Section 2, the data that is to be used in the thesis is described. More precisely, the content of the constructed portfolios is presented, the data-sets of the factors that are to be used in the analysis, while there is also a description of the way that the data-sets are used in order to achieve the best possible analysis. Section 3 presents the main models that will be used throughout the present thesis, as well as how the model to perform the analysis is chosen. Section 3, presents the (Bayesian) methods that the estimation and predictions will be based on, too. Finally, Section 4 includes the main results and Section 5 concludes the thesis.

2 Data

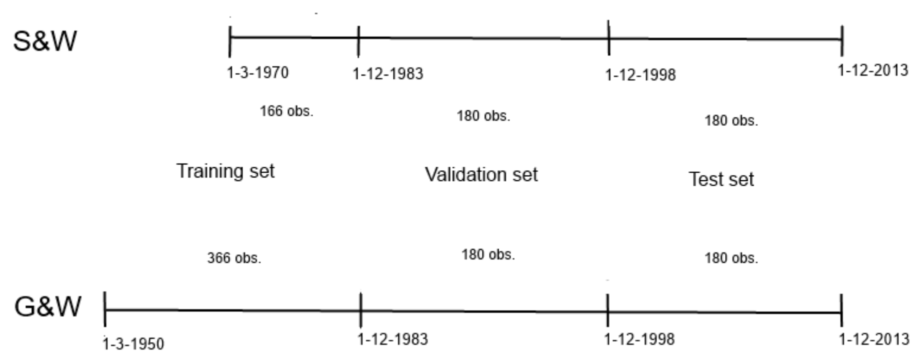
The constructed portfolios consists of monthly returns of a Treasury bill, one stock and one bond. We have three asset classes on the constructed portfolios. Firstly, our benchmark asset which will be the 3-Month real US Treasury bill log-return denoted as r_{1t} at time t, calculated by taking the difference of log return on the 3-month Treasury bill and the log inflation. Moreover, our portfolios consist of the log excess stock return of $SP\&500$ over the risk-free return denoted as r_{2t} and the log excess return of U.S. 10-Year Bond Yield over the risk-free return denoted as r_{3t}. From now on, we denote as r_t the $K_r \times 1$ vector (where here $K_r = 3$) of the three aforementioned time series returns at time t, where $r_t = (r_{1t}, r_{2t}, r_{3t})'$. Our goal is to built a dynamic factor model where those asset classes are explained by factors.

The first data-set of factors will be constructed, as Stock and Watson (2002b) proposed, from principal components analysis of macroeconomic and financial time series from the following countries: France, Germany, Italy, Japan, The Netherlands, Spain, The United States of America and The United Kingdom. The intuition behind the use of data from several countries is that we want to construct a data-set of factors that captures the the co-movements of the global economy in order to explain several asset returns and then a data-set of factors based on asset specific characteristics that could capture micro-movements and trends of the asset returns (which correspond to the second data-set of factors). In that sense we would also be able to test if there is bias of the performance of our several models, given the type of state variables. All data is retrieved from the OECD and FRED databases, combined with Bloomberg data on financial series.

The second data-set of factors is based on the factor data-set of Welch and Goyal (2008) and it is constructed as Diris (2014) proposed, where several variables from Welch and Goyal (2008) are used. The predictor variables will be assets' specific characteristics, which will be: the default risk premium, the log dividend-to-price ratio, the log book-to-market ratio, the log price-earnings ratio, the detrended log nominal yield, the yield spread, the credit spread, the ratio of 12-month moving sums of net issues by NYSE stocks and the total end-of-year market capitalization of these stocks and the log of the stock return variance. All data is retrieved from Professor Goyal's website.

The first data-set, which will be denoted as $S\&W$, starting from February 1970 until December 2013. The second data-set, which will be denoted as $G\&W$, starting from February 1950 until December 2013. The data will be split in three parts, where each subset will be used from a different scope. The purpose of this data-handling is to achieve a more objective view of how an investor could use and invest in reality, given this data. Thus, we assume that she uses a first part for model selection and a second part for parameters optimization. After that, the only part that is left is the investing part. In other words, the training set will be used at Subsection (3.1) for in-sample statistics. At Subsection (3.2.5), we use the Validation set and the Training set of data for hyper-parameters optimization. Finally at Section (4), we use the Training and Validation set and apply an expanding window for estimating the various models at the Test set, where we obtain and interpret the final results.

Figure 1: Data split



This figure presents the data split in 3 parts for both data-sets S&W (based on Stock and Watson (2002b)) and G&W (based on Welch and Goyal (2008)). The usage of 3 parts of data aims to achieve the most objective view of final results.

2.1 Stock and Watson factor construction

A set of global predictor variables based on Stock and Watson (2002b) is constructed which will be denoted as *S&W* factors, where the predictability of several asset returns as our asset classes was shown. We retrieve several Financial and Macroeconomic data such as: output, employment, exchange rates, interest rates, monetary aggregates, stock prices, price indices, exports and imports, housing-related series and commodity indices. The sampling is done for several countries such as: France, Germany, Italy, Japan, The Netherlands, Spain, The United States of America and The United Kingdom. It is worth mentioning that the contribution of The United States of America is almost by 50% to our database. The final database consists of 100 time series. More analytically, the list of series that are included in our data-set can be found on Appendix in Table 32. Since a large portion of the gathered time-series start around 1970, we only use series that are available from January 1970 through

December 2013.

After collecting the data, our scope is to reduce dimensionality and create factors that explain most of the variation in our Database via principal components Analysis. Firstly, some potential data problems we have to deal with are: non-Stationary of time series, the presence of outliers and missing observations. We are solving these problems in line with Stock and Watson (2002b). We transform the non-Stationary series to Stationary. A description of these transformations can be found on Appendix in Table 6.1. We define outliers as observations that are ten times greater in absolute value than the absolute value of the sample median. We set outliers to missing observations. Finally, we forecast the missing values by applying an EM-algorithm. The idea behind EM-algorithm to our set-up is to have an iterated and interactive process where, firstly, the missing data is approximated and, then, by using this approximation, we estimate the factors and so on. We repeat until the change in estimated factor values is below some threshold. A Mathematical formulation of this algorithm can be found on Appendix at 6.1.1.

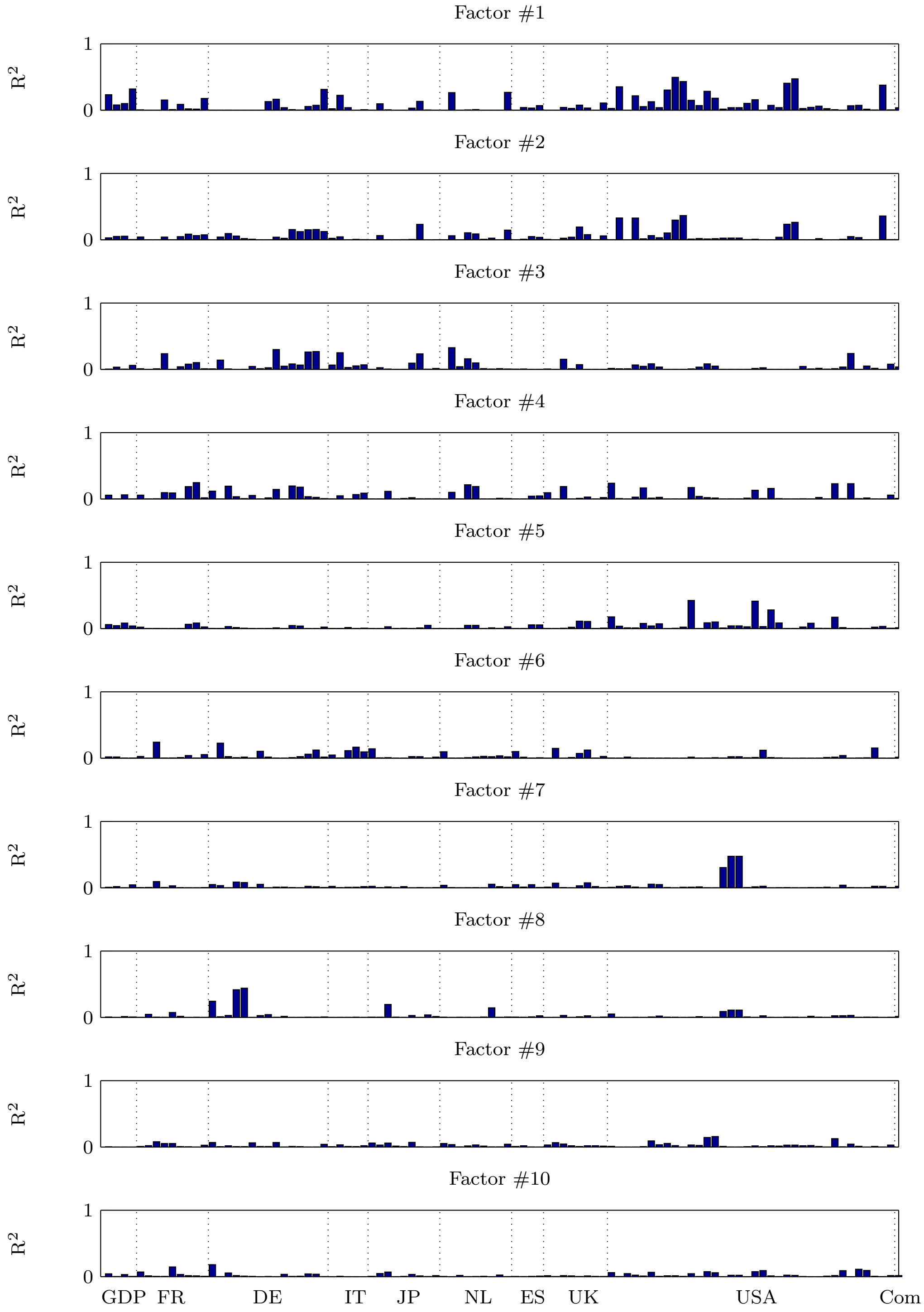
After dealing with the aforementioned problem, we are ready to perform principal component analysis. The question that remains to be answered is the number of factors to keep. One perspective of this issue regards the time period, which means that the optimal number of factors to keep might vary over time. Another perspective of this issue concerns the optimal selection of the initial series which would be used in each point in time to perform principal component analysis. However, these tasks will make the modeling implementation more complicated and time consuming. We choose to have a fixed number of factors in our model which will make our already complicated calculations more simple, without having any bias in our research question.

Finally, we base our selection of number of factors on a trade-off between the percentage of explained variation[1] and an empirical in-sample fit of the factor structure in full-sample set. For five factors the percentage explained is 30%. This increases to 42% when we extend the selection to ten factors. Moreover, we evaluate the R^2 obtained by regressing each of the macroeconomic series to each of the factors, as it can be seen at Figure 2. The first factor captures most of the variation of US economy, whereas further than 5 factors we can see that we do not obtain strong evidence of explanation. The percentage of explained variation is hardly increased when we double up the number of selected factors, whereas in order to achieve a percentage greater than 60%, we have to choose at least 15 factors. However, by using a big amount of factors, we lose a lot of efficiency in our models and, as a matter of fact,

[1] Calculated as $(\lambda_1 + \cdots + \lambda_k)/\sum_{i=1}^{N} \lambda_i$, with λ_i the eigenvalue of the respective factor

variable selection will tend to outperform against a model without variable selection which could easily lead to wrong conclusions. Therefore, it is outside the scope of this research to further investigate the optimal amount of factors' selection. Consequently, we prefer to use 5 factors, as we are able to explain quite a lot without having dimensionality problems.

Figure 2: R^2 of series regressed on the factors



Each of the 100 macroeconomic series is regressed upon each of the first 10 factors. The R^2 is then graphed. The macroeconomic series are grouped per country, one for GDP and one for commodities. (FR = France, DE = Germany, IT = Italy, JP = Japan, NL = Netherlands, ES = Spain, UK = United Kingdom, USA = United States, Com = Goldman Sachs commodity index)

2.2 Goyal and Welch factor construction

For Goyal and Welch factor construction, the computations are straightforward. We construct a set of predictor variables in line with Diris (2014), where several converted -and not- variables from Welch and Goyal (2008) were used. The data-set of factors consists of: the default risk premium (Defpr), the log dividend-to-price ratio (DP), the log book-to-market ratio (BM), the log price-earnings ratio (PE), the detrended log nominal yield (Ynom), the yield spread (Yspr), the credit spread (Crspr), the ratio of 12-month moving sums of net issues by NYSE stocks and the total end-of-year market capitalization of these stocks (ntis) and the log of the stock return variance (Var), which is calculated by Welch and Goyal (2008) as the summation over every quarter of the squared daily returns on the $S\&P500$. Below, we describe in a Mathematical formulation some definitions of our predictor variables, where following the same notation as Welch and Goyal (2008), we denote as tbl the Treasury Bill (Risk-Free Short Rate), as corpr the High-Quality Corporate Bond Rate, as lty the long-Term Treasury Yield, as D12 the $S\&P500$ dividend yield, as E12 the $S\&P500$ earnings yield, as b/m the book-to-market ratio, as AAA the High-Quality Corporate Bond Yield (usually long-term), as BAA the Mid-Quality Corporate Bond Yield (usually long-term) and as index the $S\&P500$ index.

- Defpr=corpr-lty
- DP=log(D12/Index)
- BM=log(b/m)
- PE=log(Index/E12)
- Ynom=log(1+tbl)-$MovingAverage_{12}$[log(1+tbl)] [2]
- Yspr=log(lty/tbl)
- Crspr=log(1+BAA)-log(1+AAA)

Finally in Table (1), we provide some summary statistics of the asset returns and the factors. We calculate the mean, the standard deviation, the minimum value, the maximum value and the AR(1) coefficient for each time series in the full sample.

[2] The abbreviation denotes the moving average of $log(1 + tbl)$ in a rolling window of 12 months

Table 1: Full-sample descriptive Statistics of returns, *S&W* factors and *G&W* factors

	mean	**std**	**min**	**max**	**AR(1)**
r_1	0,001	0,003	-0,011	0,019	0,561
r_2	0,004	0,045	-0,248	0,149	0,060
r_3	0,002	0,031	-0,119	0,135	0,042
factor(1)	-0,004	2,942	-18,074	10,080	0,485
factor(2)	-0,004	2,634	-8,423	14,099	-0,274
factor(3)	0,006	2,510	-13,814	7,067	0,442
factor(4)	-0,007	2,218	-6,706	7,539	0,695
factor(5)	0,008	1,871	-7,293	5,882	0,498
Defpr	0,0001	0,015	-0,098	0,074	-0,066
DP	-3,596	0,435	-4,524	-2,753	0,995
BM	-0,846	0,590	-2,116	0,188	0,994
PE	2,816	0,479	1,899	4,836	0,990
Ynom	-0,001	0,009	-0,037	0,037	0,899
Yspr	0,019	0,014	-0,032	0,044	0,951
Crspr	0,010	0,004	0,005	0,032	0,960
Ntis	0,010	0,020	-0,058	0,046	0,976
Var	-6,507	0,854	-8,484	-2,646	0,701

This table presents the descriptive Statistics (the mean, the standard deviation, the minimum value, the maximum value and the AR(1) coefficient for each time series) of our asset returns r_{1t}, r_{2t} and r_{3t}, the five factors obtained by principal component analysis of S&W (based on Stock and Watson (2002b)) and the nine factors of G&W (based on Welch and Goyal (2008)). The descriptive statistics of asset returns have been calculated in the time frame-work of S&W (starting from 1-2-1970, until 1-12-2013)

3 Model specification & Methodology

In this section, we are going to analyse the model specification that will be used for the rest of the analysis. We are going to select the model formulation, the amount of lags within our models and several input parameters for the Bayesian modeling. Moreover, we are going to fully analyse the methodology that will be used for the execution and the evaluation of our models. A contentious assumption that we apply here, is that we analyse and select among the different models for Bayesian applications, by evaluating several statistics in a Frequentist approach. However, the final model selection is quite strict with the factors' and lags' selection and thus we are not giving an easy mission to variable selection, and this does not -and should not- have a clear impact in our research.

In that section, all of our in-sample tests are performed in the Training set. The hyper-parameters' optimizations are implemented in the Training set and the Validation set, where we apply an expanding window for estimation of our various models. The extension starts from the beginning of the Validation set.

As a first step, we perform several tests to evaluate the fit of several factors. The crucial question in that step was whether we should include the lags of the returns as additional predictors, where we concluded to omit them from our factor data-set due to deterioration in the performance. Next, we test which amount of lags is well-fitted for the given models. Subsequently, we introduce our Bayesian methodology for the selected model. Lastly, having concluded to the exact formula of the models that will be used, we calculate the optimal hyper-parameters for the priors of the Bayesian approach used. Finally, after this section, we will be ready for kick off on our Test set, where we will test the Statistical and Financial out-of-sample performance of our models, by estimating the several variations by using an expanding window and forecasting each point in time of the Test set. That is, we apply an expanding window for the estimation of our models to obtain the results, which means that in our estimation, we include observations from the Training and Validation set.

3.1 Model specification - factor selection

Let r_t be the vector of our asset classes and F_t the vector of our factors from $S\&W$ or $G\&W$ at time t. We are firstly testing for the robustness of the following three models, where in the first case we assume that returns are autoregressive, in the second case we assume that returns are autoregressive but they are also explained by lags of factors and in the third case we assume that the returns are only explained by lagged factors:

$$r_t = \alpha + \Pi r_{t-1} + \xi_t \tag{3.1}$$

$$r_t = \alpha + \Pi r_{t-1} + \beta F_{t-1} + \xi_t \tag{3.2}$$

$$r_t = \alpha + \beta F_{t-1} + \xi_t \tag{3.3}$$

To begin with, we are going to have an in-sample evaluation and test which of the three models fit best in the training set. This evaluation will be based on BIC criterion, likelihood ratio tests and F-tests between those models. The models implied by the three above-mentioned equations will be referred as models A, B and C.

Bellow we provide the results of these tests for the $S\&W$ factors.

Table 2: Log-likelihood and BIC values of each model

	A	**B**	**C**
logL	1319.7	1380.3	1335.8
BIC	-2578.5	-2623.5	-2580.2

Statistics for Models of (3.1), (3.2) and (3.3) for S&W factors

Table 3: LR test between models A-B and B-C

	A-B	**B-C**
LR-stat	121.2	89.04
p-value	$<$0.001	$<$0.001

Statistics for Models of (3.1), (3.2) and (3.3) for S&W factors

Table 4: F test between models A-B and B-C

	r_1	r_2	r_3
		A-B	
F-stat	2.94	9.43	42.09
p-value	0.03	$<$0.001	$<$0.001
		B-C	
F-stat	9.55	0.34	9.06
p-value	$<$0.001	0.89	$<$0.001

Statistics for Models of (3.1), (3.2) and (3.3) for S&W factors. Note here that r_1, r_2 and r_3 are the three asset classes we are using and thus we calculate the statistics per time series.

For the second data-set of factors by $G\&W$ same tests are applied.

Table 5: Log-likelihood and BIC values of each model

	A	**B**	**C**
logL	1495.6	1518.6	1502.2
BIC	-2930.4	-2839.3	-2852.1

Statistics for Models of (3.1), (3.2) and (3.3) for G&W factors

Table 6: LR test between models A-B and B-C

	A-B	**B-C**
LR-stat	45.92	32.90
p-value	<0.001	<0.001

Statistics for Models of (3.1), (3.2) and (3.3) for G&W factors

Table 7: F test between models A-B and B-C

	r_1	r_2	r_3
	A-B		
F-stat	6.25	7.80	1.60
p-value	<0.001	<0.001	0.19
	B-C		
F-stat	3.25	0.10	0.25
p-value	0.0012	0.99	0.98

Statistics for Models of (3.1), (3.2) and (3.3) for G&W factors. Note here that r_1, r_2 and r_3 are the three asset classes we are using and thus we calculate the statistics per time series.

In tables (2) and (5), we see that model B -which is the unrestricted- has the highest log-likelihood. Moreover, in the same tables we see that the BIC values of model A indicate a better fit compared to the other two models. One potential explanation could be that there is high penalty for the number of parameters in models B and C. Continuing our analysis and focusing on the comparison of models A and B, we see that in tables (3), (6), (4) and (7) both the LR-test and F-test show us that the factors give significant additional explanatory power. In addition, if we see this result in a financial perspective, we would strongly agree that assets returns could not be explained by themselves (Lettau and Ludvigson (2001),Fama and French (1988)). We conclude that we are not taking into account model A in our final selection. Continuing with model selection analysis, we focus on the comparison of models B and C. Although in tables (3) and (6) LR-test does not reject the null hypothesis that the lags of the returns is an odd addition, we see in tables (4) and (7) that this is not the case, which is a first indication that lags of returns might be odd for the explanation of returns. On the other hand, apart from the F statistics we do not see any other statistic that indicate that we should prefer model C. However, as it has already mentioned our research question is whether variable selection will add value to our portfolio performance. We do not want to use additional insignificant factors that will give advantages to the performance of variable selection. Thus, we have no reason not to believe in model C. It looks that this set of factors contains a solid aggregation of the co-movement of the returns, whereas including the lags of the returns turns out to be a quite "heavy" choice.

A generalization of what have been already discussed could be described by the following model:

$$r_t = \alpha + \beta_1 F_{t-1} + ... + \beta_p F_{t-p} + \xi_t \tag{3.4}$$

$$F_t = \mu + \Phi_1 F_{t-1} + ... + \Phi_q F_{t-q} + u_t \tag{3.5}$$

Given the high persistence of factors which we could also see in the appendix at (6.2.1), a question that remains to be answered is the number of lags we should include in the model. We test the robustness of the model for different amounts of lags. In Table (8), we see that the lowest BIC value is achieved for $p = 1$ and $q = 2$ for the corresponding Equations (3.4) and (3.5) for $S\&W$ factors, whereas for $G\&W$ factors is achieved for $p = 1$ and $q = 1$. Furthermore, by monitoring the BIC values for further amount of lags at (6.2.2), we see that the optimum amount of lags remains the same. However, in most of the literature where common approaches are allocated, the choice of the simple VAR(1) is regularly preferred. In that sense, it would also be interesting to go in line with that "choice" of literature, so as to also have clearer comparisons. Last but not least, it should be pointed out that as long as the research question is whether variable selection should be included in parameter uncertainty approach, by adding many lags of factors in our model -which seems not to be well-fitted-, we will arbitrarily guide our results, as long as variable selection will tend to perform better when lots of close-to-insignificant factors are present. For these reasons, we choose to test the lags combinations of $p = 1$,$q = 1$ for both sets of factors.

Table 8: log-likelihood and BIC values of Equations (3.4) and (3.5) for several lags and for both sets of factors

S&W				
lags	logL-(3.4)	logL-(3.5)	BIC-(3.4)	BIC-(3.5)
1	1335,765	-1638.83	-2580,18	3429.92
2	1355,438	-1562.76	-2543,4	3404.66
3	1364,273	-1535.63	-2484,94	3477.28
4	1371,804	-1511.83	-2423,87	3556.54
G&W				
lags	logL-(3.4)	logL-(3.5)	BIC-(3.4)	BIC-(3.5)
1	1502.18	5158.08	-2852.11	-9859.40
2	1514.02	5294.20	-2738.75	-9720.55
3	1529.95	5365.26	-2633.59	-9451.58
4	1541.19	5422.25	-2519.03	-9154.47

This table presents the calculated Log-likelihood (logL) and Bayesian information criterion (BIC) statistics for several lags of Models based on (3.4) and (3.5) for S&W and G&W factors. The calculations of the statistics were done on the Training set.

Finally, after our conclusions for model and lag selection, we convert equations (3.4) and (3.5) as follows and we provide the OLS coefficients of the following model, estimated on the full data-set.

$$z_t = \begin{pmatrix} r_t \\ F_t \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \alpha \\ \mu \end{pmatrix}, \quad \mathbf{\Phi} = \begin{pmatrix} 0 & \beta_1 \\ 0 & \Phi_1 \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \xi_t \\ u_t \end{pmatrix} \tag{3.6}$$

$$z_t = \mathbf{C} + \mathbf{\Phi} z_{t-1} + \varepsilon_t \tag{3.7}$$

In the following tables, we provide the OLS coefficients for both set of factors. We correct for auto-correlation and heteroskedasticity on the residuals by calculating the heteroskedastic and autocorrelation consistent (HAC) standards errors and, finally, we calculate the p-values of t-statistic for each coefficient, where *, **, and *** stands for the significance level of < 0.1, < 0.05, and < 0.01.

Table 9: α and β_1 for S&W

	α	β_1				
	intercept	factor (1)	factor (2)	factor (3)	factor (4)	factor (5)
r_1	**0,000831**	**-0,00011**	**-0,00021**	**-3,43E-06**	**-0,00015**	**0,000591**
	0,000238	9,12E-05	6,45E-05	0,000103	7,88E-05	0,000112
	***		***		*	***
r_2	**0,004249**	**-0,0012**	**0,000162**	**5,31E-03**	**-0,00381**	**0,000279**
	0,001748	6,44E-04	7,40E-04	0,000906	9,24E-04	0,000973
	**	*		***	***	
r_3	**0,002677**	**-0,00446**	**-0,00287**	**2,21E-03**	**0,00014**	**0,001414**
	0,000884	5,25E-04	5,87E-04	0,000606	4,59E-04	0,000578
	***	***	***	***		**

Calculated OLS coefficients α and β_1 for Equations (3.6) and (3.7) for S&W factors on the Training set.

Table 10: μ and Φ_1 for S&W

	μ	Φ_1				
	intercept	factor (1)	factor (2)	factor (3)	factor (4)	factor (5)
factor (1)	**-0,00528**	**0,487001**	**0,407367**	**0,241778**	**-0,11502**	**0,239597**
	0,097858	0,033067	0,037576	0,03885	0,04423	0,052211
		***	***	***	***	***
factor (2)	**-0,00223**	**0,244529**	**-0,26728**	**0,070213**	**-0,04505**	**0,077141**
	0,105418	0,035622	0,040479	0,041852	0,047647	0,056245
		***	***	*		
factor (3)	**0,003006**	**0,107914**	**-0,00503**	**0,445846**	**-0,2821**	**0,065221**
	0,094489	0,031929	0,036283	0,037512	0,042707	0,050413
		***		***	***	
factor (4)	**-0,01659**	**0,048167**	**-0,00693**	**-0,16965**	**0,697338**	**0,31605**
	0,06124	0,020694	0,023515	0,024313	0,027679	0,032674
		**		***	***	***
factor (5)	**0,008909**	**0,094362**	**-0,03118**	**0,092081**	**0,228194**	**0,501137**
	0,065835	0,022247	0,02528	0,026137	0,029756	0,035126
		***		***	***	***

Calculated OLS coefficients μ and Φ_1 for Equations (3.6) and (3.7) for S&W factors on the Training set.

Table 11: α and β_1 for G&W

	α	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
	0.022546	**-0.03599**	**0.005766**	**-0.006**	**-0.00231**	**-0.03235**	**-0.03165**	**0.181577**	**0.014376**	**0.000148**
r_1	0.005179	0.010101	0.001222	0.001175	0.001115	0.028755	0.021385	0.097835	0.016254	0.000297
	***	***	***	***	**			*		
	0.075688	**0.258017**	**0.024188**	**-0.02682**	**-0.01346**	**-0.29395**	**0.096122**	**0.786412**	**0.080459**	**-0.00304**
r_2	0.045472	0.203644	0.011384	0.010068	0.009555	0.273355	0.183287	0.894348	0.117144	0.003247
	*		**	***						
	0.038989	**0.005923**	**-0.00196**	**0.002246**	**-0.00541**	**0.196604**	**0.428706**	**-0.62775**	**-0.13024**	**0.004107**
r_3	0.032315	0.1245	0.008846	0.008081	0.005731	0.228656	0.132601	0.623269	0.083374	0.001915
							***			**

Calculated OLS coefficients α and β_1 for Equations (3.6) and (3.7) for G&W factors on the Training set.

Table 12: μ and Φ_1 for G&W

	μ	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
	-0.00513	**-0.10456**	**0.006271**	**-0.00326**	**0.004858**	**0.019081**	**-0.03131**	**0.623031**	**0.047659**	**-0.00077**
Defpr	0.0161	0.044446	0.004289	0.003922	0.002767	0.087097	0.058183	0.253955	0.038651	0.000887
		**			*			**		
	-0.09552	**-0.26214**	**0.960323**	**0.032904**	**0.002586**	**0.366456**	**0.082195**	**-1.04559**	**-0.15858**	**0.002603**
DP	0.048592	0.134143	0.012946	0.011838	0.008351	0.262868	0.175601	0.766461	0.116651	0.002676
	**	*	***	***						
	-0.00664	**-0.05141**	**-0.01199**	**0.991641**	**-0.01608**	**0.593594**	**0.364901**	**0.175695**	**0.110023**	**0.001464**
BM	0.068992	0.190459	0.018381	0.016808	0.011858	0.373226	0.249322	1.088239	0.165624	0.003799
				***						
	0.186304	**-0.19521**	**0.03236**	**-0.06082**	**0.953448**	**-1.62992**	**-0.69969**	**4.506753**	**-0.08194**	**0.003493**
PE	0.06996	0.193132	0.018638	0.017043	0.012024	0.378464	0.252821	1.103511	0.167949	0.003852
	***		*	***	***	***	***	***		
	-0.00734	**0.011516**	**-0.00159**	**0.001758**	**0.000425**	**0.925997**	**0.060256**	**-0.08027**	**0.003599**	**-0.00023**
Ynom	0.004323	0.011933	0.001152	0.001053	0.000743	0.023385	0.015621	0.068184	0.010377	0.000238
	*			*		***	***			
	0.001881	**-0.01685**	**0.001755**	**-0.00193**	**0.000479**	**-0.06701**	**0.909365**	**0.190544**	**0.013684**	**-0.00017**
Yspr	0.00461	0.012726	0.001228	0.001123	0.000792	0.024939	0.01666	0.072716	0.011067	0.000254
				*		***	***	***		
	0.005028	**-0.02144**	**0.000182**	**6.2E-06**	**-0.00038**	**0.002459**	**-0.00121**	**0.923725**	**-0.00432**	**0.000377**
Crspr	0.001126	0.003107	0.0003	0.000274	0.000193	0.006089	0.004068	0.017755	0.002702	6.2E-05
	***	***			*			***		***
	-0.00532	**0.016984**	**0.001372**	**0.000469**	**0.002588**	**-0.04127**	**-0.0044**	**-0.01047**	**0.972893**	**-0.00058**
Ntis	0.004571	0.01262	0.001218	0.001114	0.000786	0.02473	0.01652	0.072106	0.010974	0.000252
					***	*			***	**
	-3.74146	**-5.56089**	**-0.18377**	**0.040482**	**0.143951**	**-0.37987**	**-3.324**	**24.733**	**-2.36451**	**0.609117**
Var	0.646985	1.786074	0.172367	0.157617	0.111197	3.500009	2.338071	10.20519	1.553176	0.035627
	***	***						**		***

Calculated OLS coefficients μ and Φ_1 for Equations (3.6) and (3.7) for G&W factors on the Training set.

3.2 BVAR preliminaries

In this subsection, we are going to describe the BVAR process. We had from Equations (3.4) and (3.5) that:

$$r_t = \alpha + \beta_1 F_{t-1} + ... + \beta_p F_{t-p} + \xi_t$$

$$F_t = \mu + \Phi_1 F_{t-1} + ... + \Phi_p F_{t-q} + u_t$$

The investor wants to calculate moments of r_{t+h} in order to allocate her wealth. Below, we derive the analytical formulas of these moments. In order to calculate $E_t(r_{t+h})$ and $Var_t(r_{t+h})$, it would be helpful to convert the two equations of (3.4) and (3.5) and interpolate them both in one auto-regressive process, where the returns r_t and the factors F_t are simultaneously explained. Thus, given the matrix of parameters B we have:

$$B = \begin{pmatrix} \alpha & \beta_1 & \beta_2 & \cdots & \beta_p \\ \mu & \Phi_1 & \Phi_2 & \cdots & \Phi_q \end{pmatrix} \tag{3.8}$$

$$z_t = \begin{pmatrix} r_t \\ F_t \\ F_{t-1} \\ \vdots \\ F_{t-max(p,q)+1} \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \alpha \\ \mu \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{\Phi} = \begin{pmatrix} 0 & \beta_1 & \beta_2 & \cdots & \beta_p \\ 0 & \Phi_1 & \Phi_2 & \cdots & \Phi_q \\ 0 & I & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & I & 0 \end{pmatrix}, \quad \upsilon_t = \begin{pmatrix} \xi_t \\ u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{3.9}$$

Given these definitions, we are now ready to describe the process in one single equation, which will make easier all future calculations:

$$z_t = \mathbf{C} + \mathbf{\Phi} z_{t-1} + \upsilon_t \qquad \upsilon_t \sim N(0, Q) \qquad Q = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{\xi\xi} & \Sigma_{\xi u} \\ \Sigma_{u\xi} & \Sigma_{uu} \end{pmatrix} \tag{3.10}$$

Here, Q is the covariance matrix of υ_t and Σ the covariance matrix of the vector that contains both ξ_t and u_t. Finally, we derive the moments for portfolio construction problem following Campbell and Viceira (2005):

$$E_t(z_{t+h}) = \left[\sum_{i=0}^{h-1} \mathbf{\Phi}^i\right]\mathbf{C} + \left[\sum_{j=1}^{h} \mathbf{\Phi}^j\right] z_t \tag{3.11}$$

$$Var_t(z_{t+h}) = (I + \mathbf{\Phi} + \cdots + \mathbf{\Phi}^{h-1})\Sigma(I + \mathbf{\Phi} + \cdots + \mathbf{\Phi}^{h-1})'$$

It is now obvious how to take moments for r_{t+h}, whereas if someone sets $y_t = (r_t, F_t)$, she can easily take the expectations $E_t(y_{t+h})$ from $E_t(z_{t+h})$ and $Var_t(y_{t+h})$ from $Var_t(z_{t+h})$.

However, in order to describe the BVAR process we transform Equations (3.4) and (3.5) as:

$$y_t = \begin{pmatrix} r_t \\ F_t \end{pmatrix}, \quad B = \begin{pmatrix} \alpha & \beta_1 & \cdots & \beta_p \\ \mu & \Phi_1 & \cdots & \Phi_q \end{pmatrix}, \quad Z_t = \begin{pmatrix} 1 \\ F_{t-1} \\ \vdots \\ F_{t-max(p,q)} \end{pmatrix} \quad \varepsilon_t = \begin{pmatrix} \xi_t \\ u_t \end{pmatrix} \tag{3.12}$$

$$y_t = BZ_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma) \tag{3.13}$$

Let us assume that y_t is an $m \times 1$ vector of T times series observations and Σ the $m \times m$ covariance matrix of ε_t. Finally, we rewrite the above equation to a seemingly unrelated regressions (SUR), where n is the total number or parameters of matrix B and β is an $n \times 1$ vector and the symbol " $\otimes$ " is the Kronecker product :

$$\beta = vec(B'), \quad z_t = I_m \otimes Z_t \tag{3.14}$$

$$y_t = z_t\beta + \varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma) \tag{3.15}$$

The simple BVAR version of our variations is based on Equation (3.15). However, the BVAR with variable selection (BVAR-VS) is selecting which variables to include in the model. This selection is indirectly undergone, multiplying each element of β with the corresponding element of γ, which is also an $n \times 1$ vector. Each element of γ could either equal to zero or one. In that sense, the selection is forced. We define the symbol "$\odot$" as the Hadamard product.The BVAR-VS listens to:

$$y_t = z_t\theta + \varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma) \qquad \theta = \gamma \odot \beta \tag{3.16}$$

We are now ready to describe the BVAR procedure, based on the models of Equation

(3.15) and (3.16). Firstly, we will describe the prior choice. Moreover, we will show how a BVAR Gibbs sampler model draws from the posterior density of $y_{t+1}, y_{t+2}, .., y_{t+h}$ and forecasts h periods ahead and how Korobilis BVAR Gibbs sampler with variable selection model differentiate from this. Please note that having the posterior of y_{t+1},y_{t+2},...,y_{t+h}, we can isolate and use the MCMC simulations of r_{t+1},r_{t+2},...,r_{t+h}, which means that having the first posterior density, we also have the one of the returns.

3.2.1 Prior

We are going to use two kinds of priors for our parameters. We present the priors that regard the parameters in both BVAR and the Bayesian VAR with variable selection (BVAR-VS). Firstly the Ridge Regression prior, which could be written in the form: $\overline{\beta} = 0$ with prior parameter covariance matrix $V = \lambda_1 \times I$, where I is the $n \times n$ identical matrix. Secondly the Minnesota prior which could be written as: $\overline{\beta_i} = 1$ if the i^{th} parameter correspond to the first own lag and $\overline{\beta_i} = 0$ otherwise. For the prior parameter covariance matrix we transform properly the elements of the Vm at (3.17) to a diagonal $\overline{V}$ $n \times n$ matrix, where each diagonal element corresponds to the prior variance of each parameter. We denote as s_j^2 the variance of the residuals of the AR(p) model of the j^{th} time series and as r the corresponding lag of j^{th} parameter.[3] Following we present the formula of prior covariance for each element B'_{ij} of (3.12):

$$Vm_{ij}{}^{r} = \begin{matrix} 100{s_i}^2 & if & intercept \\ \lambda_2/r^2 & if & i=j \\ \lambda_3\frac{{s_i}^2}{r^2{s_j}^2} & if & i\neq \text{j} \end{matrix} \tag{3.17}$$

Moreover, we should note that the use of Minnesota prior is mainly done because it gives a more "fair" chance for less restrictions when variable selection is applied compared to the Ridge OLS prior. As it was indicated in Korobilis (2013), Minnesota prior allows for the diagonal parameters to fluctuate and shrinks more the non-diagonal parameters to zero compared to the Ridge one and, as a matter of fact, this would highlight a more clear "winner" in the comparison with the variable selection. In contrast, the Ridge OLS prior will allow for more restrictions to be efficient when variable selection is applied -especially when the number of parameters is getting bigger-, because of the fact that this prior shrinks all parameters in the same magnitude.

Furthermore, in Equation (3.15), we presented a way that the processes of Equations (3.4) and (3.5) could perform interactively in a single Equation. However, the prior use

[3] in our case $r = 1$, we do not use further lags

and the hyper-parameter optimization of β will be done separately for the parameters that correspond to (3.4) and (3.5). The reason is that the nature of the two Equations is different and, in this way, we provide more efficient results. However, for a single horizon ($h = 1$) we do not "need" Equation (3.5) and, thus, we do not divide our models for the priors that correspond to that Equation. All in all, we set the Ridge Regression prior to all of the parameters that correspond to Equation (3.4) and we never apply a Minnesota prior for those parameters, whereas for those that correspond to Equation (3.5), we apply both the Ridge Regression prior and the Minnesota prior. The reason that we do not set a Minnesota prior to the parameters that correspond to equation (3.4) is that this model is not Autoregressive.

Moreover, our prior beliefs concerns the prior covariance matrix S_p which is set to the OLS covariance matrix of residuals ε_t of the corresponding rolling window and the prior degrees of freedom v_p which is set to the length of the vector y_t ($v_p = m$).

Moreover, when we apply variable selection, we also interpolate for our prior beliefs about the exclusion of each parameter β_i. This is the prior likelihood for the i^{th} parameter β. We denote this parameter as $\overline{\pi_0}$, which is the $n \times 1$ vector, where $\overline{\pi_{0k}}$ is the prior probability that the k^{th} parameter will be included in the final model. Following Korobilis (2013), the prior $\overline{\pi_{0k}}$ does not need to be separately set for each of our parameters. In that sense, we could set a a value $\overline{\pi_0}$ for all of our parameters. However, similarly to our prior use of β and $\overline{V}$, we split our prior beliefs about $\overline{\pi_{0i}}$ separately depending on which Equation the i^{th} parameter corresponds to, as already analysed. For example, it might be the case that for (3.4) we need to restrict less parameters than (3.5). One additional important implementation of Korobilis (2013) was not to restrict parameters β_c that are intercepts, which we also implement. In practice, this could be done either by simply not running the variable selection algorithm solely for these parameters, or by setting $\pi_{0j} = 1$ when $j = c$. We modify our algorithm to run over the selection on non-intercepts.

All in all, we have various models, where the standard set-up that we have selected regards five different horizons ($h = 1, 4, 6, 12, 24$) and one kind of prior that corresponds to Equation (3.4), which is the Ridge OLS. The variations of our models for a given selected horizon and a given selected set of factors regards whether we apply variable selection or we use Ridge OLS prior for Equation (3.5) or Minnesota. Thus, for each horizon we have four different types of models, as shown in the following table.

Table 13: Possible combinations of models for given horizon

Models	Variable Selection	Ridge prior for (3.4)	Ridge prior for (3.5)	Minnesota prior for (3.5)
BVARR	0	1	1	0
BVARRVS	1	1	1	0
BVARM	0	1	0	1
BVARMVS	1	1	0	1

This table defines the models $M_{(\cdot)}$ by taking into account all the possible combinations of applying variable selection and applying Minnesota prior on Equation (3.5). The values of 0 and 1 stand for the fact of applying of not the function of the corresponding column for the model.

3.2.2 BVAR Gibbs sampler process

The joint posterior predictive distribution of $y_{t+1}, y_{t+2}, .., y_{t+h}$ follows from:

$$p(y_{t+1}, y_{t+2}, .., y_{t+h}|Y) = \int p(y_{t+1}, y_{t+2}, .., y_{t+h}|Y, \beta, \Sigma)p(\beta, \Sigma|Y)d\beta d\Sigma =$$
$$\iint_{\beta,\Sigma} p(y_{t+1}, y_{t+2}, .., y_{t+h}|Y, \beta, \Sigma)p(\beta|\Sigma, Y)p(\Sigma|Y)d\beta d\Sigma$$

In order to approximate this integration, we apply Gibbs sampler. This technique is well-known in the literature of Bayesian Econometrics techniques, where it is extensively shown in Gelfand and Smith (1990) and Dellaportas and Roberts (2003). Thus, in order to draw from the joint posterior density of $y_{t+1}, y_{t+2}, .., y_{t+h}$, we follow the Monte Carlo Integration, given the prior parameters which are: $\overline{\beta}$, which is the $n \times 1$ vector that corresponds to the prior belief of β, $\overline{V}$ which is the $n \times n$ prior parameter covariance matrix and finally S_p and v_p which are the prior covariance matrix of ε_t and the prior degrees of freedom. We describe the i^{th} iteration of the process.

- Draw $\tilde{\beta}^i$ from the posterior density:

$$\beta|\Sigma, Y \sim N(\tilde{\beta}, \tilde{V})$$

  where $\tilde{V} = (\overline{V}^{-1} + \sum_{t=1}^{T} z_t \Sigma^{-1} z_t)^{-1}$ and $\tilde{\beta} = \tilde{V}(\overline{V}^{-1}\overline{\beta} + \sum_{t=1}^{T} z_t \Sigma^{-1} y_t)$

- draw $\tilde{\Sigma}^i$ from the posterior density:

$$\Sigma|\beta, Y \sim inverted \quad Wishart(S_p + S, v_p + T)$$

where $S = \sum_{t=1}^{T}(y_t - z_t\tilde{\beta}^i)'(y_t - z_t\tilde{\beta}^i)$

- reshape $\tilde{\beta}^i$ to $\tilde{B}^i$ according to the definition of β in equation (3.14).
- Draw $\tilde{y_{t+h}}^i$ from the posterior density:

$$y_{t+h}|\beta, \Sigma, Y \sim N(\mu^i, \Omega^i)$$

where μ^i and Ω^i are the $E_t(\tilde{y_{t+h}}^i)$ and $Var_t(\tilde{y_{t+h}}^i)$ and taken from the forecast procedure as described at (3.8),(3.9),(3.10) and (3.11) by setting $B = \tilde{B}^i$ and $\Sigma = \tilde{\Sigma}^i$.

- repeat the process for N times and approximate $E_t(y_{t+h}) \approx \frac{1}{N}\sum_{i=1}^{N}\tilde{y_{t+h}}^i$

3.2.3 BVAR Gibbs sampler with variable selection

The joint posterior predictive distribution of $y_{t+1}, y_{t+2}, .., y_{t+h}$ follows from:

$$p(y_{t+1}, y_{t+2}, .., y_{t+h}|Y) = \int p(y_{t+1}, y_{t+2}, .., y_{t+h}|Y, \beta, \theta, \Sigma)p(\beta, \theta, \Sigma|Y)d\beta d\theta d\Sigma =$$

$$\iiint_{\beta,\theta,\Sigma} p(y_{t+1}, y_{t+2}, .., y_{t+h}|Y, \beta, \theta, \Sigma)p(\beta|\theta, \Sigma, Y)p(\theta|\Sigma, Y)p(\Sigma|Y)d\beta d\theta d\Sigma$$

Thus, in order to draw from the joint posterior density of $y_{t+1}, y_{t+2}, .., y_{t+h}$ we follow the Monte Carlo Integration, given the prior parameters which are: $\overline{\beta}$, $\overline{V}$, S_p, v_p, $\overline{\pi_0}$ which is the $n \times 1$ vector, where $\overline{\pi_{0k}}$ is the prior probability that the k^{th} parameter will be included in the final model. We describe the i^{th} iteration of the process.

- Draw $\tilde{\beta}^i$ from the posterior density:

$$\beta|\gamma, \Sigma, Y \sim N(\tilde{\beta}, \tilde{V})$$

where $\tilde{V} = (\overline{V}^{-1} + \sum_{t=1}^{T} z_t^* \Sigma^{-1} z_t^*)^{-1}$, $\tilde{\beta} = \tilde{V}(\overline{V}^{-1}\overline{\beta} + \sum_{t=1}^{T} z_t^* \Sigma^{-1} y_t)$ and $z_t^* = z_t \tilde{\Gamma^i}$, where $\tilde{\Gamma^i}$ is the $n \times n$ matrix that is equal to zero everywhere except from the diagonal elements where : $\tilde{\Gamma^i}_{jj} = \tilde{\gamma^i}_j$

- Draw $\tilde{\gamma^i}_j, j = 1, ..., n$ from the posterior density and j is taken in random order:

$$\gamma_j|\gamma_{|-j}, \beta, \Sigma, Y \sim Bernoulli(\tilde{\pi_j})$$

where $\tilde{\pi_j} = \frac{l_{0j}}{l_{0j}+l_{1j}}$ and:

Define $\theta^* = \theta$, but set the j^{th} parameter $\theta_j = \beta_j$ and
$\theta^{**} = \theta$, but set the j^{th} parameter $\theta_j = 0$

$l_{0j} = log[p(y|\theta_j.\gamma_{|-j}, \gamma_j = 1)]\pi_{0j}$
Thus $l_{0j} \propto \Big(-\frac{1}{2}\sum_{t=1}^{T}(y_t - z_t\theta^*)'\Sigma^{-1}(y_t - z_t\theta^*)\Big)\pi_{0j}$
$l_{1j} = log[p(y|\theta_j.\gamma_{|-j}, \gamma_j = 0)](1 - \pi_{0j})$
Thus $l_{1j} \propto \Big(-\frac{1}{2}\sum_{t=2}^{T}(y_t - z_t\theta^{**})'\Sigma^{-1}(y_t - z_t\theta^{**})\Big)(1 - \pi_{0j})$

Here, following Diris (2014), we differentiate from Korobilis (2013) and we use log Marginal likelihood instead of Marginal likelihood in our procedure. The log-likelihood itself is quite big in our set-up and, thus, might cause numerical issues in the calculation of Marginal likelihood. Moreover, note here that the the exact l_{oj} and l_{1j} are proportional to the formula given, because we ignore the integrating constant.

- Set $\tilde{\theta^i} = \tilde{\gamma^i} \odot \tilde{\beta^i}$.
- draw $\tilde{\Sigma^i}$ from the posterior density:

$$\Sigma|\beta, Y \sim inverted \quad Wishart(S_p + S, v_p + T)$$

where $S = \sum_{t=1}^{T}(y_t - z_t\theta)'(y_t - z_t\theta)$

- reshape $\tilde{\beta^i}$ to $\tilde{B^i}$ according to the definition of β in equation (3.14) and similarly reshape θ^i to $\tilde{\Theta^i}$.
- Draw $\tilde{y_{t+h}}^i$ from the posterior density:

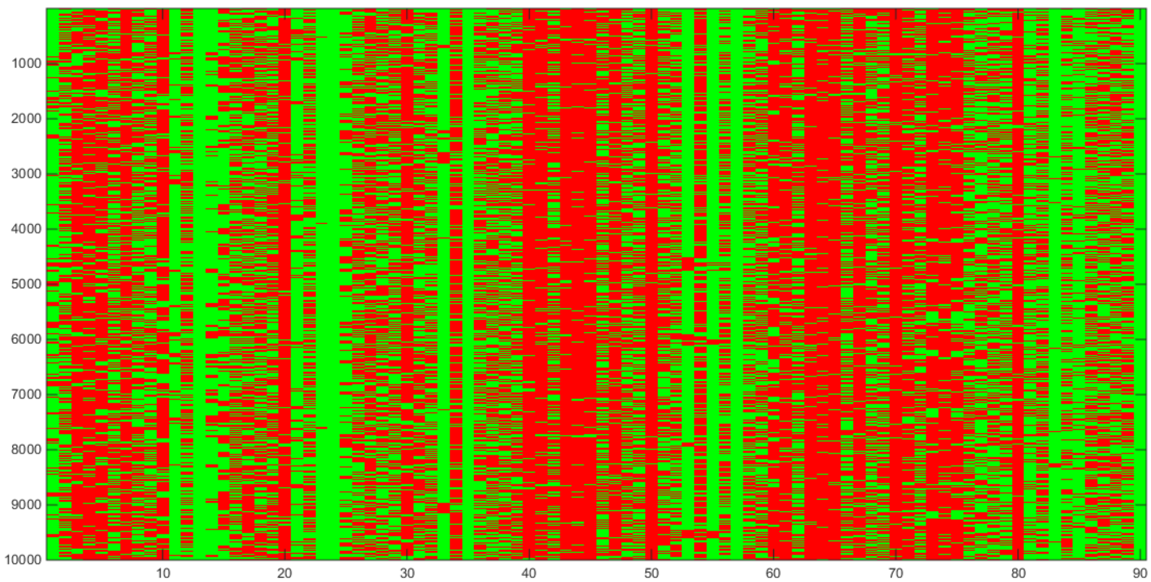
$$y_{t+h}|\beta, \gamma, \Sigma, Y \sim N(\mu^i, \Omega^i)$$

where μ^i and Ω^i are the $E_t(\tilde{y_{t+h}}^i)$ and $Var_t(\tilde{y_{t+h}}^i)$ and taken from the forecast procedure as described at (3.8),(3.9),(3.10) and (3.11) by setting $B = \tilde{\Theta^i}$ and $\Sigma = \tilde{\Sigma^i}$.

- repeat the process for N times and approximate $E_t(y_{t+h}) = \frac{1}{N}\sum_{i=1}^{N}\tilde{y_{t+h}}^i$

The reader might find helpful the following visualization of draws of γ as we see in Figure (3). In that figure we see the behavior of variable selection. We clearly see that there are some variables that our algorithm almost always select to include, others that we almost always exclude and some that switch over and over again.

Figure 3: Draws of γ



This Figure plots the results of 10,000 simulations of drawn coefficients. We perform a BVAR-VS(1) of (3.5) with $q = 1$ (only for the factors), including 90 parameters and we set for all of our parameters $\overline{\pi_{0i}} = 0.5$. For each of the 10,000 simulations that we can see in vertical axis, we monitor the draws of γ on horizontal axis and we set green $= 1$ and red $= 0$.

3.2.4 BVAR optimal weights

Let $\boldsymbol{w(t_h)}$ be the $K_r \times 1$ vector (where here $K_r = 3$ as long as we have 3 assets), such that the k^{th} element of that vector is the optimal proportion of wealth that the investor allocates to the k^{th} asset at time t and expects to maximize the power Utility function at time $t + h$. Let us firstly assume the following:

$U(\cdot)$ the power utility function of the investor where $U(x) = (1/\gamma)x^{1-\gamma}$.

$r_{\cdot t \to t+h} = \sum_{j=1}^{K} r_{\cdot t+j}$.

Let $cgr_2(t_h) = (r_{2t \to t+h} + r_{1t \to t+h})$ $cgr_3(t_h) = (r_{3t \to t+h} + r_{1t \to t+h})$ and $cgr_1(t_h) = r_{1t \to t+h}$

Then the cumulative Gross log-returns will be: $cgr(t_h) = \begin{pmatrix} cgr_1(t_h) \\ cgr_2(t_h) \\ cgr_3(t_h) \end{pmatrix}$

The cumulative wealth: $W_{t\to t+h} = W_t \times [\boldsymbol{w(t_h)}' exp(cgr(t_h))]$

In order to obtain the BVAR optimal weights for the buy-and-hold investor, we are solving the following problem:

$\max_{\boldsymbol{w(t_h)}} E_t[U(W_{t\to t+h})]$

$\max_{\boldsymbol{w(t_h)}} \iiint_{\beta,\Sigma,cgr(t_h)} U(W_{t\to t+h}) p(cgr(t_h)|\beta,\Sigma,Y) p(\beta|Y) p(\Sigma|\beta,\gamma,Y) d(\beta) d(\Sigma) d(cgr(t_h))$

Whereas, in order to obtain the BVAR optimal weights with variable selection for the buy-and-hold investor, we are solving the following problem:

$\max_{\boldsymbol{w(t_h)}} E_t[U(W_{t\to t+h})]$

$\max_{\boldsymbol{w(t_h)}} \iiiint_{\beta,\Sigma,\theta,cgr(t_h)} U(\cdot) p(cgr(t_h)|\beta,\Sigma,\theta,Y) p(\beta|Y) p(\theta|\beta,Y) p(\Sigma|\beta,\theta,Y) d(\beta) d(\Sigma) d(\theta) d(cgr(t_h))$

In practice, the above equations mean that the optimal weights $\boldsymbol{w(t_h)}$ should be obtained as the vector that maximizes the expected Utility at the simulated MCMC paths of the joint posterior of r_{t+1},r_{t+2},...,r_{t+h}. At subsection (3.2.3), we presented how to draw from this density in both cases. Thus, we choose the weight $\boldsymbol{w(t_h)}$, so that it maximizes the expected power Utility function h periods ahead and no sort selling is allowed, which in mathematical form could be written as: $0 \leq \boldsymbol{w(t_h)} \leq 1$. Thus, our utility function is:

$$U(W_{t\to t+h}) = E_t\left[\frac{[\boldsymbol{w(t_h)}' exp(cgr(t_h))]^{1-\gamma}}{1-\gamma}\right]$$

A proxy of this expectation could be the following:

$$U(W_{t\rightarrow t+h}) \approx \frac{1}{N}\sum_{i=1}^{N}\left(\frac{[\boldsymbol{w(t_h)}'exp(cgr(t_h))]^{1-\gamma}}{1-\gamma}\right)$$

We simulate grid points, as possible values of $\boldsymbol{w(t_h)}$. Thus, we simulate values on the linespace between 0 and 1 with a particular step and we take into account all possible combinations for our elements of our vector $\boldsymbol{w(t_h)}$ and we calculate realized utility for each of our simulated paths as:

$$U^i{}_w = \frac{[\boldsymbol{w(t_h)}'exp(cgr^i(t_h))]^{1-\gamma}}{1-\gamma}$$

We approximate conditional expected utility for each combination of grid points w as:

$$CU_w \approx \frac{1}{N}\sum_{i=1}^{N} U^i{}_w$$

Finally, we select $\boldsymbol{w(t_h)}$ that maximizes the conditional expected utility for h periods ahead. That will be the optimal portfolio weights for a buy-and-hold investor who invests h periods ahead, whereas for the myopic investor who invest in a single period ahead, we can obtain the optimal weights by setting $h = 1$.

3.2.5 Parameters and inputs handling on BVAR variations

We are now going to analyse the prior hyper-parameter optimization. This analysis will be done for each horizon that we are interested in. Thus, for each horizon we simulate different grid points from our hyper-parameters and we calculate for which combination of simulated values we obtain the minimum average [4] mean squared prediction errors of $cgr(t_h)$

.

Finally, we perform the analysis that has been described for the prior parameter optimization for our several models by simulating hyper-parameters from the corresponding grid points of table (14) for each model and each horizon.

[4]note here that $cgr(t_h)$ is a vector. We calculate the MSPE for each element and we optimize for the average of those MSPE's

Table 14: Allowance of hyper-parameters' grids

hyper-parameters	**grid**
λ_1^r	$\{10^{-4}, 10^{-3}, 0.01, 0.1, 1\}$
λ_1^F	$\{10^{-3}, 0.01, 0.1, 1, 10\}$
λ_2^r	$\{10^{-3}, 0.01, 0.1, 0.5, 1\}$
λ_3^F	$\{10^{-3}, 0.01, 0.1, 0.5, 1\}$
${\pi_0}^r$	$\{0.5, 0.8\}$
${\pi_0}^F$	$\{0.5, 0.8\}$

This table specifies under which values the hyper-parameters optimazation will be undergo. Note here that λ^r and λ^F stands for the corresponding parameter of Equations (3.4) and (3.5) respectively.

Finally, we find the following optimal hyper-parameters for each data-set of factors and for each horizon we are interested in. We present the result in Table (15).

Table 15: Optimal hyper-parameters for $S\&W$ factors.

	1	**4**	**6**	**12**	**24**
BVARR	$\lambda_1^r = 10^{-3}$	$\lambda_1^r = 0.01$ $\lambda_1^F = 0.01$	$\lambda_1^r = 0.01$ $\lambda_1^F = 0.1$	$\lambda_1^r = 0.01$ $\lambda_1^F = 0.1$	$\lambda_1^r = 0.01$ $\lambda_1^F = 0.1$
BVARRVS	$\lambda_1^r = 10^{-3}$ ${\pi_0}^r = 0.5$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.5$ $\lambda_1^F = 0.01$ ${\pi_0}^F = 0.5$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.5$ $\lambda_1^F = 0.1$ $\lambda\_4^F = 0.5$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_1^F = 0.1$ $\lambda\_4^F = 0.5$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_1^F = 0.1$ ${\pi_0}^F = 0.8$
BVARM		$\lambda_1^r = 0.01$ $\lambda_2^F = 0.1$ $\lambda_3^F = 0.01$	$\lambda_1^r = 0.01$ $\lambda_2^F = 0.1$ $\lambda_3^F = 0.1$	$\lambda_1^r = 0.01$ $\lambda_2^F = 1$ $\lambda_3^F = 0.1$	$\lambda_1^r = 0.01$ $\lambda_2^F = 1$ $\lambda_3^F = 0.1$
BVARMVS		$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.5$ $\lambda_2^F = 0.1$ $\lambda_3^F = 0.01$ ${\pi_0}^F = 0.5$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_2^F = 0.5$ $\lambda_3^F = 0.1$ ${\pi_0}^F = 0.5$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_2^F = 1$ $\lambda_3^F = 0.1$ ${\pi_0}^F = 0.8$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_2^F = 1$ $\lambda_3^F = 0.1$ ${\pi_0}^F = 0.8$

This table presents the results of optimal hyper-parametes optimazation of S&W factors, for each combination of our models and for each horizon we are interested in. Note here that the missing values from cells that correspond to models BVARM and BVARMVS for horizon=1, are values that is meaningless to optimize, since for horizon=1 we do not use Equation (3.5) and BVARR is identical to BVARM and BVARRVS is identical to BVARMVS for a single horizon.

Table 16: Optimal hyper-parameters for $G\&W$ factors

	1	**4**	**6**	**12**	**24**
BVARR	$\lambda_1^r = 10^{-4}$	$\lambda_1^r = 0.01$ $\lambda_1^F = 0.01$	$\lambda_1^r = 0.01$ $\lambda_1^F = 0.01$	$\lambda_1^r = 0.01$ $\lambda_1^F = 0.01$	$\lambda_1^r = 0.01$ $\lambda_1^F = 0.01$
BVARRVS	$\lambda_1^r = 10^{-4}$ ${\pi_0}^r = 0.5$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_1^F = 0.01$ ${\pi_0}^F = 0.5$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_1^F = 0.01$ ${\pi_0}^F = 0.8$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_1^F = 0.01$ ${\pi_0}^F = 0.8$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_1^F = 0.1$ ${\pi_0}^F = 0.8$
BVARM		$\lambda_1^r = 0.01$ $\lambda_2^F = 0.01$ $\lambda_3^F = 0.01$	$\lambda_1^r = 0.01$ $\lambda_2^F = 1$ $\lambda_3^F = 0.01$	$\lambda_1^r = 0.01$ $\lambda_2^F = 1$ $\lambda_3^F = 0.1$	$\lambda_1^r = 0.01$ $\lambda_2^F = 1$ $\lambda_3^F = 0.1$
BVARMVS		$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_2^F = 0.01$ $\lambda_3^F = 0.01$ ${\pi_0}^F = 0.8$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_2^F = 1$ $\lambda_3^F = 0.01$ ${\pi_0}^F = 0.8$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_2^F = 1$ $\lambda_3^F = 0.1$ ${\pi_0}^F = 0.8$	$\lambda_1^r = 0.01$ ${\pi_0}^r = 0.8$ $\lambda_2^F = 1$ $\lambda_3^F = 0.1$ ${\pi_0}^F = 0.8$

This table presents the results of optimal hyper-parametes optimazation of G&W factors, for each combination of our models and for each horizon we are interested in. Note here that the missing values from cells that correspond to models BVARM and BVARMVS for horizon=1, are values that is meaningless to optimize, since for horizon=1 we do not use Equation (3.5) and BVARR is identical to BVARM and BVARRVS is identical to BVARMVS for a single horizon.

It should be noted here that the compared approaches will be quite strict to the prior beliefs, especially for a single horizon. In other words, this might be an additional reason that even when we apply variable selection in our model, we do not expect huge differences, also given that we are testing without using lags in the returns and with lagged one factors, two features that probably affect the comparisons.

3.3 Comparison and Evaluation of the performance

Finally, after concluding on the model specification and parameter optimization, we are ready to test the out of sample performance of our various models on the Test set. The first comparison of the two approaches will be based on Statistical evaluation. We calculate the mean squared prediction error (MSPE) between the actual cumulative gross returns $cgr(t_h)$ as defined at (3.2.4) and the forecasted ones $c\hat{g}r(t_h)$, for each horizon h we are interested in and for each of our various models. We compare the performance of the two corresponding models by dividing the mspe's of BVAR-VS with BVAR and, finally, performing a Bayesian variation of Diebold-Mariano following Garratt et al. (2012) to test even deeper if there is significant forecasting improvement by including variable selection and we also perform our own approach on Bayesian percentage of correct signs test.

Let t belong to the Test set and τ denote the number of observations in the Test set. For the calculation of MSPE of each element j of $\hat{cgr}(t_h)$, we use the following formula:

$$MSPE_j = \frac{1}{\tau - h + 1} \sum_{t=1}^{\tau - h} (\hat{cgr}^j(t_h) - cgr^j(t_h))^2$$

Next to it, we calculate the Bayesian version of Diebold-Mariano described at Garratt et al. (2012). Enhancing to our application of this test, we test two models $m = 1, 2$, where Model 1 is the one with variable selection and Model 2 is the one without. Let us denote as y_t the actual value and as $\tilde{y_{it}^m}$ our posterior density of y for model m. We calculate the density of forecast errors as: $e^m_{i(t+h)} = y_t - \tilde{y^m_{i(t+h)}}$. Furthermore, we calculate the density of $p(d(t+h))$, where $d(t+h) = (e^1_{i(t+h)})^2 - (e^2_{i(t+h)})^2$. Thus, for each point in time $t+h$ we can calculate $Pr(d_{t+h} > 0 \mid Y)$, which will give us which model has lower posterior residuals at time t. Finally, we calculate the BDM test statistic , which will indicate which model has consistently lower forecast error in the Test Set.:

$$BDM = \frac{1}{\tau - h + 1} \sum_{t=1}^{\tau - h} Pr(d_{t+h} > 0 \mid Y)$$

The importance of non-parametric tests for the prediction of correct signs of returns, is well known in the literature (see for example Pesaran and Timmermann (1992) or Dueker and Neely (2007)). To that purpose, we introduce our Bayesian approach of percentage of correctly predicted signs, similarly to the Bayesian version of Diebold-Mariano. We calculate $Pr^j(I(y_{t+h} > 0)I(\tilde{y^j}_{i(t+h)} > 0) + I(y_{t+h} < 0)I(\tilde{y^j}_{i(t+h)} < 0) \mid Y)$ for each point in time $t+h$ and for both models $j = 1, 2$, which roughly speaking will indicate which model posterior predictive distribution has more mass to the correct sign of the actual cumulative return at time $t+h$. We calculate the Bayesian percentage of correctly predicted signs as:

$$BPCS = \frac{1}{\tau - h + 1} \sum_{t=1}^{\tau - h} Pr^j(I(y_{t+h} > 0)I(\tilde{y^j}_{i(t+h)} > 0) + I(y_{t+h} < 0)I(\tilde{y^j}_{i(t+h)} < 0) \mid Y)$$

Moreover, to test the actual portfolio performance and compare possible differences we calculate the cumulative portfolio return of each portfolio as:

$$cpr_{t \to t+h} = cpr_t + \boldsymbol{w(t_h)}' cgr(t_h)$$

Finally we evaluate the maximum drawdowns of cumulative return of our portfolios over a perior $[t-p, t]$

$$Max(Drawdown_t) = \frac{d-u}{u} \quad where: \quad d = min(cpr_{t_1}) : t_1 \in [t-p, t]$$

$$and \quad u = max(cpr_{t_2}) : t_2 \in [t_1 - q, t_1]$$

3.3.1 Illustrative example for using multiple metrics for statistical evaluation

We are now going to explain with several examples, why we are using those metrics (MSPE, BDM, BPCS) and what they are actually measuring (let us denote here as MSPE the mspe ration between two models).

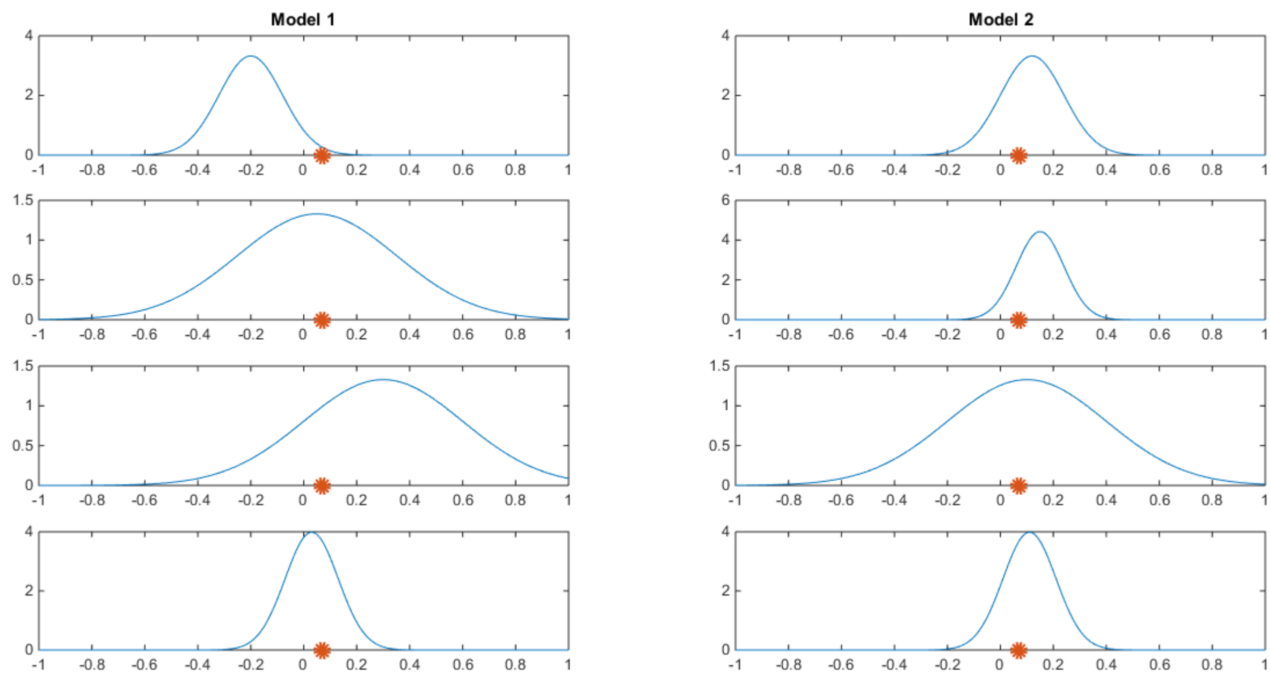
A crucial point on evaluating the performance of predictive posterior distributions is that we should not only see one perspective of performance. There is not one metric that will make clear the actual fit of our predictive distributions.

In Figure (3.3.1), we see four cases of normal distributions, produced by two models. Let us assume that we have four scenarios, where for each point in time the predictive posterior distributions, generated by two models for an asset return, is the horizontal plots of (3.3.1) and the actual return in each point in time is the same as the one indicated in the same Figure. Each row of plots corresponds to a possible scenario for two models.

To begin with, it is obvious that in these hypothetical scenarios, Model 2 will have clearly better performance compared to Model 1. In the first case, we can imagine that all three metrics will indicate that Model 2 will have better performance. In the second scenario, the MSPE will indicate that the first model will perform better, because on average the mean of the predictive distribution of Model 1 will be closer to the actual return, compared to the mean of Model 2, which will consistently be a bit further. In contrast, both BPCS and BDM will indicate Model 2, as better. The BPCS will "see" that there is great mass of the distribution on the actual return's sign, whereas BDM will "see" that on average the forecast errors are much lower for Model 2. In the third case, BPCS will indicate Model 1 as clear winner as long as it contains more mass in the actual return's sign, whereas the other two statistics will show us that Model 2 is better. Lastly, in the fourth case, BDM will show a "draw" for the comparison of the two Models and we should obtain a statistics equals 0.5. The reason is that on average, the distributions are skewed to the left and right of the actual value, they are symmetric and they have the same standard deviation. In other words the absolute errors would be exactly the same and the squared forecast errors on average should be the same. Moreover, the MSPE will also indicate draw between the two models for the same reason. Thus, we can proudly conclude, that only our BPCS statistic will indicate that Model 2 performs better, in that case.

All in all, with these examples, we would like to show that the only way to have clear indications for outperformance of a model to another, is only when all statistics agree. Thus, we have to combine portfolio performance with all these statistics to have a better image of what we actually forecast and how our forecasts "behave".

Figure 4: Four examples of posterior predictive distributions generated by two models.



This figure plots four cases of normal distributions, produced by two models. These distributions are products of four hypothetical scenarios for the predictive posterior distributions for an asset return, generated by two models. The red dot represents the actual return

4 Results

In this Section, the main results of the thesis are presented. In subsection (4.1), the results of the Gibbs sampler of Korobilis (2013) are also introduced. The approach of Stock and Watson (2002a) is followed, where common factors have been used as predictors in order for the shifts and the co-movements of global financial and macroeconomic time series to be captured. Subsection (4.2) presents the corresponding results in the case where as predictor variables the variable of Welch and Goyal (2008) have been used, where several asset specific characteristics such as the dividend-to-price ratio, the book-to-market ratio and others have been used. The risk aversion of the investor in all versions is assumed to be equal to five.

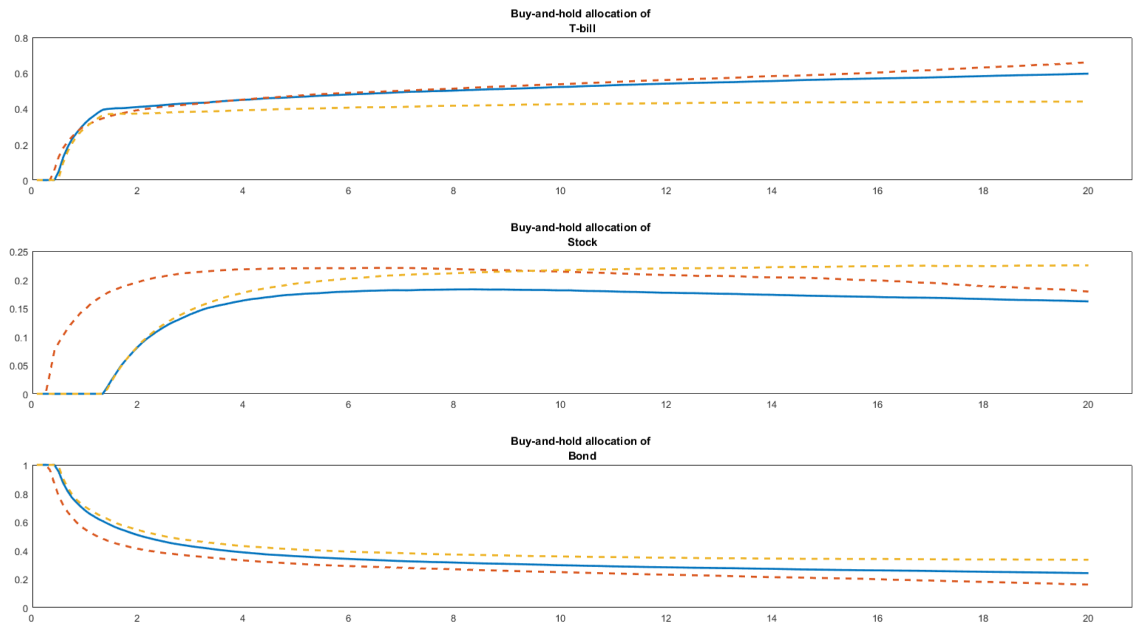
4.1 Results for Stock & Watson factors

In this subsection, the results of our algorithm is presented, Gibbs sampler with variable selection for a VAR model (Korobilis (2013)). In the VAR model, the common factors have been used as predictor variables, constructed by Stock and Watson (2002b). The allocation of weights of our method is also compared with the simple Gibbs sampler for a VAR model and with the frequentist approach of OLS estimation. For the conclusions, the portfolio performance is also taken into account which could be considered as a financial evaluation and can be seen in Appendix at (6.5.3) and (6.5.4)

To begin with, our first step into looking up our models is to provide with some end-of-sample statistics, where our results are based on 100,000 iterations of the described procedures. Figure (4.1) presents the end-of-sample allocation of weights until 20 years ahead for buy-and-hold strategy for three different assets a T-bill, a stock and a bond with our approach, the BVAR model and the OLS approach. It should be noted that there are not significant differences between the different BVAR approaches and the OLS. This is a surprising result since taking into account the parameter uncertainty should had at least differentiate the allocation of the investor in comparison to the OLS plug-in approach.

Figure 5:

Figure 6: End-of-sample allocation of weights for $S\&W$ factors.



This figure plots the end-of-sample allocation of weights for S&W factors. The investment horizon that is undetaken is each month until twenty years ahead, beggining from the end of our sample. The portfolio investment strategy is a buy and hold strategy for the T-bill (top), for a stock (middle) and for a bond (bottom). We apply the BVAR model (blue lines), the BVAR-VS method (red lines) and the OLS method (yellow lines).

Table (17) presents the posterior means and the standard deviations of the posterior distribution of the coefficients of the BVAR model that was selected for this end-of-sample implementation. We do not observe any considerable difference with OLS estimates of Tables (9) and (10). Most of the signs are in line with the OLS estimates. Table (18) presents the posterior mean for the γ coefficients from which we can decide for the predictor variables that have to be included in the model. Note here that variable selection seems to be forced and our model does not have any preference on which coefficient to select. In other words, it looks like it randomly selects the restricted factors.

Table 17: Posterior mean and standard deviations of BVAR coefficients

	factor(1)	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
r_1	-0.0001061	-0.0002083	0.0000025	-0.0001601	0.0006059
	0.0000488	0.0000557	0.0000575	0.0000653	0.0000778
r_2	-0.0011049	0.0001779	0.0052299	-0.0038923	0.0000810
	0.0006336	0.0007222	0.0007445	0.0008524	0.0010097
r_3	-0.0043893	-0.0028975	0.0022325	0.0000089	0.0016225
	0.0003941	0.0004525	0.0004639	0.0005311	0.0006265
factor(1)	0.4775899	0.4097199	0.2501671	-0.1138982	0.2495335
	0.0331442	0.0379354	0.0390752	0.0445361	0.0523748
factor(2)	0.2387316	-0.2629316	0.0654096	-0.0467852	0.0707170
	0.0360650	0.0411749	0.0422378	0.0481582	0.0569879
factor(3)	0.1133728	-0.0045780	0.4327157	-0.2973292	0.0603518
	0.0323086	0.0367015	0.0377332	0.0430644	0.0510567
factor(4)	0.0491683	-0.0112873	-0.1710603	0.6925132	0.3092577
	0.0212379	0.0241156	0.0249366	0.0282262	0.0335201
factor(5)	0.0960096	-0.0252434	0.0894832	0.2191982	0.5066257
	0.0224525	0.0256100	0.0263646	0.0301102	0.0354854

This table presents the full sample posterior means and standard deviations of the BVAR coefficients that correspond to Equations (3.4) and (3.5) for the S&W factors. Note here that we have selected p=1 and q=1. These results are based on 100,000 draws from the posterior distribution.

Table 18: Posterior mean of BVAR-VS γ coefficients

	factor(1)	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
r_1	0.0872	0.0972	0.1090	0.0359	0.9473
r_2	0.5791	0.5540	0.6258	0.6524	0.5096
r_3	0.7866	0.8382	0.4581	0.3123	0.6257
factor(1)	0.5960	0.5900	0.5373	0.3635	0.3836
factor(2)	0.4502	0.4652	0.4698	0.4060	0.4064
factor(3)	0.4818	0.4154	0.6192	0.2965	0.3927
factor(4)	0.3714	0.3413	0.5036	1.0000	0.9672
factor(5)	0.4861	0.3414	0.4188	0.8920	0.9991

This table presents the full sample posterior means of the BVAR-VS γ coefficients that correspond to Equations (3.4) and (3.5) for the S&W factors. Note here that we have selected p=1 and q=1. These results are based on 100,000 draws from the posterior distribution.

We continue our analysis by evaluating the performance of our models in the Test set, where the results are based on 20,000 iterations in each point in time.

Tables (19) and (20) present the Mspe's ratios for the comparisons of the four models. We do not observe any significant difference, since most of the values are close to one. The only characteristic that we could highlight is that variable selection seems to have better performance for forecasting cumulative portfolio returns of bonds on long horizons (12 and 24). Moreover, although for most horizons the predictability of factors is roughly the same, we observe some indications of improvement for horizons of 12 and 24. This might also make sense by looking the correlograms at (6.2.1), where we see that for longer horizons only the fourth and the fifth factor are highly correlated. Thus, when variable selection is undertaken, it might indeed produce better forecasts for those factors in longer horizons. This result highlights here the importance of variable selection when some of our state variables are highly cross-correlated.

Table 19: Mspe's ratios bewtween models BVARRVS and BVARR

horizons	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**factor(1)**	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
h=1	1.0246	0.9930	1.0880	1.2117	1.0574	1.0254	1.0795	1.0568
h=4	0.9847	1.0274	1.0151	1.0692	0.9938	1.0289	1.0863	1.0362
h=6	0.9926	1.0656	1.0124	1.0004	0.9932	1.0169	1.0513	1.0042
h=12	1.0133	1.0724	0.9460	0.9828	0.9916	1.0046	0.9974	0.9806
h=24	1.0209	1.0754	0.8724	0.9983	0.9985	1.0022	0.9891	0.9941

This table presents the mean squared prediction errors between the BVARRVS model and the BVARR model for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the S&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

Table 20: Mspe's ratios bewtween models BVARMVS and BVARM

horizons	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**factor(1)**	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
h=1	-	-	-	-	-	-	-	-
h=4	0.9764	1.0033	0.9836	1.0869	0.9918	1.0306	1.1740	1.0543
h=6	0.9814	1.0456	0.9795	0.9897	0.9899	1.0332	1.1532	1.0194
h=12	0.9985	1.0811	0.8811	0.9627	0.9990	1.0162	1.0678	0.9732
h=24	1.0010	1.1050	0.7894	1.0037	1.0035	1.0177	0.9769	0.9871

This table presents the mean squared prediction errors between the BVARMVS model and the BVARM model for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the S&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

The BDM statistics are given in Tables (21) and (22). We note here that the BVAR model has on average the lower posterior forecast errors for the prediction of the factors. In other words, we observe that BVAR-VS does not outperform in any horizon. However, we do see here that BVAR deteriorates its performance as well for the cumulative portfolio return of bond for longer horizons. However, we do not see an improvement of performance of variable selection for the fourth and the fifth factor for longer horizons as the Mspe's ratios indicated previously and as a matter of fact we do not consider that this improvement was significant. Lastly, we also observe in those two tables that when Minessota prior is applied, the performance of variable selection becomes even worse.

Table 21: BDM statistics bewtween models BVARRVS and BVARR

horizons	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**factor(1)**	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
h=1	0.5090	0.5115	0.5192	0.5394	0.5213	0.5113	0.5184	0.5154
h=4	0.4976	0.5149	0.4911	0.5159	0.5042	0.5089	0.5202	0.5093
h=6	0.4865	0.5186	0.4996	0.5076	0.5039	0.5059	0.5183	0.5054
h=12	0.5011	0.5150	0.4696	0.5059	0.5037	0.5052	0.5095	0.5004
h=24	0.5121	0.5163	0.4556	0.5090	0.5055	0.5039	0.5081	0.5024

This table presents the Bayesian Diebold-Mariano statistics between the BVARRVS model and the BVARR model for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the S&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

Table 22: BDM statistics bewtween models BVARMVS and BVARM

horizons	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**factor(1)**	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
h=1	-	-	-	-	-	-	-	-
h=4	0.4963	0.5152	0.5006	0.5227	0.5097	0.5176	0.5367	0.5158
h=6	0.4934	0.5188	0.4997	0.5110	0.5085	0.5148	0.5370	0.5126
h=12	0.4953	0.5178	0.4886	0.5067	0.5097	0.5127	0.5301	0.5059
h=24	0.4982	0.5231	0.4863	0.5156	0.5100	0.5164	0.5244	0.5058

This table presents the Bayesian Diebold-Mariano statistics between the BVARMVS model and the BVARM model for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the S&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

In Table (23), the corresponding BPCS are presented. When variable selection is applied here as well, we see that on average the posterior predictive returns have roughly the same proportion of correct sign as when the simple BVAR is undertaken. However, for the cumulative return of bond in longer horizons, we see again that variable selection outperform of BVAR. Note, also, that by applying Minessota prior, clearly increases the statistics of BPS.

Table 23: BPCS statistics for all models and all horizons

horizons	**Models**	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$
h=1	BVARR	0.5557	0.5509	0.5879
	BVARRVS	0.5494	0.5465	0.5672
	BVARM	-	-	-
	BVARMVS	-	-	-
h=4	BVARR	0.5287	0.5829	0.5410
	BVARRVS	0.5351	0.5702	0.5354
	BVARM	0.5323	0.5808	0.5444
	BVARMVS	0.5393	0.5709	0.5388
h=6	BVARR	0.4927	0.5886	0.5334
	BVARRVS	0.4964	0.5700	0.5283
	BVARM	0.5012	0.5883	0.5361
	BVARMVS	0.5062	0.5703	0.5318
h=12	BVARR	0.4682	0.6149	0.5797
	BVARRVS	0.4667	0.5969	0.5983
	BVARM	0.4854	0.6192	0.5802
	BVARMVS	0.4827	0.5984	0.5979
h=24	BVARR	0.4277	0.5671	0.6629
	BVARRVS	0.4268	0.5540	0.6902
	BVARM	0.4386	0.5753	0.6551
	BVARMVS	0.4387	0.5581	0.6804

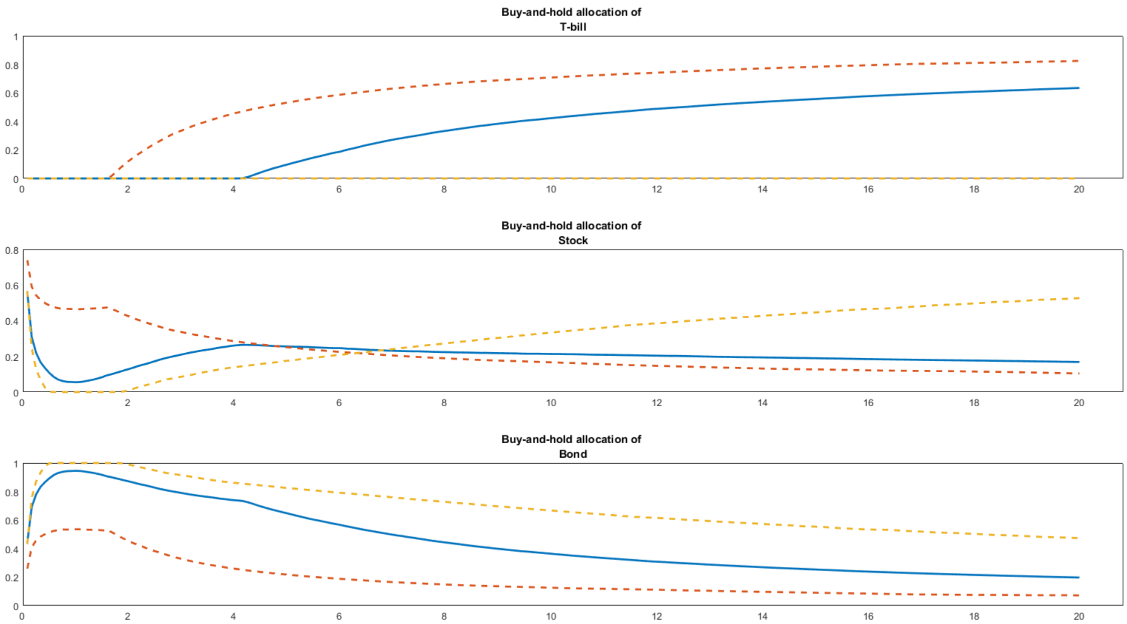
This table presents the Bayesian percentage of correct signs statistics for all of our models for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the S&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

To conclude, as far as the results of our algorithm are concerned, applied on the data set with factors based in Stock and Watson (2002a), we note that this set of predictor variables seems to be a well-constructed set for the description of the connections between the components of the portfolio. This fact is also reflected in the posterior means of γ coefficients. Moreover, in the portfolio performance we hardly see any difference between the two main approaches. Given also the above statistical analysis of our results, it seems that the variable selection technique does not any advantage against the simple approach. In conclusion, we do not observe any significant improvement in the results (allocation of weights, significant variables, etc.) of our approach and of the existing in the literature approaches.

4.2 Results for Goyal & Welch factors

In this subsection, we present the results of our algorithm, Gibbs sampler with variable selection for a VAR model (Korobilis (2013)). In this VAR model, we use the predictor variables that have been used in Diris (2014), based on Welch and Goyal (2008). We also compare the allocation of weights of our method with the simple Gibbs sampler for a VAR model and with the frequentist approach of OLS estimation. For once again, we also consider the portfolio performance which could be seen in Appendix at (6.5.3) and (6.5.4).

Figure 7: End-of-sample allocation of weights for $G\&W$ factors.



This figure plots the end-of-sample allocation of weights for G&W factors. The investment horizon that is undetaken is each month until twenty years ahead, beggining from the end of our sample. The portfolio investment strategy is a buy and hold strategy for the T-bill (top), for a stock (middle) and for a bond (bottom). We apply the BVAR model (blue lines), the BVAR-VS method (red lines) and the OLS method (yellow lines).

The initial step into looking up our models is to provide with some end-of-sample statistics, where again the results are based on 100,000 iterations of the described procedures. Figure (4.2) presents the allocation of weights up to 20 years ahead for buy and hold strategy for three different assets, the T-bill, a stock and a bond with our approach, the BVAR model and the OLS approach. It is clear that the allocation of the weights is quite different for the three approaches. The allocation of the weights using our algorithm is in line with the work of Diris (2014), where the model uncertainty is incorporated. More precisely, we can see in Figure (4.2) that the investor's uncertainty is increasing and, as a result, he invests less to risky assets like stocks and bonds. We can also note in this Figure that, as expected, as the time horizon grows, the differences among the three methods are becoming smaller. Finally, in Figure 6 we can see that the main difference between our method and the method of the simple Gibbs sampler for a VAR model is that the investment in riskless assets, such as the T-bill, is taking place earlier.

Table (24) presents the posterior mean and the standard deviation of the posterior distribution of the coefficients of the BVAR model that was selected for this end-of-sample application. Most of the signs of our posterior means are in line with the OLS coefficients of Tables (3.1) and (3.1), whereas the magnitude of the effects had shrunk. Table (25) presents the posterior mean for the γ coefficients from which we can decide for the predictor variables that have to be included in the model. Surprisingly, many of the popular predictors, such as the divident-to-price ratio, are out of the preferences of our algorithm. Variable selection tends to prefer the detrended log nominal yield (Ynom), the yield spread (Yspr) and the credit spread (Crspr) as the most popular predictors in order to forecast the asset returns.

Taking into account both the results illustrated in Figure (4.2) and the mean posterior selection-coefficients in Table (25), we note that our method in which the variable selection uncertainty in a VAR model has been incorporated, converges towards an optimal model. This result is in contrast with the results from the previous subsection where only some variables have been picked up. Roughly speaking, in this case it looks that our model is quite picky on factor selection.

Table 24: Posterior mean and standard deviations of BVAR coefficients

	Defpr	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
r_1	-0.0259	0.0048	-0.0043	-0.0010	-0.0274	-0.0363	0.1034	0.0023	0.0000
	0.0082	0.0008	0.0007	0.0005	0.0163	0.0103	0.0435	0.0073	0.0001
r_2	0.1807	0.0238	-0.0213	-0.0101	-0.2834	0.1791	0.2602	-0.0036	-0.0026
	0.1136	0.0115	0.0100	0.0064	0.2199	0.1420	0.5211	0.1000	0.0020
r_3	-0.0120	0.0004	0.0007	-0.0040	0.2246	0.3764	-0.4836	-0.1113	0.0032
	0.0758	0.0076	0.0066	0.0043	0.1487	0.0947	0.3760	0.0663	0.0014
Defpr	-0.1062	0.0055	-0.0017	0.0057	-0.0088	-0.0380	0.5004	0.0359	-0.0007
	0.0382	0.0038	0.0033	0.0022	0.0754	0.0480	0.1993	0.0337	0.0007
DP	-0.1738	0.9628	0.0283	0.0017	0.3411	-0.0430	-0.7836	-0.0757	0.0021
	0.1145	0.0115	0.0101	0.0065	0.2220	0.1432	0.5265	0.1009	0.0020
BM	-0.0315	-0.0081	0.9919	-0.0093	0.4222	0.1026	0.0120	0.1856	-0.0002
	0.1563	0.0157	0.0138	0.0089	0.3080	0.1962	0.7833	0.1383	0.0028
PE	-0.2015	0.0343	-0.0432	0.9738	-1.8228	-0.8965	2.8817	-0.1890	0.0007
	0.1585	0.0160	0.0140	0.0090	0.3138	0.1994	0.7999	0.1410	0.0029
Ynom	0.0112	-0.0015	0.0020	0.0006	0.9276	0.0621	-0.1241	0.0010	-0.0002
	0.0097	0.0010	0.0009	0.0005	0.0193	0.0122	0.0515	0.0087	0.0002
Yspr	-0.0144	0.0015	-0.0021	0.0001	-0.0663	0.9149	0.2233	0.0133	-0.0001
	0.0103	0.0010	0.0009	0.0006	0.0204	0.0129	0.0545	0.0091	0.0002
Crspr	-0.0188	0.0001	0.0000	-0.0003	0.0017	-0.0034	0.9317	-0.0042	0.0003
	0.0025	0.0003	0.0002	0.0001	0.0050	0.0031	0.0133	0.0022	0.0000
Ntis	0.0160	0.0015	0.0002	0.0023	-0.0205	0.0006	-0.0053	0.9725	-0.0004
	0.0108	0.0011	0.0009	0.0006	0.0214	0.0136	0.0574	0.0096	0.0002
Var	-4.7232	-0.1408	-0.1850	-0.0752	2.3409	-0.8944	42.7131	-0.8132	0.6174
	1.6799	0.1702	0.1486	0.0957	3.3283	2.1138	8.8794	1.4876	0.0308

This table presents the full sample posterior means and standard deviations of the BVAR coefficients that correspond to Equations (3.4) and (3.5) for the G&W factors. Note here that we have selected p=1 and q=1. These results are based on 100,000 draws from the posterior distribution.

Table 25: Posterior mean of BVAR-VS γ coefficients

	Defpr	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
r_1	0.4302	0.0294	0.0176	0.0549	0.7930	0.6002	0.6307	0.2005	0.0127
r_2	0.7303	0.1258	0.1537	0.3853	0.8041	0.7998	0.7961	0.7048	0.0787
r_3	0.6446	0.0733	0.0991	0.2822	0.8063	0.6573	0.8012	0.5775	0.0635
Defpr	0.1617	0.0831	0.1373	0.2845	0.4993	0.2509	0.4317	0.1519	0.5950
DP	0.3065	1.0000	0.0885	0.2325	0.4882	0.1664	0.2078	0.0984	0.3364
BM	0.2843	0.0880	1.0000	0.2591	0.4928	0.1919	0.1999	0.1004	0.3520
PE	0.2633	0.0545	0.0907	1.0000	0.4745	0.1477	0.2515	0.1296	0.3332
Ynom	0.3794	0.1008	0.1455	0.3142	0.3563	0.2553	0.2893	0.1325	0.4218
Yspr	0.3515	0.0667	0.1167	0.3314	0.4925	1.0000	0.2101	0.1077	0.3962
Crspr	0.2878	0.0712	0.1151	0.3412	0.4896	0.2177	1.0000	0.0985	0.6251
Ntis	0.4074	0.0730	0.1087	0.3977	0.4957	0.2441	0.3466	1.0000	0.4427
Var	0.5249	0.0869	0.1292	0.3315	0.4957	0.2599	0.2628	0.1303	0.1160

This table presents the full sample posterior means of the BVAR-VS γ coefficients that correspond to Equations (3.4) and (3.5) for the G&W factors. Note here that we have selected p=1 and q=1. These results are based on 100,000 draws from the posterior distribution.

Continuing with the statistical and financial performance of our models in the Test set, where the results are based on 20,000 iterations in each point in time, Tables (26)-(30) present the Mspe's ratios, the BDM statistics and the BPCS statistics for the comparisons of the four models.

It is clear that the Mspe's ratios in Tables (26) and (27) are smaller or close to one for short horizons, indicating that the variable selection technique is more advantageous compared to the simple approach. We do observe any improvement for the BVAR-VS, but this is the case only for a single horizon and a quarter horizon of investment. We also see this trend on the prediction of the factors. For example, it seems that variable selection does a good job for predicting one of the most important predictors which is the detrended log nominal yield (Ynom), but only for short horizons ($h = 1, 4$). For once again by looking the correlograms at (6.2.1), this result makes perfectly sense because we see that these factors are strongly correlated. Again it is clear, that, variable selection is is specialist on detecting when some of our state variables are highly cross-correlated.

Table 26: Mspe's ratios bewtween models BVARRVS and BVARR

horizons	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
h=1	0.9908	0.9871	0.9852	1.0007	0.9973	1.0041	0.9989	0.9860	1.0004	0.9979	1.0016	0.9715
h=4	0.9880	0.9820	0.9693	0.9970	0.9759	1.0193	0.9970	0.9484	0.9938	1.0122	1.0446	0.9263
h=6	0.9955	1.0062	0.9873	1.0104	0.9995	1.0178	0.9838	0.9802	0.9969	1.0137	1.0870	0.9313
h=12	1.0029	1.0232	0.9881	1.0121	1.0477	1.0449	1.0531	0.9889	1.0236	0.9758	1.0162	0.9701
h=24	0.9635	1.2241	1.0197	0.9743	1.0575	1.0916	0.9773	1.0196	0.9723	1.0326	1.1096	0.9590

This table presents the mean squared prediction errors between the BVARRVS model and the BVARR model for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the G&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

Table 27: Mspe's ratios bewtween models BVARMVS and BVARM

horizons	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
h=1	-	-	-	-	-	-	-	-	-	-	-	-
h=4	0.9864	0.9977	0.9640	0.9961	0.9714	1.0218	0.9985	0.9534	0.9955	0.9917	1.0438	0.9363
h=6	1.0062	1.0042	1.0151	1.0065	1.0035	1.0379	0.9841	0.9825	0.9983	1.0081	1.1089	0.9601
h=12	1.0036	1.0064	0.9222	1.0245	1.0477	1.0407	0.9787	0.9645	0.9959	1.0268	1.0383	0.9805
h=24	0.9956	1.2195	1.0166	1.0471	1.0118	1.0730	0.9820	0.9761	0.9979	1.0098	1.3731	0.9619

This table presents the mean squared prediction errors between the BVARMVS model and the BVARM model for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the G&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

Continuing our analysis and looking at the BDM statistics in Tables (28) and (29), we note that the BVAR-VS model produces on average the lower posterior residuals and it performs better, as documented in the aforementioned description, only for horizons $h = 1, 4$. Moreover, the BVAR-VS has a great performance on predicting some important predictors (Ynom and Yspr) and this is directly reflected on the accurate prediction of cumulative asset returns for short horizons. Here, the predictive power of variable selection is also confirmed by the BDM statistics.

Table 28: BDM statistics bewtween models BVARRVS and BVARR

	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
h=1	0.4914	0.4986	0.4985	0.4949	0.4949	0.5182	0.5041	0.4627	0.497	0.5036	0.5063	0.4883
h=4	0.4826	0.4908	0.4739	0.4941	0.4766	0.5278	0.5031	0.4547	0.5191	0.5034	0.5123	0.4466
h=6	0.4969	0.4962	0.5068	0.5092	0.5006	0.5211	0.5037	0.4830	0.4951	0.5076	0.5363	0.4681
h=12	0.5018	0.5106	0.5017	0.5025	0.4949	0.5229	0.4959	0.4924	0.5006	0.5087	0.5194	0.4875
h=24	0.4948	0.5449	0.5015	0.5027	0.5079	0.4966	0.5038	0.5073	0.4972	0.5008	0.5103	0.5045

This table presents the Bayesian Diebold-Mariano statistics between the BVARRVS model and the BVARR model for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the G&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

Table 29: BDM statistics bewtween models BVARMVS and BVARM

	r_1	r_2	r_3	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
h=1	-	-	-	-	-	-	-	-	-	-	-	-
h=4	0.4873	0.4908	0.4781	0.4967	0.4789	0.5299	0.4949	0.4576	0.5047	0.5091	0.5212	0.4444
h=6	0.5026	0.5095	0.5162	0.5144	0.5036	0.5233	0.5071	0.4914	0.4963	0.5095	0.5422	0.4656
h=12	0.4945	0.5295	0.5067	0.5195	0.5086	0.5198	0.5006	0.4954	0.5033	0.4948	0.5216	0.4886
h=24	0.4985	0.5475	0.5045	0.5089	0.5022	0.5088	0.5023	0.5044	0.4991	0.4999	0.5163	0.5056

This table presents the Bayesian Diebold-Mariano statistics between the BVARMVS model and the BVARM model for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the G&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

Finally, the BPCS statistics are provided in Table (30). Although the previous two metrics had shown the variable selection as a clear winner for short horizons, this is roughly the case here, since the differences are rather small. However, we still see here some small advanced performance of variable selection for predicting all asset classes at short horizons. As a matter of fact, we consider that this advanced performance of variable selection is significant. Moreover, once more we notice that the Minessota prior gives a great boost to the BPCS statistics.

Table 30: BPCS statistics for all models and all horizons

horizons	**Models**	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$
h=1	BVARR	0.5280	0.5008	0.5154
	BVARRVS	0.5232	0.5187	0.5185
	BVARM	-	-	-
	BVARMVS	-	-	-
h=4	BVARR	0.5309	0.5027	0.5225
	BVARRVS	0.5310	0.5172	0.5272
	BVARM	0.5309	0.5048	0.5232
	BVARMVS	0.5311	0.5176	0.5279
h=6	BVARR	0.5134	0.4720	0.5288
	BVARRVS	0.5103	0.4648	0.5290
	BVARM	0.5185	0.4809	0.5289
	BVARMVS	0.5186	0.4761	0.5293
h=12	BVARR	0.5041	0.4648	0.6281
	BVARRVS	0.5099	0.4421	0.6288
	BVARM	0.5084	0.4669	0.6241
	BVARMVS	0.5099	0.4474	0.6330
h=24	BVARR	0.5036	0.4175	0.7129
	BVARRVS	0.5040	0.3952	0.6766
	BVARM	0.5054	0.4348	0.6879
	BVARMVS	0.5049	0.4069	0.6581

This table presents the Bayesian percentage of correct signs statistics for all of our models for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the G&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

Concluding this subsection, we note that for h=1 and h=4, most statistics and the portfolio performance indicate that incorporating variable selection in a Gibbs sampler for a VAR model improves the results. This result is, probably, due to the fact that for short horizons not all the predictor variables are necessary, while the posterior predictive distribution is affected by the ineffective predictor variables. Therefore, taking into account the variable uncertainty, we might come out with predictive distribution with narrow credible intervals. It seems that the Gibbs sampler with Variable selection produces predictive distributions with narrow credible intervals with mean which is closer to the real value in short time horizons, while for longer horizons the mean is quite different from the real value. We end up with this conclusion, since it is clear from the Tables above that all statistics are better for the returns and for the factors for short horizons. The most important result in this subsection, and in general in the present thesis, is the Gibbs sampler which incorporates the variable selection technique, improves the prediction of very important factors for h=4, resulting in improved prediction of the returns. More importantly, this result is also reflected by the actual portfolio perfomance, where in our back-test portfolio perfomance we see at (6.5.3) and (6.5.4), that variable selection again outperforms. Finally, we note that variable selection is a useful contribution, since it generates biased but efficient estimations which in practice predicts better both our return and the factor for short horizons. For longer horizons, we either need all factors or both models loose their predictive power.

4.3 Conclusion remarks of results

All in all, in this Section we tested the out of sample performance of two algorithms. The Gibbs sampler and the Gibbs sampler with variable selection applied on a vector autoregressive model (VAR). We also applied for the same set-ups of the two algorithms with a Minnesota prior. All of our methods applied in two sets of factors ($S\&W$ and $G\&W$) and at five different horizons ahead (1,4,6,12,24). We evaluate the performance of our BVAR and BVAR-VS by several statistics and by the portfolio performance. The evaluation of the models for the $S\&W$ factors indicates a clear draw between the two approaches in all versions. None of the two algorithms succeeds on consistently outperform in any horizon and with any kind of prior. This feature could also be observed on the allocation of the weights of the two approaches and on the portfolio performance. On the other hand, the evaluation of the two models for the $G\&W$ factors indicates that in all versions, variable selection is a clear winner for short horizons (h=1,4). This contradicting result between the two data-sets might be caused for several reasons. Firstly, the $S\&W$ is a small data-set (5 factors) compared to $G\&W$ (9 factors) and thus BVAR underperforms only when over-parametrization is a big problem. Secondly, $S\&W$ factors is a product of principal component analysis. This might affect the variable selection, if by construction these factors are very solid. Both of these comments are reflected also in Table (18), where variable selection seems not to be able to consistently pick up which factors to include. On the other hand, $G\&W$ seems to be a bigger and more vivacious data-set of factors, with many cross-correlations and eventually variable selection does a good job on converging which are really usefull. However, for the $G\&W$ factors we see that this advantage is lost for longer horizons, where is seems that all factors are needed. Finally, in all versions we see that when we apply the Minessota prior in both of our algorithms, the relative performance of the BVAR to BVAR-VS compared to the relative performance of the BVAR to BVAR-VS when we do not apply Minessota prior, is almost always improved. This result is also in line with Korobilis (2013) and it seems that Minessota prior shrinks the BVAR coefficients and improves the performance in a bigger magnitude compared to the pre-shrunk BVAR-VS coefficients.

5 Conclusion

The present thesis contributes to the problem of portfolio choice. It proposes an alternative approach of choosing the optimal weights, using recent results from the statistical literature of Bayesian analysis, while the approach is compared with existing ones. More precisely, the use of computationally efficient algorithms for stochastic variable selection is suggested. Using the proposed algorithms, the estimations and predictions for the optimal portfolio weights are taking into account both the parameter and the variable selection uncertainty, working in line with the work in Diris (2014) where both parameter and model uncertainty is taken into account.

In particular, Markov Chain Monte Carlo (MCMC) methods are used, and more precisely, an efficient Gibbs sampler for the Bayesian estimation of a VAR model which includes predictor variables related to two different approaches in the literature (Stock and Watson (2002a) and Welch and Goyal (2008)). The main contribution of this thesis is the use of a technique for automatic selection of variables in Bayesian VAR models using the Gibbs sampler, which is proposed by Korobilis (2013). The algorithm of Korobilis (2013) does not rely on estimation of all the possible VAR model combinations with a method which is independent of the prior assumptions about the coefficients of the VAR model, as in the case of the traditional Gibbs sampler. In contrast, this algorithm decides for the most suitable variables to be included in the model with goodness-of-fit techniques. In other words, the algorithm used, visits only the most probable models in a stochastic manner rather than enumerating and estimating all possible models. A nice and useful feature of the algorithm of Korobilis (2013) is the introduction of random variables (indicators) which indicate the variables that are significant for the VAR model in an automatic way, during the MCMC algorithm.

This thesis evaluates the performance of our two approaches from a statistical and a financial perspective. The main results can be summarized as follows. Using the data set with predictor variables which is constructed as in the paper of Stock and Watson (2002a), we do not view any significant differences on the allocation of the portfolio weights using our method for the estimation of the model and the usual Gibbs sampler for the estimation of a VAR model. The most significant result of the thesis is noted when the data set of predictor variables is used, constructed by Diris (2014) and based on Welch and Goyal (2008). In this case, with our method and using the Gibbs sampler proposed by Korobilis (2013), we have found that incorporating variable-selection uncertainty increases investor's uncertainty and invests less to stocks and bonds when the time horizon goes further and further. More precisely, for this data-set of predictor variables, we have found significant improvement both in factor and returns predictability for a single horizon's investments and four horizon's investments. This improvement is, also, observed in the final financial portfolio evaluation, whereas for longer horizons, we have found that all the variables are significant for improved returns and/or factor predictability. This is a result which is in line with the result of Diris (2014), where both parameter and model uncertainty is taken into account in the proposed estimation procedure. The big advantage of our proposed method, compared with the existing Gibbs samplers for the estimation of VAR models, is the provision of the matrix with the posterior means of restricted variables, which could be interpreted as the mean posterior selection-coefficients for including a predictor variable in the final VAR model. Visual comparison of the factor's correlograms and the estimated mean posterior selection-coefficients of including a specific factor in our model give a lot of evidence for the practical use of our method. We further note that for small time horizons, our posterior distributions have smaller range than the posteriors of the usual Gibbs sampler for a VAR model, based on our statistical results.

Thus, concluding the present thesis, we would like to highlight the trade off between efficiency and bias that variable selection techniques introduce to the estimation and prediction procedures in the problem of choosing the optimal weights in financial portfolios.

Future work in the context of the present thesis could be related to the exploration of the properties of the Gibbs sampler proposed, in more complicated models, where the number of factors that are used is an unknown parameter to be estimated. Specifically, future research could focus on the impact of recent advances in the Bayesian literature of dynamic factor models in the optimal portfolio choice problem. Moreover, methods of relaxing the identifiability constraints of the models (invariance under rotation of the factors and label and sign switching of the factor loadings) could be investigated in the future, while a method for choosing the number of factors used could also be proposed. Moreover, one starting point for further investigation could be the sensitivity analysis of lags of factor selection for the final performance of variable selection. Additionally, another perspective of further sensitivity analysis of this research regards the range of portfolio constrains and the risk-aversion γ (in our version we do not allow for sort selling and we also set $\gamma = 5$ which might had been too strict for the final portfolio evaluation). In summary, future work in the problem of finding the optimal allocation strategy could be in the direction of investigating recent advances of Bayesian dynamic factor models, where also the paper of Aßmann, Boysen-Hogrefe, and Pape (2016) could be used as a starting point.

References

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics 18*(3), 338–357.

Aït-Sahalia, Y. and M. W. Brandt (2001). Variable selection for portfolio choice. *The Journal of Finance 56*(4), 1297–1351.

Aßmann, C., J. Boysen-Hogrefe, and M. Pape (2016). Bayesian analysis of static and dynamic factor models: An ex-post approach towards the rotation problem. *Journal of Econometrics*.

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

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of finance*, 225–264.

Brandt, M. (2009). Portfolio choice problems. *Handbook of financial econometrics 1*, 269–336.

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

Campbell, J. Y. (1990). A variance decomposition for stock returns. Technical report, National Bureau of Economic Research.

Campbell, J. Y. and R. J. Shiller (1988). Stock prices, earnings, and expected dividends. *The Journal of Finance 43*(3), 661–676.

Campbell, J. Y. and L. M. Viceira (2005). The term structure of the risk-return trade-off. *Financial Analysts Journal 61*(1), 34–44.

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

Dellaportas, P. and G. O. Roberts (2003). An introduction to mcmc. In *Spatial statistics and computational methods*, pp. 1–41. Springer.

Diris, B. F. (2014). Model uncertainty for long-term investors. *Available at SSRN 1786587*.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews 3*(1), 1–100.

Dueker, M. and C. J. Neely (2007). Can markov switching models predict excess foreign exchange returns? *Journal of Banking & Finance 31*(2), 279–296.

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

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of financial economics 22*(1), 3–25.

Frost, P. A. and J. E. Savarino (1988). For better performance: Constrain portfolio weights.

Garratt, A., G. Koop, E. Mise, and S. P. Vahey (2012). Real-time prediction with uk monetary aggregates in the presence of model uncertainty. *Journal of Business & Economic Statistics*.

Gelfand, A. E. and A. F. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association 85*(410), 398–409.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2014). *Bayesian data analysis*, Volume 2. Taylor & Francis.

George, E. I., D. Sun, and S. Ni (2008). Bayesian stochastic search for var model restrictions. *Journal of Econometrics 142*(1), 553–580.

Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the stock and bond markets. *Journal of financial Economics 17*(2), 357–390.

Korobilis, D. (2013). Var forecasting using bayesian variable selection. *Journal of Applied Econometrics 28*(2), 204–230.

Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 815–849.

Markowitz, H. (1952). Portfolio selection*. *The journal of finance 7*(1), 77–91.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, 247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of economic theory 3*(4), 373–413.

Pesaran, M. H. and A. Timmermann (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics 10*(4), 461–465.

Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *The review of economics and statistics*, 239–246.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–162.

Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics 24*(4), 630–650.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

6 Appendix

6.1 Stock and Watson factor construction details

To deal with stationary problems, we transform the original time series to non-stationary, when it is needed. To go in line with Stock and Watson (2002b) we use various methods of transformations for each type of time series. These transformations are indicated as T-codes in Table 6.1. We apply these defined transformations in our final database. You can find the analytical description of variables used in Table 32. Other abbreviations that are used are: SA stands for seasonal adjustment; if the data is seasonally adjusted the value is 1, otherwise it is 0. The sources we use are: FRED = Federal Reserve Economic Data, ECB = European Central Bank, OECD = Organization of Economic Cooperation and Development, IMF = International Monetary Fund, BB = Bloomberg, BIS = Bank for International Settlements, BoJ = Bank of Japan, BLS = U.S. Bureau of Labor Statistics, S&W = Database of Stock and Watson (2002b). Used indicates whether or not the variable is used (1) or not (0) in factor construction.

Table 31: Transformations

T-code	Transformation	Formula
1	No transformation	
2	First difference	$Y_t = \Delta X_{t,t-1}$
3	Second difference	$Y_t = \Delta X_{t,t-1} - \Delta X_{t-1,t-2}$
4	Logarithm	$Y_t = \ln X_t$
5	First difference of the logarithms	$Y_t = \Delta \ln X_{t,t-1}$
6	Second difference of the logarithms	$Y_t = \Delta \ln X_{t,t-1} - \Delta \ln X_{t-1,t-2}$
7	Log growth rates	$Y_t = \ln(1 + \frac{X_t}{100})$

This table presents the transformations of time series for the corresponding T-codes for the constructions of initials data-set proposed by Stock and Watson (2002a). Y_t is the transformed variable. X_t is the raw data.

Table 32: Data Series

	Variable Name	Title	Unit	T-code	SA	So
		FRANCE				
1	UNEMFR	Unemployment Rate	Rate	1	1	EC
2	NBFRBIS	Broad Effective Exchange Rate	Index (2010=100)	5	0	BI
3	NNFRBIS	Narrow Effective Exchange Rate	Index (2010=100)	5	0	BI
4	TRESEGFRM052N	Total Reserves	Dollars	6	1	IM
5	PRINTO01FR	Industrial Production	Index (2010=100)	5	1	OE
6	ODCNPI03FR	Permits Issued for Dwelling	Index (2010=100)	4	1	OE
7	LOLITOAAFR	Leading Indicators	Index (2010=100)	5	0	OE
8	XTEXVA01FR	Exports	US Dollar, millions	5	1	OE
9	XTIMVA01FR	Imports	US Dollar, millions	5	1	OE
10	CAC	Stock Index	Price	5	0	BE
11	GTFRF6M	6-Month Bond	Percent	2	0	BE
12	GTFRF1Y	1-Year Bond	Percent	2	0	BE
13	GTFRF2Y	2-Year Bond	Percent	2	0	BE
14	CPIFR	Consumer Price Index	Index (2010=100)	6	0	OE
15	IRLTLT01FRM156N	10-Year Bond	Percent	2	0	OE
		GERMANY				
16	UNEMDE	Unemployment Rate	Rate	1	1	EC
17	DEUPPDMMINMEI	Domestic Producer Prices Index: Manufacturing	Index (2010=100)	6	1	OE
18	NBDEBIS	Broad Effective Exchange Rate	Index (2010=100)	5	0	BI
19	NNDEBIS	Narrow Effective Exchange Rate	Index (2010=100)	5	0	BI
20	PIEATI01DEM661N	Producer Price Index: All Industrial Activity	Index (2010=100)	6	1	OE

21	TRESEGDEM052N	Total Reserves	Dollars	5	1	IM
22	XFORSD01DEM194N	Reserve Assets	Special Drawing Rights	5	0	OE
23	DAX	Stock Index	Price	5	0	BE
24	GETB2	6-Month Bond	Percent	2	0	BE
25	GDBR1	1-Year Bond	Percent	2	0	BE
26	GDBR2	2-Year Bond	Percent	2	0	BE
27	PRINTO01DE	Industrial Production	Index (2010=100)	5	1	OE
28	ODCNPI03DE	Permits Issued for Dwelling	Index (2010=100)	4	1	OE
29	LOLITOAADE	Leading Indicators	Index (2010=100)	5	0	OE
30	IR3TIB01DE	Interbank Interest Rate 3-Month (Euribor after '99)	Percentage	2	0	OE
31	XTEXVA01DE	Exports	US Dollar, millions	5	1	OE
32	XTIMVA01DE	Imports	US Dollar, millions	5	1	OE
33	CPIDE	Consumer Price Index	Index (2010=100)	6	0	FR
34	SX5E	Euro Stoxx 50 Index	Price	5	0	BE
35	MYAGM1EZM196N	EUROZONE M1 Money Stock	Euros	6	0	IM
	Variable Name	Title	Unit	T-code	SA	So
36	MYAGM2EZM196N	EUROZONE M2 Money Stock	Euros	6	0	IM
37	MYAGM3EZM196N	EUROZONE M3 Money Stock	Euros	6	0	IM
38	IRLTLT01DEM156N	10-Year Bond	Percent	2	0	OE
		ITALY				
39	UNEMIT	Unemployment Rate	Rate	1	1	EC
40	PRINTO01IT	Industrial Production	Index (2010=100)	5	1	OE
41	LOLITOAAIT	Leading Indicators	Index (2010=100)	5	0	OE
42	XTEXVA01IT	Exports	US Dollar, millions	5	1	OE
43	XTIMVA01IT	Imports	US Dollar, millions	5	1	OE

44	FTSEMIB	Stock Index	Price	5	0	BE
45	GTITL6M	6-Month Bond	Percent	2	0	BE
46	GTITL1Y	1-Year Bond	Percent	2	0	BE
47	GTITL2Y	2-Year Bond	Percent	2	0	BE
48	CPIIT	Consumer Price Index	Index (2010=100)	6	0	OE
49	IRLTLT01ITM156N	10-Year Bond	Percent	2	0	OE
		JAPAN				
50	CSCICP02JPM460S	Consumer Opinion Surveys: Composite Confidence Indicators	Net Percent	2	1	OE
51	JPNCPIALLMINMEI	Consumer Price Index (All Items)	Index (2010=100)	6	0	OE
52	JPNPERMITMISMEI	Permits Issued for Dwelling	Index (2005=100)	4	1	OE
53	JPNPROMANMISMEI	Production in total Manufacturing	Index (2010=100)	5	1	OE
54	LRHUTTTTJPM156S	Harmonized Unemployment: Total	Rate	1	1	OE
55	MABMM301JPM189S	M3	Yen	6	1	OE
56	MYAGM1JPM189S	M1	Yen	6	1	OE
57	MYAGM2JPM189S	M2	Yen	6	1	OE
58	SPASTT01JPM661N	Total Share Prices for All Shares	Index (2010=1.00)	5	0	OE
59	VALEXPJPM052N	Goods, Value of Exports	US Dollar	5	0	IM
60	XFORSD01JPM194N	Reserve Assets	Special Drawing Rights	5	0	OE
61	JPNASSETS	Bank of Japan: Total Assets	100 Million Yen	5	0	Bo
62	IRLTLT01JPM156N	10-Year Bond	Percent	2	0	OE
63	INTGSTJPM193N	T-Bills Japan	Percent per Annum	2	0	IM
64	NIKKEI	Nikkei Index	Price	5	0	BE
		NETHERLANDS				
65	UNEMNL	Unemployment Rate	Rate	1	1	EC

66	PRINTO01NL	Industrial Production	Index (2010=100)	5	1	OE
67	ODCNPI03NL	Permits Issued for Dwelling	Index (2010=100)	4	1	OE
68	LOLITOAANL	Leading Indicators	Index (2010=100)	5	0	OE
69	XTEXVA01NL	Exports	US Dollar, millions	5	1	OE
	Variable Name	Title	Unit	T-code	SA	So
70	XTIMVA01NL	Imports	US Dollar, millions	5	1	OE
71	AEX	Stock Index	Price	5	0	BE
72	GTNLG6M	6-Month Bond	Percent	2	0	BE
73	GTNLG1Y	1-Year Bond	Percent	2	0	BE
74	GTNLG2Y	2-Year Bond	Percent	2	0	BE
75	CPINL	Consumer Price Index	Index (2010=100)	6	0	OE
76	NLDPPDMMINMEI	Domestic Producer Prices Index: Manufacturing	Index (2010=100)	6	0	OE
77	NLDSARTMISMEI	Total Retail Trade	Index (2010=100)	5	1	OE
78	XFORSD01NLM194N	Reserve Assets	Special Drawing Rights	5	0	OE
79	LCEAMN01_IXOBSANL	Average Hourly Earnings Manufacturing	Index (2010=100)	6	1	OE
80	IRLTLT01NLM156N	10-Year Bond	Percent	2	0	OE
		SPAIN				
81	UNEMES	Unemployment Rate	Rate	1	1	EC
82	PRINTO01ES	Industrial Production	Index (2010=100)	5	1	OE
83	ODCNPI03ES	Permits Issued for Dwelling	Index (2010=100)	4	1	OE
84	LOLITOAAES	Leading Indicators	Index (2010=100)	5	0	OE
85	XTEXVA01ES	Exports	US Dollar, millions	5	1	OE
86	XTIMVA01ES	Imports	US Dollar, millions	5	1	OE
87	IBEX	Stock Index	Price	5	0	BE
88	GTESP6M	6-Month Bond	Percent	2	0	BE

89	GTESP1Y	1-Year Bond	Percent	2	0	BE
90	GTESP2Y	2-Year Bond	Percent	2	0	BE
91	CPIES	Consumer Price Index	Index (2010=100)	6	0	OE
92	IRLTLT01ESM156N	10-Year Bond	Percent	2	0	OE
		UNITED KINGDOM				
93	UNEMUK	Unemployment Rate	Rate	1	1	EC
94	PRINTO01UK	Industrial Production	Index (2010=100)	5	1	OE
95	LOLITOAAUK	Leading Indicators	Index (2010=100)	5	0	OE
96	IR3TIB01UK	Interest Rate 3-Month	Percentage	2	0	OE
97	XTEXVA01UK	Exports	US Dollar, millions	5	1	OE
98	XTIMVA01UK	Imports	US Dollar, millions	5	1	OE
99	UKX	Stock Index	Price	5	0	BE
100	GUKG6M	6-Month Bond	Percent	2	0	BE
101	GUKG1	1-Year Bond	Percent	2	0	BE
102	GUKG2	2-Year Bond	Percent	2	0	BE
103	CPIUK	Consumer Price Index	Index (2010=100)	6	0	OE
104	LCEAMN01_IXOBSAUK	Hourly Earnings Manufacturing	Index (2010=100)	6	1	OE
105	MABMM201GBM189N	M1 Money Stock	Pounds	6	0	OE
	Variable Name	Title	Unit	T-code	SA	So
106	MABMM301GBM189N	M2 Money Stock	Pounds	6	0	OE
107	MANMM101GBM189N	M3 Money Stock	Pounds	6	0	OE
108	IRLTLT01UKM156N	10-Year Bond	Percent	2	0	OE
		UNITED STATES OF AMERICA				
109	UNEMUS	Unemployment rate	Rate	1	1	BL

110	AAA	Moody's Seasoned AAA Corporate Bond Yield	Percent	2	0	FR
111	AMBSL	St. Louis Adjusted Money Base	Billion US Dollars	6	0	FR
112	BAA	Moody's Seasoned BAA Corporate Bond Yield	Percent	2	0	FR
113	BUSINV	Total Business Inventories	Million US Dollars	5	1	FR
114	BUSLOANS	Commercial and Industry Loans, All Commercial Banks	Billion US Dollars	5	1	FR
115	CMRMTSPL	Real Manufacturing and Trade Industries Sales	Millions Chained 2009 Dollars	5	1	FR
116	CPF3M	3-Month AA Financial Commercial Paper Rate	Percent	2	1	FR
117	RPCE	Real Personal Consumption Expenditures	Index 2009=100	5	1	FR
118	FEDFUNDS	Effective Federal Funds Rate	Percent	2	0	FR
119	GS1	1-Year Treasury Constant Maturity Rate	Percent	2	0	FR
120	GS5	5-Year Treasury Constant Maturity Rate	Percent	2	0	FR
121	HOUST	Housing Starts: Total: New Privately Owned Housing Units	1000s	4	1	FR
122	ICLAIM	Initial Claims (unemployment)	Number	5	0	FR
123	INDPRO	Industrial Production Index	Index 2007=100	5	1	FR
124	IPFINAL	Industrial Production: Final Products	Index 2007=100	5	1	FR
125	M1SL	M1 Money Stock	Billion US Dollars	6	1	FR
126	M2SL	M2 Money Stock	Billion US Dollars	6	1	FR
127	M3US	M3 Money Stock	Billion US Dollars	6	1	FR
128	NONREVSL	Total Nonrevolving Credit Owned and Securitized, Outstanding	Billion US Dollars	5	1	FR
129	PERMIT	New Private Housing Units Authorized by Building Permits	Thousands of Units	4	1	FR
130	PPIFGS	Producer Price Index: Finished Goods	Index 1982=100	6	1	FR
131	REALLN	Real Estate Loans, All Commercial Banks	Billions of U.S. Dollars	5	1	FR
132	RPI	Real Personal Income	Billions of Chained 2009 Dollars	5	1	FR
133	TB3MS	3-Month Treasury Bill: Secondary Market Rate	Percent	2	0	FR
134	TB6MS	6-Month Treasury Bill: Secondary Market Rate	Percent	2	0	FR
135	TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies	Index March 1973=100	5	1	FR

136	UEMP15T26	Number of Civilians Unemployed for 15 to 26 Weeks	Thousands of Persons	5	1	FR
137	UEMP27OV	Number of Civilians Unemployed for 27 Weeks and Over	Thousands of Persons	5	1	FR
138	UEMP5TO14	Number of Civilians Unemployed for 5 to 14 Weeks	Thousands of Persons	5	1	FR
139	UEMPLT5	Number of Civilians Unemployed for under 5 Weeks	Thousands of Persons	5	1	FR
140	UEMPMEAN	Mean Duration of Unemployment	Weeks	1	1	FR
141	UMCSENT	Consumer Sentiment, University of Michigan	Index 1st Quarter 1966=100	5	1	FR
142	CPIUS	Consumer Price Index	Index (2010=100)	6	0	FR
143	LOLITOAAUS	Leading Indicators	Index (2010=100)	5	0	OE
	Variable Name	Title	Unit	T-code	SA	So
144	XTEXVA01US	Exports	US Dollar, millions	5	1	OE
145	XTIMVA01US	Imports	US Dollar, millions	5	1	OE
146	IPS11	Industrial Production Index - Products, Total	Index (2000=100)	5	0	S&
147	PMP	NAPM Production Index	Index (between 0 and 100)	2	0	S&
148	PI071	Personal Consumption Expenditures	Index (2000=100)	6	0	S&
149	FSDXP	S&P'S Composite Common Stock: Dividend Yield	Percent per annum	2	0	S&
150	FSPXE	S&P'S Composite Common Stock: Price-Earnings Ratio	Percent	2	0	S&
151	PMI	Purchasing Managers' Index	Index (between 0 and 100)	2	1	S&
152	PMNO	NAPM New Orders Index	Index (between 0 and 100)	2	0	S&
153	MSONDQ	New Orders, Non-Defense Capital Goods	400s 1996 US Dollar	5	0	S&
154	LCEAPR_IXOBSAUS	Hourly Earnings Private Sector	Index (2010=100)	6	1	OE
155	IRLTLT01USM156N	10-Year Bond	Percent	2	0	OE
156	SP500	S&P 500 index	Price	5	0	BE
		EXCHANGE RATES				
157	USDEUR	US Dollar / Euro	Curr Rate	5	0	BE
158	GBPUSD	GB Pound / US Dollar	Curr Rate	5	0	BE

159	USDCHF	US Dollar / Swiss Franc	Curr Rate	5	0	BE
160	EURGBP	Euro / GB Pound	Curr Rate	5	0	BE
161	USDCAD	US Dollar / Canadian Dollar	Curr Rate	5	0	BE
162	USDDEM	US Dollar / Deutsche Mark (euro-altered after 2001)	Curr Rate	5	0	BE
		GROSS DOMESTIC PRODUCT				
163	GDPFR	Quarterly GDP Growth France	Growth Rates	7	0	OE
164	GDPUK	Quarterly GDP Growth United Kingdom	Growth Rates	7	0	OE
165	GDPDE	Quarterly GDP Growth Germany	Growth Rates	7	0	OE
166	GDPNL	Quarterly GDP Growth Netherlands	Growth Rates	7	0	OE
167	GDPES	Quarterly GDP Growth Spain	Growth Rates	7	0	OE
168	GDPIT	Quarterly GDP Growth Italy	Growth Rates	7	0	OE
169	GDPUS	Quarterly GDP Growth United States of America	Growth Rates	7	0	OE
		COMMODITIES				
170	SPGSCI	Goldman Sachs Commodities Inde(GSCI)	Price	5	0	BE
171	BCOM	BB Commodity Index	Price	5	0	BE
172	CRY	Thomson Reuters CRB Index	Price	5	0	BE
173	CNXMET	NSE CNX Metal Index	Price	5	0	BE
174	CL1 COMB	US Crude Oil Futures	Price	5	0	BE

6.1.1 EM-algorithm for PCA factor construction

Let us denote $\hat{X}^{\{j\}}$ is the j-th iteration estimate of the full data. X is the entire observed data-set with missing values.

1. Create an initial (naive) estimate of $\hat{X}^{\{0\}}$. Such that $\hat{X}^{\{0\}}_{t,i} = X_{t,i}$ if observed. We set the normal missing observations equal to the unconditional mean of their respective series. We set the missing observations for the monthly GDP growth rate equal to the quarterly GDP growth rate divided by 3 for those columns i that are the GDP series:

$$\hat{X}^{\{0\}}_{t,i} = X^{q}_{\tau,i}/3$$

   where $X^{q}_{\tau,i}$ is the observed quarterly growth series, and $\tau = 3$ for $t = 1, 2, 3$, $\tau = 6$ for $t = 4, 5, 6$, etc. We calculate the first h estimated principal components $\hat{F}^{\{0\}}$ and their loadings $\hat{\Lambda}^{\{0\}}$ with PCA on $\hat{X}^{\{0\}}$.

2. E-step: For iteration j we use the estimates $\hat{F}^{\{j-1\}}$ and $\hat{\Lambda}^{\{j-1\}}$ from the last iteration to update our estimate of the data $\hat{X}^{\{j\}}$. Where:

$$\hat{X}^{\{j\}}_{t,i} = X_{t,i} \quad \text{If } X_{t,i} \text{ observed}$$

$$\hat{X}^{\{j\}}_{t,i} = \hat{\lambda}^{\{j-1\}\prime}_{i} \hat{F}^{\{j-1\}}_{t} \quad \text{If } X_{t,i} \text{ is a normal missing value}$$

$$\hat{X}^{\{j\}}_{t,i} = \hat{\lambda}^{\{j-1\}\prime}_{i} \hat{F}^{\{j-1\}}_{t} + \hat{e}_{t,i} \quad \text{If } i \text{ is a GDP series}$$

$$\text{where:} \quad \hat{e}_{t,i} = \left(X^{q}_{\tau,i} - \hat{\lambda}^{\{j-1\}\prime}_{i} (\hat{F}^{\{j-1\}}_{\tau} + \hat{F}^{\{j-1\}}_{\tau-1} + \hat{F}^{\{j-1\}}_{\tau-2}) \right) /3$$

$$\tau = 3 \text{ for } t = 1, 2, 3,\ \tau = 6 \text{ for } t = 4, 5, 6, \text{ etc.}$$

3. M-step: The updated $\hat{X}^{\{j\}}$ is used to calculate the new estimates for $\hat{F}^{\{j\}}$ and $\hat{\Lambda}^{\{j\}}$ via PCA. The algorithm goes back to (2).

This is repeated until the change in factor values per iteration is below some threshold value, or the algorithm exceeds some preset number of iterations. The factors from the last iteration are taken as the estimated factors: $\hat{F} = \hat{F}^{\{J\}}$.

A big advantage of the robustness of the above EM-algorithm in creating the factors is that we make use of the underlying factor structure of the data to estimate the missing values. This takes cross-sectional as well time interdependence into account. This as opposed to, for example, a simple interpolation of the values which only take the observations close to the missing value of the same series into account.

6.2 Factors' characteristics

6.2.1 Factors' correlograms

Below we monitor the absolute values of the cross-autocorrelations of both set of factors that we are using. The insignificant correlation have been set to zero.

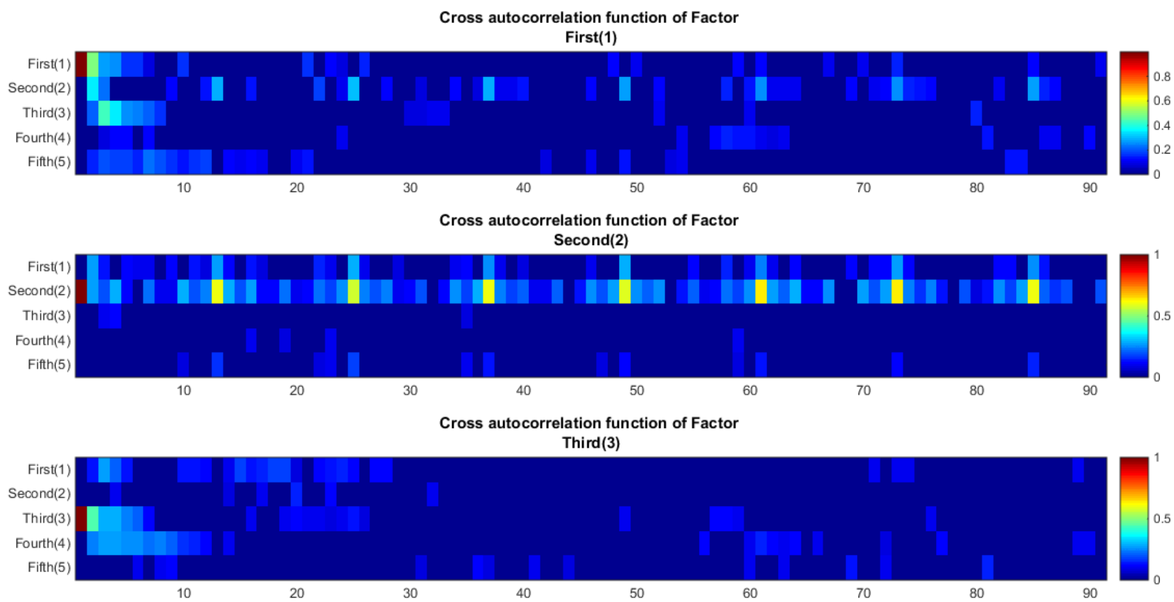


Figure 8: Autocorrelation function of S&W factors. Each figure plots the absolute values of cross autocorrelations between a factor highlighted and all 5 $S\&W$ factors. The correlations have been corrected for significance, i.e. a (cross)correlation of 0 indicates that the value was not significantly different from 0.

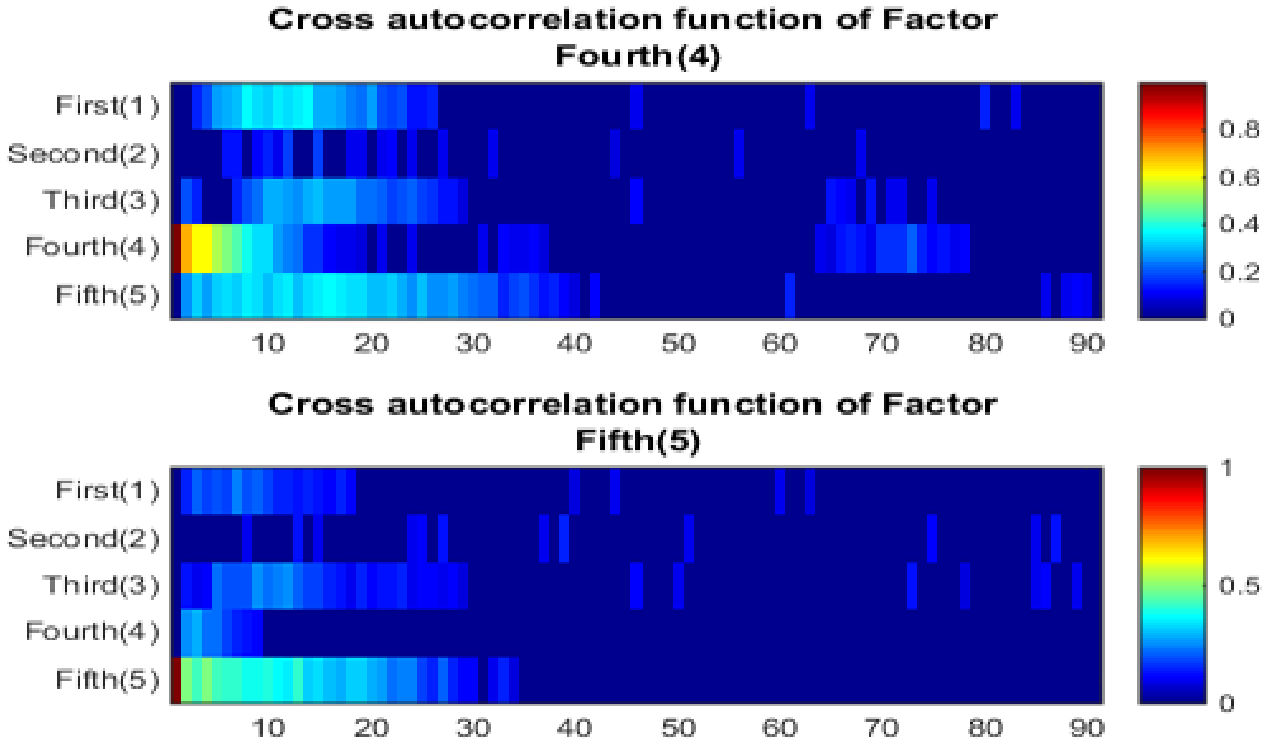


Figure 9: Autocorrelation function of S&W factors. Each figure plots the absolute values of cross autocorrelations between a factor highlighted and all 5 $S\&W$ factors. The correlations have been corrected for significance, i.e. a (cross)correlation of 0 indicates that the value was not significantly different from 0.

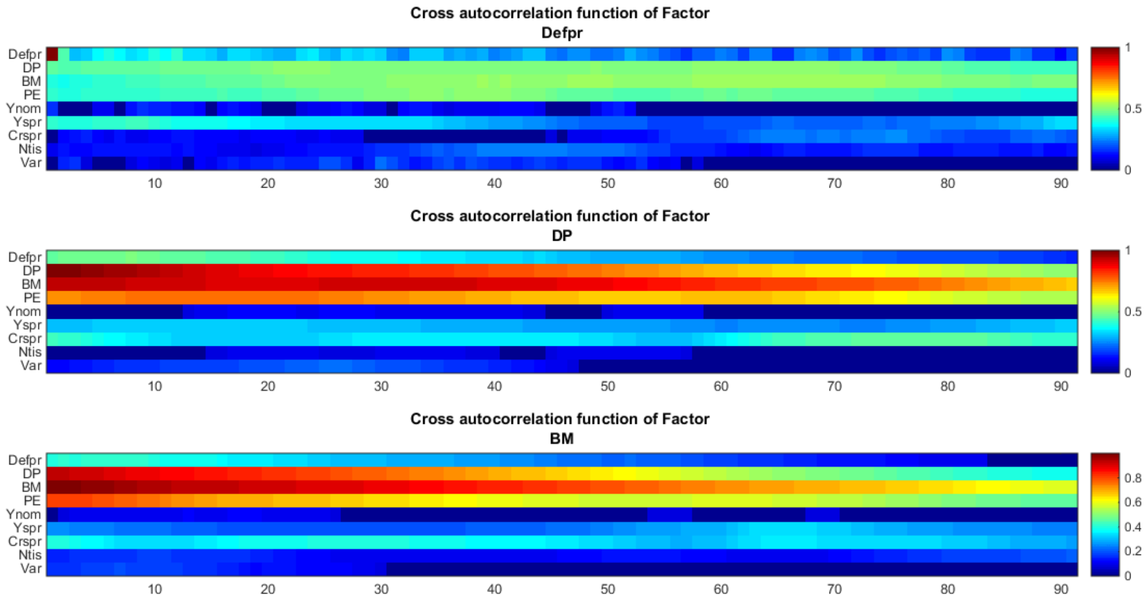


Figure 10: Autocorrelation function of G&W factors. Each figure plots the absolute values of cross autocorrelations between a factor highlighted and all 5 $S\&W$ factors. The correlations have been corrected for significance, i.e. a (cross)correlation of 0 indicates that the value was not significantly different from 0.

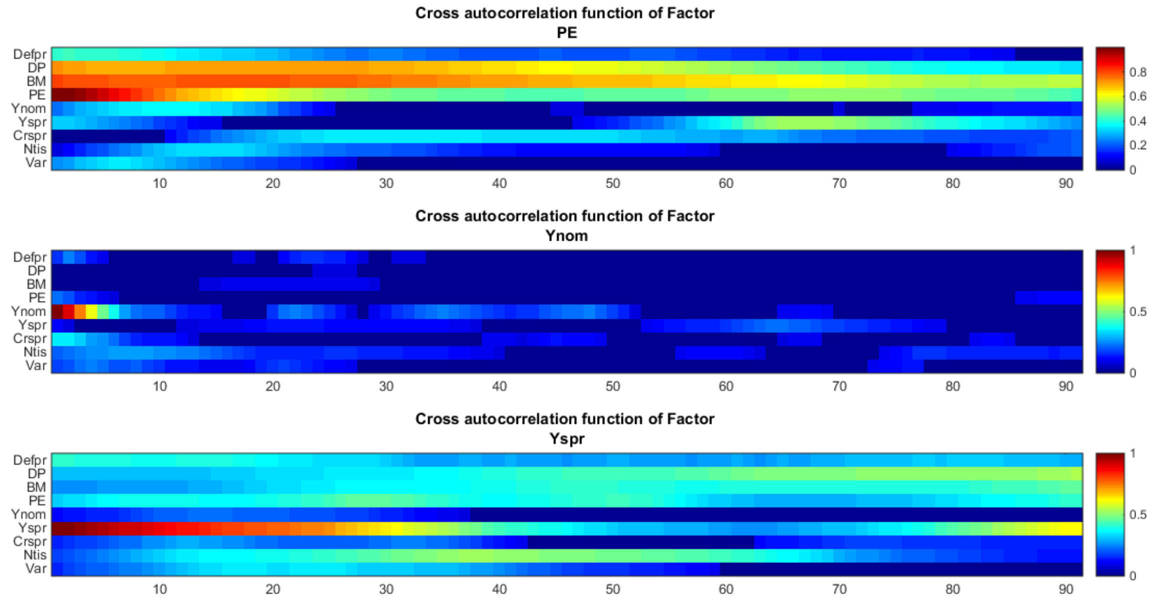


Figure 11: Autocorrelation function of G&W factors. Each figure plots the absolute values of cross autocorrelations between a factor highlighted and all 5 $S\&W$ factors. The correlations have been corrected for significance, i.e. a (cross)correlation of 0 indicates that the value was not significantly different from 0.

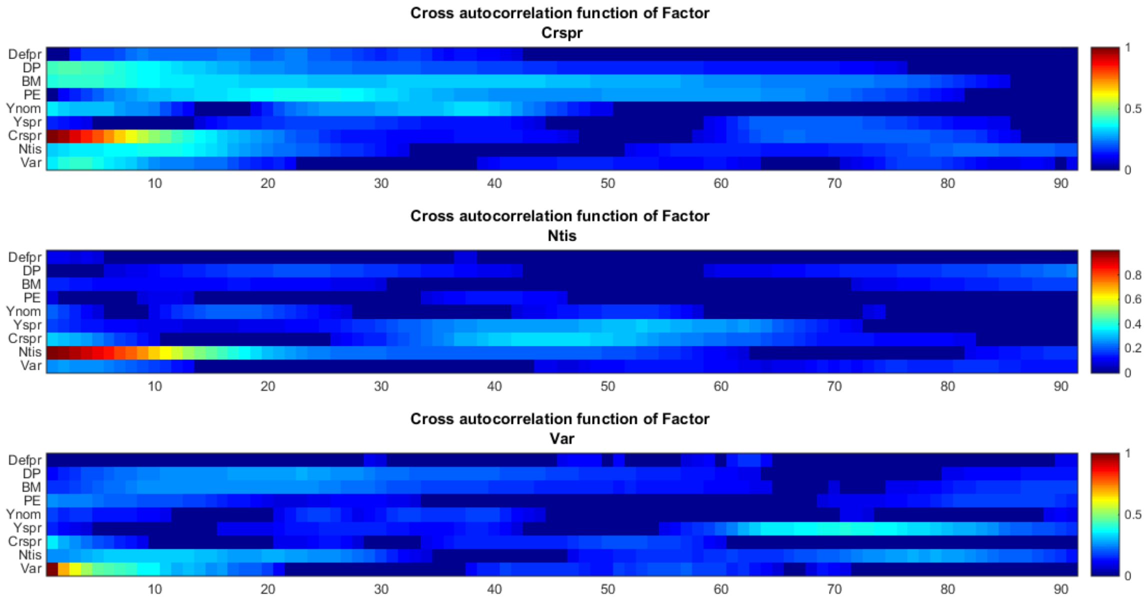


Figure 12: Autocorrelation function of G&W factors. Each figure plots the absolute values of cross autocorrelations between a factor highlighted and all 5 $S\&W$ factors. The correlations have been corrected for significance, i.e. a (cross)correlation of 0 indicates that the value was not significantly different from 0.

6.2.2 BIC and log-likelihood values for various lags

Below we monitor the BIC values and the log-likelihood values of (3.4) and (3.5) for several amount of lags p and q and for both set of factors.

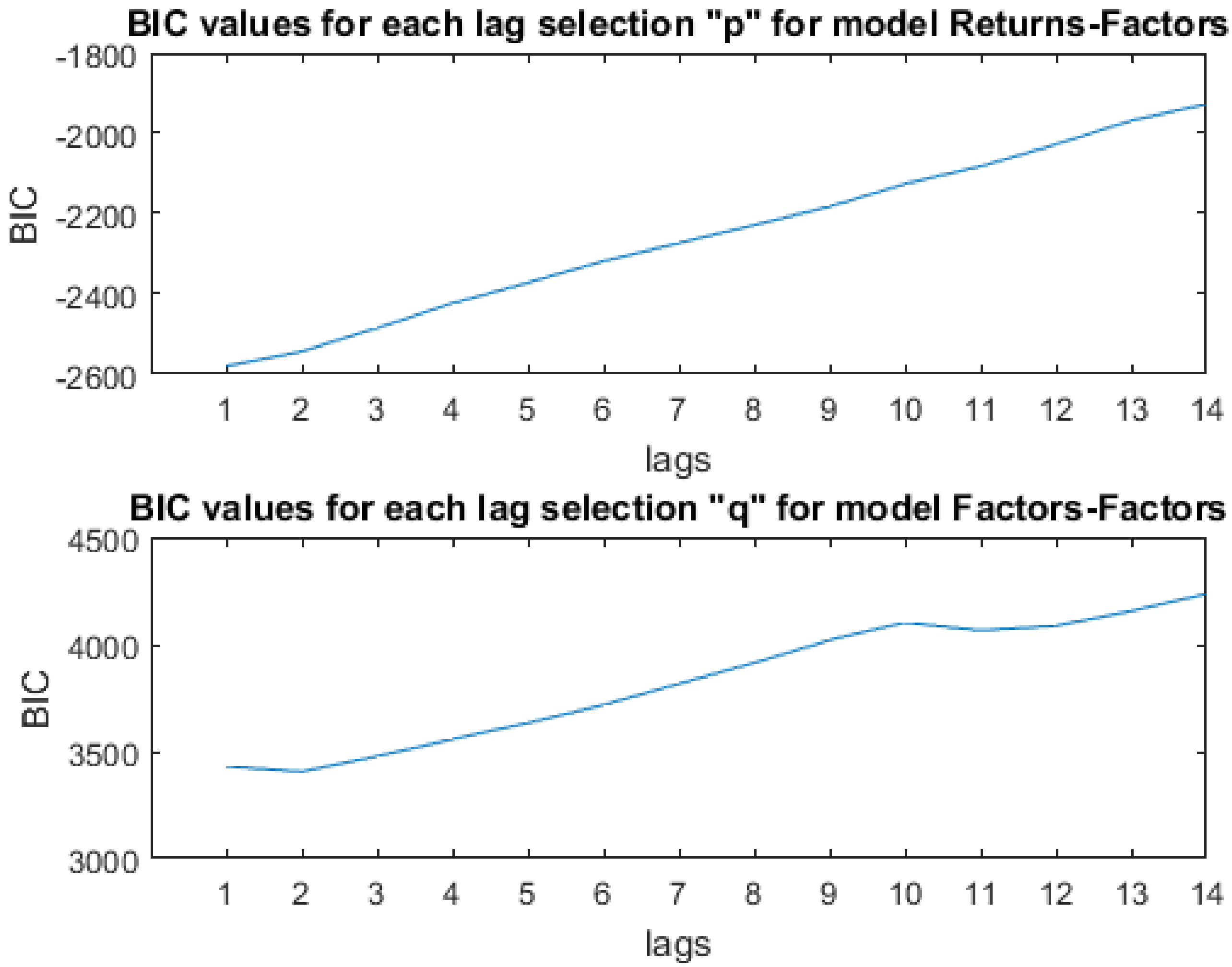


Figure 13: BIC values for models (3.4) & (3.5) of S&W factors. Each figure plots the Bayesian information criterion statistics for a given amount of lag that correspond to p and q on Equations (3.4) & (3.5) respectively.

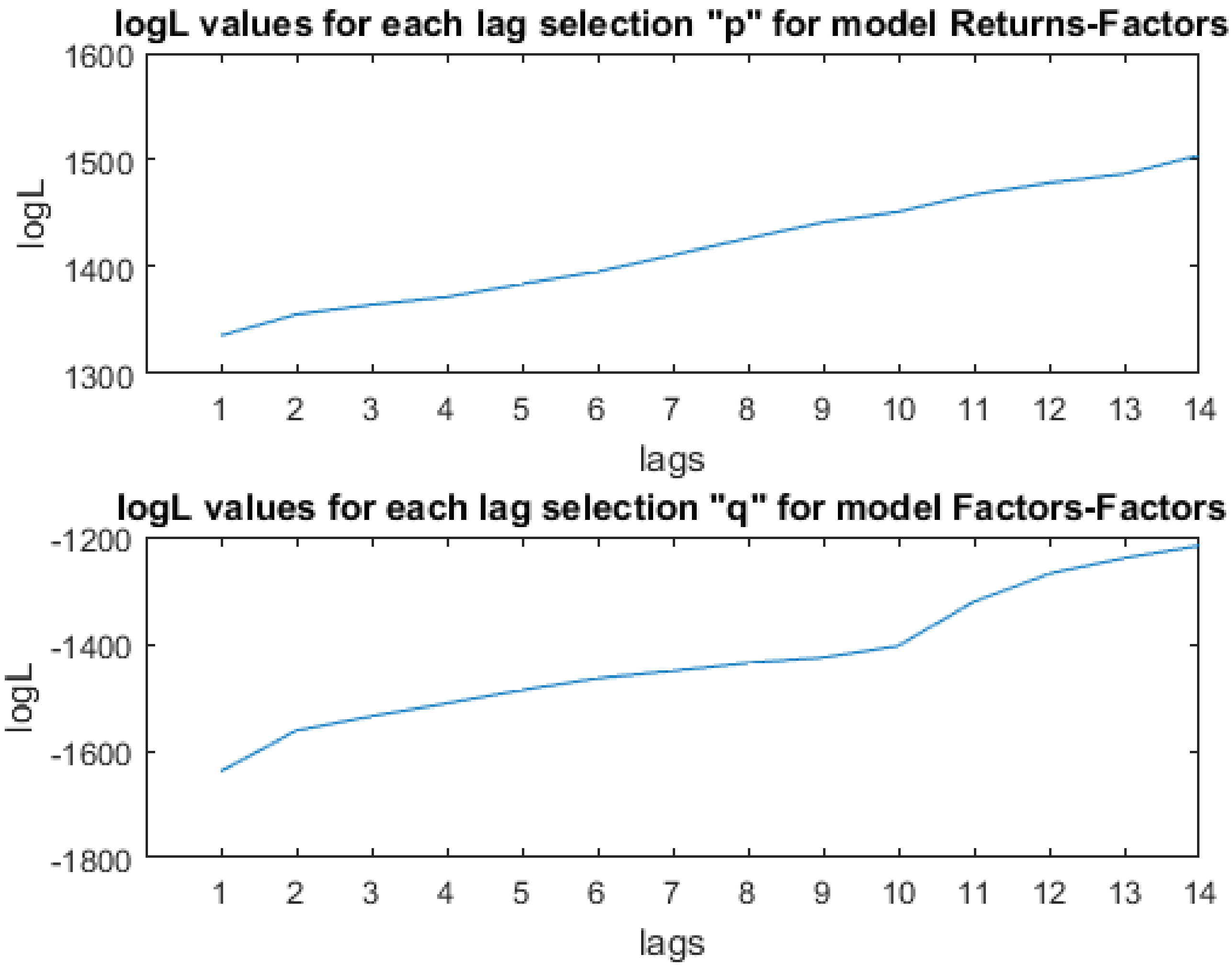


Figure 14: Log-likelihood values for models (3.4) & (3.5) of S&W factors. Each figure plots the Bayesian information criterion statistics for a given amount of lag that correspond to p and q on Equations (3.4) & (3.5) respectively.

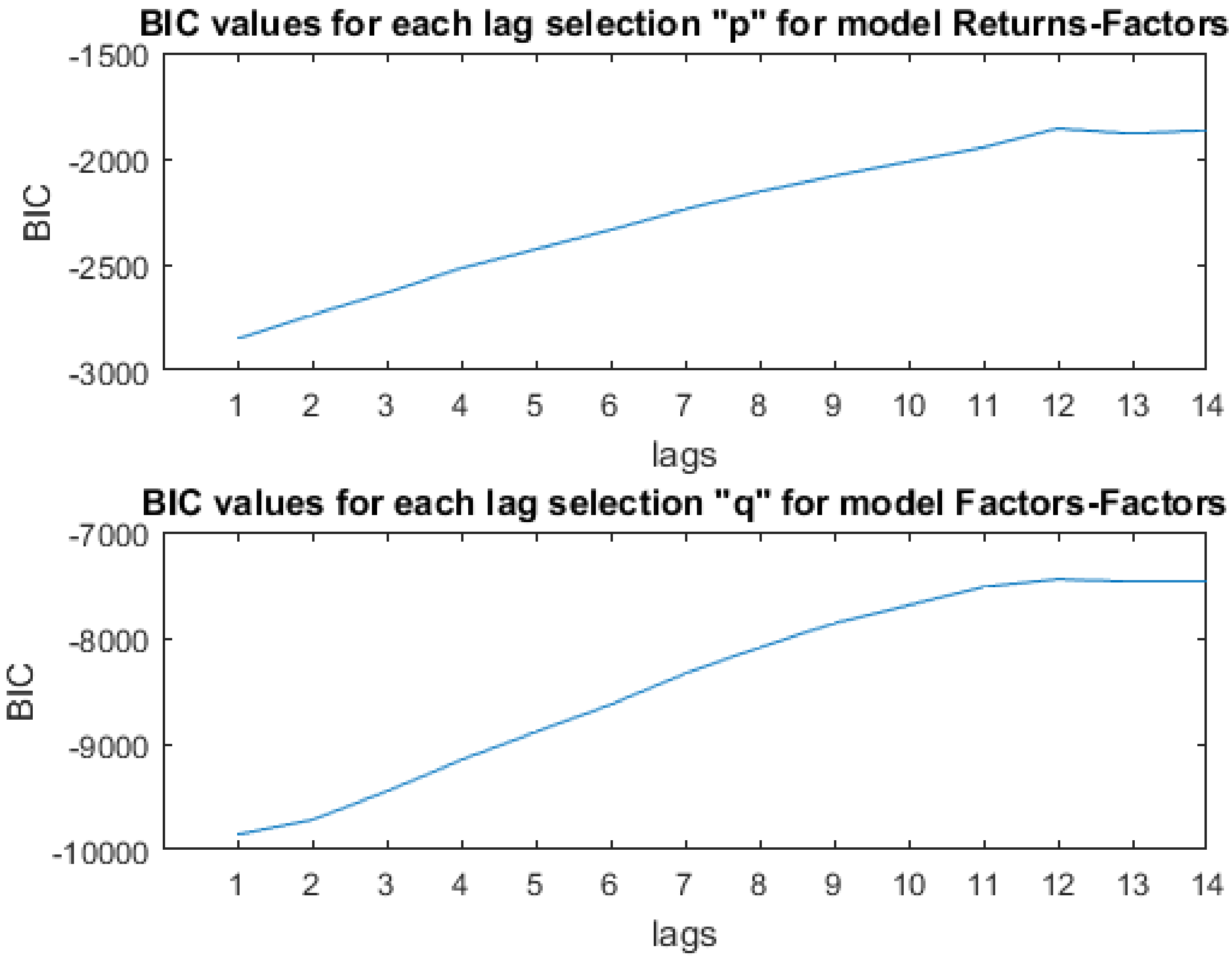


Figure 15: BIC values for models (3.4) & (3.5) of G&W factors. Each figure plots the Bayesian information criterion statistics for a given amount of lag that correspond to p and q on Equations (3.4) & (3.5) respectively.

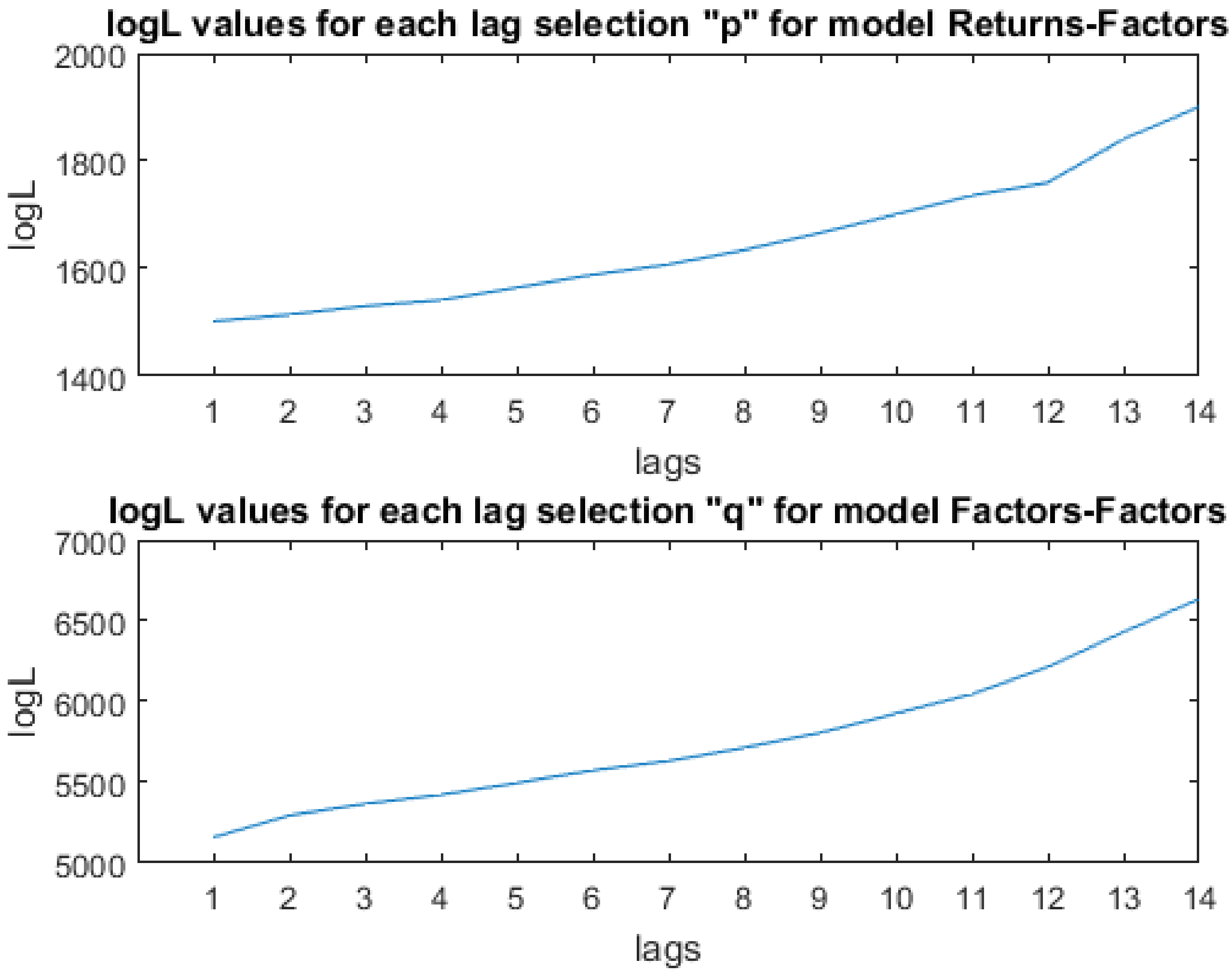


Figure 16: Log-likelihood values for models (3.4) & (3.5) of G&W factors. Each figure plots the Bayesian information criterion statistics for a given amount of lag that correspond to p and q on Equations (3.4) & (3.5) respectively.

6.3 Vector Autoregressive - Frequentist Plug-in Approach with simulation techniques and portfolio constrains -Ignore parameter uncertainty

Using the set-up Equation (3.7), we estimate $\hat{C}$,$\hat{\Phi}$ and $\hat{\Sigma}$. That is performed in any point of time by a looking back period of several months. In other words, we are using a rolling window to estimate those parameters. Then, we simulate N paths of h periods for asset returns and predictor variables. The algorithm of this iterated process is described below. We repeat N times the following procedure.

- Assume that we are in the i^{th} iteration.
- Draw ${\varepsilon^i}_{t+1}$ from $N(0, \hat{\Sigma})$ and set: ${z^i}_{t+1} = \hat{C} + \hat{\Phi} z_t + {\varepsilon^i}_{t+1}$
- Draw ${\varepsilon^i}_{t+2}$ from $N(0, \hat{\Sigma})$ and set: ${z^i}_{t+2} = \hat{C} + \hat{\Phi} {z^i}_{t+1} + {\varepsilon^i}_{t+2}$
- Keep on in the same logic until you come to the K^{th} iteration.
- Draw ${\varepsilon^i}_{t+h}$ from $N(0, \hat{\Sigma})$ and set: ${z^i}_{t+h} = \hat{C} + \hat{\Phi} {z^i}_{t+h-1} + {\varepsilon^i}_{t+h}$

Having these simulated paths of z_{t+1},z_{t+2},...,z_{t+h}, we isolate and use only the Monte Caro simulations of r_{t+1},r_{t+2},...,r_{t+h} [5].

6.4 Additional statistical results

6.4.1 Additional statistical results for $S\&W$ factors

Table 33: BVAR posterior means of covariance matrix Σ

	r_1	r_2	r_3	**factor(1)**	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
r_1	1.07E-05	5.83E-06	7.71E-06	-1.82E-03	1.11E-03	3.93E-04	-7.19E-04	8.41E-04
r_2	5.83E-06	1.78E-03	6.17E-05	6.65E-03	8.05E-03	4.22E-03	-1.13E-02	7.92E-03
r_3	7.71E-06	6.17E-05	6.98E-04	-2.14E-02	-1.17E-02	1.97E-02	-6.30E-05	7.37E-03
factor(1)	-1.82E-03	6.65E-03	-2.14E-02	4.99E+00	-4.32E-01	-1.40E+00	2.19E-01	-8.25E-01
factor(2)	1.11E-03	8.05E-03	-1.17E-02	-4.32E-01	5.89E+00	-5.74E-01	4.89E-02	-4.21E-01
factor(3)	3.93E-04	4.22E-03	1.97E-02	-1.40E+00	-5.74E-01	4.68E+00	1.40E+00	-2.15E-01
factor(4)	-7.19E-04	-1.13E-02	-6.30E-05	2.19E-01	4.89E-02	1.40E+00	2.01E+00	-1.26E+00
factor(5)	8.41E-04	7.92E-03	7.37E-03	-8.25E-01	-4.21E-01	-2.15E-01	-1.26E+00	2.26E+00

This table presents the full sample posterior means of the covariance matrix of errors that correspond to Equation (3.14), for the S&W factors. These results are based on 100,000 draws from the posterior distribution.

[5] Bare in mind that $z_t = \begin{pmatrix} r_t \\ F_t \\ F_{t-1} \\ \vdots \\ F_{t-p+1} \end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} \alpha \\ \mu \\ \vdots \\ 0 \end{pmatrix}$

Table 34: BVAR-VS posterior means of covariance matrix Σ

	r_1	r_2	r_3	**factor(1)**	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
r_1	1.15E-05	7.84E-06	9.34E-06	-2.31E-03	1.22E-03	5.78E-04	-6.77E-04	8.29E-04
r_2	7.84E-06	2.19E-03	1.02E-04	9.62E-03	9.93E-03	1.06E-02	-1.25E-02	9.91E-03
r_3	9.34E-06	1.02E-04	8.92E-04	-2.39E-02	-1.24E-02	2.23E-02	-1.35E-03	8.65E-03
factor(1)	-2.31E-03	9.62E-03	-2.39E-02	7.50E+00	-2.32E-01	-9.66E-01	2.30E-01	-7.60E-01
factor(2)	1.22E-03	9.93E-03	-1.24E-02	-2.32E-01	7.06E+00	-3.63E-01	1.34E-01	-3.15E-01
factor(3)	5.78E-04	1.06E-02	2.23E-02	-9.66E-01	-3.63E-01	6.12E+00	1.43E+00	-7.99E-02
factor(4)	-6.77E-04	-1.25E-02	-1.35E-03	2.30E-01	1.34E-01	1.43E+00	2.31E+00	-1.35E+00
factor(5)	8.29E-04	9.91E-03	8.65E-03	-7.60E-01	-3.15E-01	-7.99E-02	-1.35E+00	2.60E+00

This table presents the full sample posterior means of the covariance matrix of errors that correspond to Equation (3.16), for the S&W factors. These results are based on 100,000 draws from the posterior distribution.

Table 35: Mean squared prediction errors for all models and all horizons $S\&W$ factors

horizons	**Models**	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**factor(1)**	**factor(2)**	**factor(3)**	**factor(4)**	**factor(5)**
h=1	BVARR	0.0000152	0.0018515	0.0008121	2.4780628	3.0109949	3.9033522	1.8893461	2.2836298
	BVARRVS	0.0000156	0.0018385	0.0008835	3.0026844	3.1837936	4.0026573	2.0394749	2.4133514
	BVARM	-	-	-	-	-	-	-	-
	BVARMVS	-	-	-	-	-	-	-	-
h=4	BVARR	0.0001391	0.0082027	0.0035680	5.4688908	3.3459556	6.4259593	2.9586063	3.8047037
	BVARRVS	0.0001370	0.0084277	0.0036218	5.8471038	3.3252292	6.6118076	3.2140365	3.9425313
	BVARM	0.0001405	0.0086155	0.0037169	5.3035675	3.3936630	6.5356235	2.8070079	3.6000967
	BVARMVS	0.0001371	0.0086438	0.0036561	5.7645616	3.3657758	6.7355987	3.2954611	3.7956861
h=6	BVARR	0.0002323	0.0148404	0.0056497	6.1827698	3.3991290	7.2272588	3.7139339	4.4040754
	BVARRVS	0.0002306	0.0158140	0.0057200	6.1853073	3.3759299	7.3493821	3.9043005	4.4227180
	BVARM	0.0002340	0.0160349	0.0060685	6.5754805	3.4707043	7.6889509	3.5102038	4.1976025
	BVARMVS	0.0002297	0.0167655	0.0059442	6.5076339	3.4355100	7.9443661	4.0480379	4.2791126
h=12	BVARR	0.0005845	0.0382027	0.0086883	6.1698208	3.4274169	7.4126907	4.6647947	5.2620621
	BVARRVS	0.0005923	0.0409698	0.0082189	6.0639807	3.3986432	7.4465993	4.6526211	5.1598495
	BVARM	0.0005824	0.0434203	0.0105319	6.6385844	3.4647733	7.7967946	4.4483709	5.3181260
	BVARMVS	0.0005815	0.0469437	0.0092800	6.3911363	3.4612596	7.9232936	4.7498744	5.1757681
h=24	BVARR	0.0023390	0.0879081	0.0112774	6.3102661	3.3580349	7.6973063	5.1551257	5.7961327
	BVARRVS	0.0023879	0.0945354	0.0098383	6.2995256	3.3529847	7.7144725	5.0989580	5.7616644
	BVARM	0.0023592	0.0978627	0.0147770	6.4095077	3.3593750	7.8004223	4.8905037	6.0777417
	BVARMVS	0.0023615	0.1081403	0.0116655	6.4335114	3.3711560	7.9382013	4.7774244	5.9993214

This table presents the mean squared prediction errors of all of our models for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the S&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

6.4.2 Additional statistical results for $G\&W$ factors

Table 36: BVAR posterior mean of covariance matrix Σ

	r_1	r_2	r_3	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
r_1	9.07E-06	5.13E-06	1.51E-05	-1.75E-06	-5.59E-06	-2.15E-05	1.54E-05	4.82E-07	-1.77E-06	3.58E-07	-1.19E-07	-1.59E-04
r_2	5.13E-06	1.81E-03	1.40E-04	1.23E-04	-1.81E-03	-1.64E-03	1.60E-03	-2.06E-05	5.82E-06	-8.97E-07	-1.15E-05	-8.45E-03
r_3	1.51E-05	1.40E-04	7.80E-04	-1.78E-04	-1.42E-04	-2.02E-04	3.22E-04	-3.36E-05	-3.46E-05	6.70E-06	-1.17E-06	9.20E-04
Defpr	-1.75E-06	1.23E-04	-1.78E-04	1.97E-04	-1.24E-04	-5.67E-05	3.97E-05	2.57E-06	1.23E-05	-7.81E-07	5.88E-06	-1.74E-03
DP	-5.59E-06	-1.81E-03	-1.42E-04	-1.24E-04	1.84E-03	1.65E-03	-1.56E-03	1.90E-05	-3.95E-06	7.65E-07	1.12E-05	8.37E-03
BM	-2.15E-05	-1.64E-03	-2.02E-04	-5.67E-05	1.65E-03	3.39E-03	-1.56E-03	2.03E-05	-1.98E-06	-4.69E-07	3.45E-05	5.88E-03
PE	1.54E-05	1.60E-03	3.22E-04	3.97E-05	-1.56E-03	-1.56E-03	3.47E-03	-1.86E-05	-9.39E-06	7.29E-06	-6.51E-06	-7.51E-03
Ynom	4.82E-07	-2.06E-05	-3.36E-05	2.57E-06	1.90E-05	2.03E-05	-1.86E-05	1.29E-05	-1.03E-05	-8.11E-07	1.12E-06	-2.41E-06
Yspr	-1.77E-06	5.82E-06	-3.46E-05	1.23E-05	-3.95E-06	-1.98E-06	-9.39E-06	-1.03E-05	1.45E-05	2.28E-07	-8.85E-07	-5.67E-05
Crspr	3.58E-07	-8.97E-07	6.70E-06	-7.81E-07	7.65E-07	-4.69E-07	7.29E-06	-8.11E-07	2.28E-07	8.51E-07	-5.76E-08	5.94E-05
Ntis	-1.19E-07	-1.15E-05	-1.17E-06	5.88E-06	1.12E-05	3.45E-05	-6.51E-06	1.12E-06	-8.85E-07	-5.76E-08	1.57E-05	2.17E-05
Var	-1.59E-04	-8.45E-03	9.20E-04	-1.74E-03	8.37E-03	5.88E-03	-7.51E-03	-2.41E-06	-5.67E-05	5.94E-05	2.17E-05	3.87E-01

This table presents the full sample posterior means of the covariance matrix of errors that correspond to Equation (3.14), for the G&W factors. These results are based on 100,000 draws from the posterior distribution.

Table 37: BVAR-VS posterior mean of covariance matrix Σ

	r_1	r_2	r_3	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
r_1	2.78E-05	8.27E-06	1.71E-05	-2.42E-06	-1.07E-05	-2.34E-05	2.81E-05	-1.76E-06	-8.81E-07	4.97E-07	1.40E-07	1.13E-04
r_2	8.27E-06	2.12E-03	1.35E-04	1.26E-04	-1.82E-03	-1.64E-03	1.63E-03	-3.34E-05	8.85E-06	-1.36E-06	-8.92E-06	-9.89E-03
r_3	1.71E-05	1.35E-04	1.00E-03	-1.91E-04	-1.59E-04	-2.12E-04	3.51E-04	-5.29E-05	-3.35E-05	7.06E-06	-2.14E-06	3.00E-03
Defpr	-2.42E-06	1.26E-04	-1.91E-04	2.87E-04	-1.20E-04	-5.20E-05	2.54E-05	7.67E-06	1.21E-05	-1.00E-06	6.47E-06	-3.47E-03
DP	-1.07E-05	-1.82E-03	-1.59E-04	-1.20E-04	2.30E-03	1.67E-03	-1.62E-03	4.78E-05	-7.85E-06	1.32E-06	9.44E-06	7.07E-03
BM	-2.34E-05	-1.64E-03	-2.12E-04	-5.20E-05	1.67E-03	4.20E-03	-1.62E-03	4.01E-05	-4.37E-06	-6.38E-07	3.40E-05	5.20E-03
PE	2.81E-05	1.63E-03	3.51E-04	2.54E-05	-1.62E-03	-1.62E-03	4.50E-03	-1.00E-04	7.88E-06	9.59E-06	-2.86E-06	-2.64E-03
Ynom	-1.76E-06	-3.34E-05	-5.29E-05	7.67E-06	4.78E-05	4.01E-05	-1.00E-04	6.49E-05	-1.47E-05	-9.93E-07	1.76E-07	-9.08E-04
Yspr	-8.81E-07	8.85E-06	-3.35E-05	1.21E-05	-7.85E-06	-4.37E-06	7.88E-06	-1.47E-05	2.13E-05	3.80E-07	-4.91E-07	2.37E-04
Crspr	4.97E-07	-1.36E-06	7.06E-06	-1.00E-06	1.32E-06	-6.38E-07	9.59E-06	-9.93E-07	3.80E-07	1.28E-06	-7.55E-08	1.62E-04
Ntis	1.40E-07	-8.92E-06	-2.14E-06	6.47E-06	9.44E-06	3.40E-05	-2.86E-06	1.76E-07	-4.91E-07	-7.55E-08	2.45E-05	-9.10E-06
Var	1.13E-04	-9.89E-03	3.00E-03	-3.47E-03	7.07E-03	5.20E-03	-2.64E-03	-9.08E-04	2.37E-04	1.62E-04	-9.10E-06	1.04E+00

This table presents the full sample posterior means of the covariance matrix of errors that correspond to Equation (3.16), for the G&W factors. These results are based on 100,000 draws from the posterior distribution.

Table 38: Mean squared prediction errors for all models and all horizons for G & W factors

horizons	**Models**	$cgr_1(t_h)$	$cgr_2(t_h)$	$cgr_3(t_h)$	**Defpr**	**DP**	**BM**	**PE**	**Ynom**	**Yspr**	**Crspr**	**Ntis**	**Var**
h=1	BVARR	1.80E-05	2.13E-03	1.14E-03	3.99E-04	2.16E-03	7.53E-03	8.53E-03	3.59E-06	9.46E-06	1.14E-06	2.57E-05	4.24E-01
	BVARRVS	1.78E-05	2.11E-03	1.12E-03	3.99E-04	2.15E-03	7.56E-03	8.52E-03	3.54E-06	9.47E-06	1.13E-06	2.58E-05	4.12E-01
	BVARM	-	-	-	-	-	-	-	-	-	-	-	-
	BVARMVS	-	-	-	-	-	-	-	-	-	-	-	-
h=4	BVARR	1.71E-04	1.17E-02	4.53E-03	3.88E-04	1.41E-02	3.51E-02	8.69E-02	2.61E-05	5.46E-05	1.10E-05	1.28E-04	8.25E-01
	BVARRVS	1.69E-04	1.15E-02	4.39E-03	3.87E-04	1.37E-02	3.58E-02	8.66E-02	2.48E-05	5.43E-05	1.11E-05	1.34E-04	7.64E-01
	BVARM	1.70E-04	1.17E-02	4.49E-03	3.86E-04	1.39E-02	3.50E-02	8.65E-02	2.58E-05	5.45E-05	1.09E-05	1.26E-04	8.07E-01
	BVARMVS	1.68E-04	1.16E-02	4.33E-03	3.85E-04	1.35E-02	3.58E-02	8.64E-02	2.46E-05	5.43E-05	1.08E-05	1.31E-04	7.55E-01
h=6	BVARR	3.03E-04	2.24E-02	6.77E-03	3.99E-04	2.65E-02	6.01E-02	1.83E-01	3.82E-05	9.54E-05	1.80E-05	1.76E-04	8.77E-01
	BVARRVS	3.02E-04	2.25E-02	6.68E-03	4.03E-04	2.64E-02	6.12E-02	1.80E-01	3.74E-05	9.51E-05	1.83E-05	1.91E-04	8.17E-01
	BVARM	3.01E-04	2.24E-02	6.57E-03	3.97E-04	2.64E-02	5.88E-02	1.81E-01	3.81E-05	9.52E-05	1.82E-05	1.67E-04	8.82E-01
	BVARMVS	3.03E-04	2.25E-02	6.66E-03	3.99E-04	2.65E-02	6.10E-02	1.79E-01	3.74E-05	9.50E-05	1.84E-05	1.85E-04	8.46E-01
h=12	BVARR	6.83E-04	6.00E-02	1.23E-02	4.34E-04	6.72E-02	1.30E-01	5.92E-01	4.92E-05	2.29E-04	3.39E-05	4.49E-04	1.26E+00
	BVARRVS	6.85E-04	6.14E-02	1.22E-02	4.39E-04	7.04E-02	1.36E-01	6.23E-01	4.86E-05	2.34E-04	3.31E-05	4.56E-04	1.22E+00
	BVARM	6.83E-04	5.97E-02	1.31E-02	4.35E-04	6.61E-02	1.39E-01	6.06E-01	5.13E-05	2.28E-04	3.14E-05	4.46E-04	1.23E+00
	BVARMVS	6.85E-04	6.01E-02	1.21E-02	4.45E-04	6.92E-02	1.45E-01	5.93E-01	4.95E-05	2.27E-04	3.22E-05	4.63E-04	1.20E+00
h=24	BVARR	2.91E-03	1.94E-01	2.55E-02	4.55E-04	2.40E-01	2.42E-01	1.39E+00	8.07E-05	4.21E-04	4.23E-05	1.61E-03	1.51E+00
	BVARRVS	2.80E-03	2.37E-01	2.60E-02	4.43E-04	2.54E-01	2.64E-01	1.36E+00	8.22E-05	4.09E-04	4.37E-05	1.78E-03	1.44E+00
	BVARM	2.91E-03	1.83E-01	2.49E-02	4.35E-04	2.40E-01	2.18E-01	1.43E+00	8.49E-05	4.11E-04	4.32E-05	1.58E-03	1.50E+00
	BVARMVS	2.90E-03	2.23E-01	2.53E-02	4.56E-04	2.43E-01	2.34E-01	1.41E+00	8.28E-05	4.10E-04	4.36E-05	2.17E-03	1.44E+00

This table presents the mean squared prediction errors of all of our models for the predictions of Test set observations of the time series of the cumulative gross returns and the factors, where here we are using the G&W factors (also see Table (13) for definition of each model. In order to predict each point in time of the Test set, for each horizon ahead, we use 20,000 draws from the posterior distribution.

6.5 Cumulative portfolio returns and maximum drawdowns

Below, we monitor the cumulative portfolio returns of the two models (with and without variable selection) for each horizon and for each set of factors.

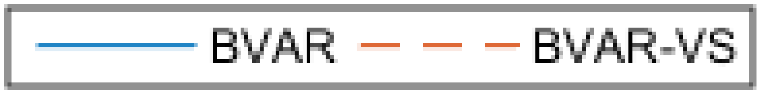


Figure 17:

6.5.1 Cumulative portfolio returns and maximum drawdowns for BVARR and BVARRVS models and for $S\&W$ factors

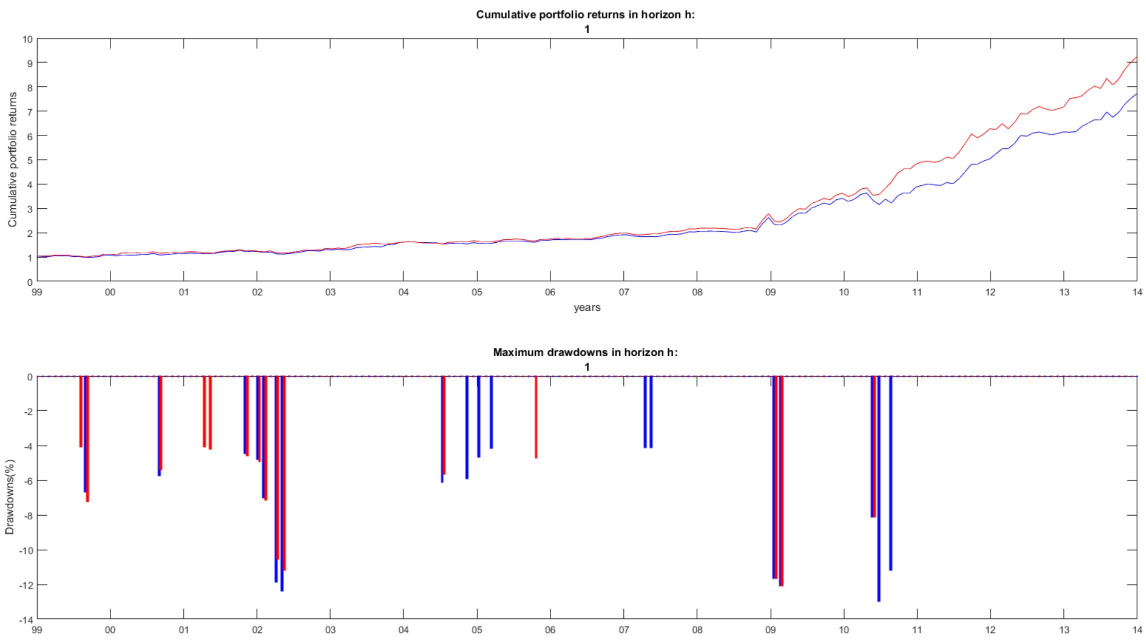


Figure 18: Cumulative portfolio returns with Ridge prior for Equation (3.5) and h=1. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions one month ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

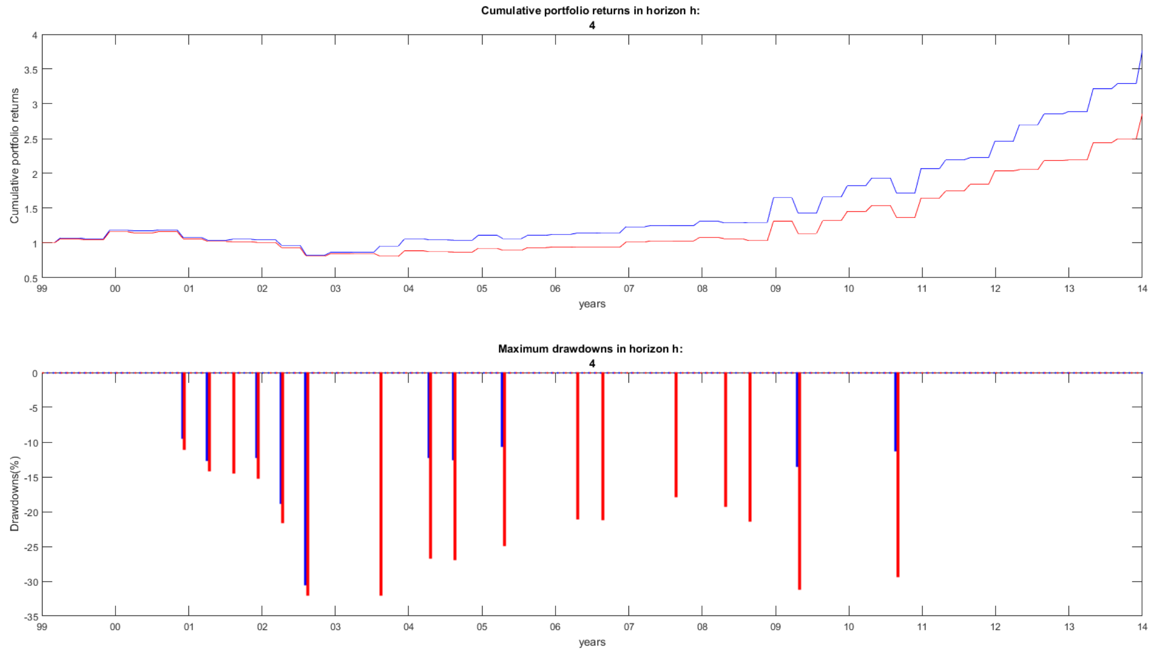


Figure 19: Cumulative portfolio returns with Ridge prior and h=4. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions four months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

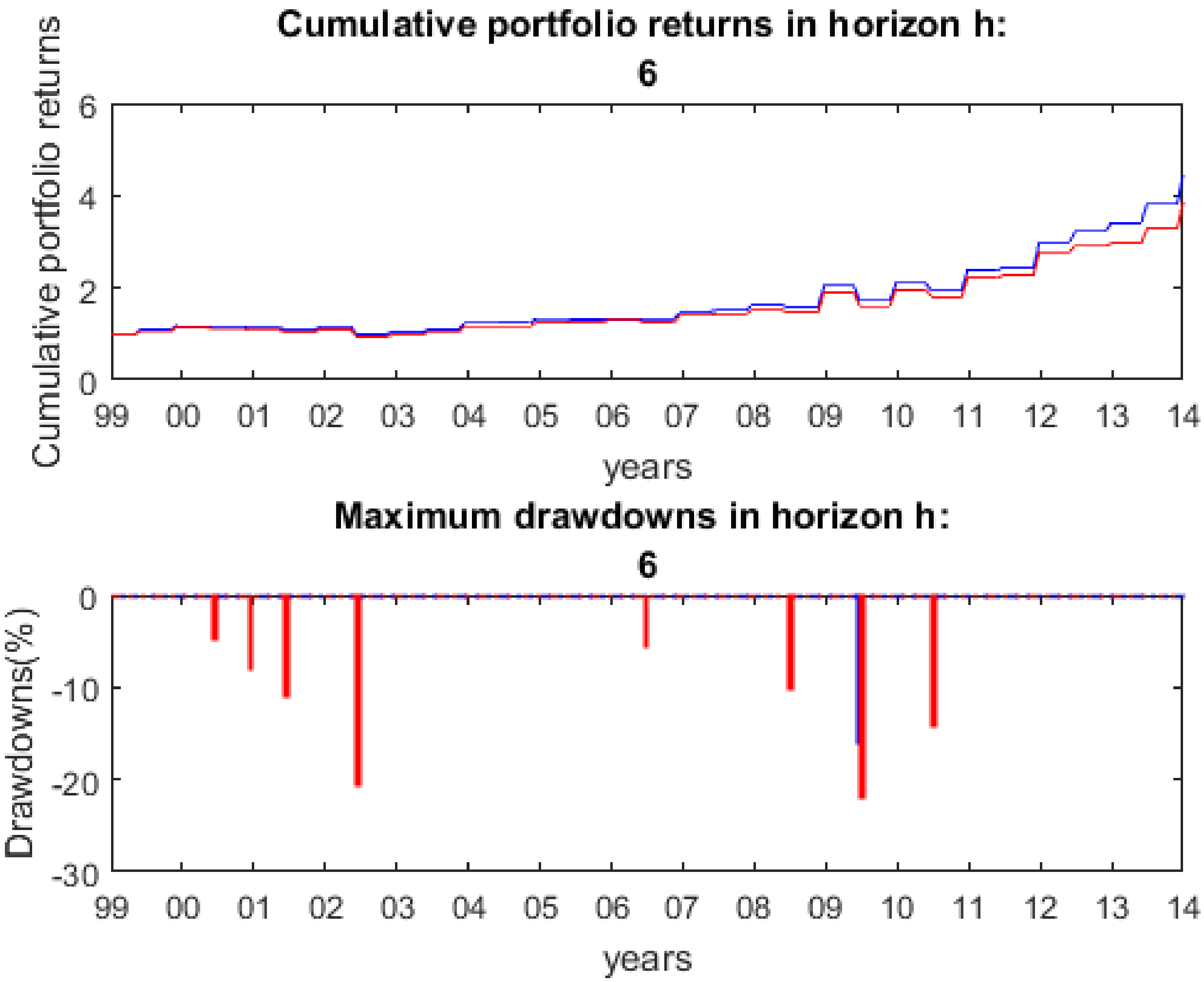


Figure 20: Cumulative portfolio returns with Ridge prior and h=6. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions six months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

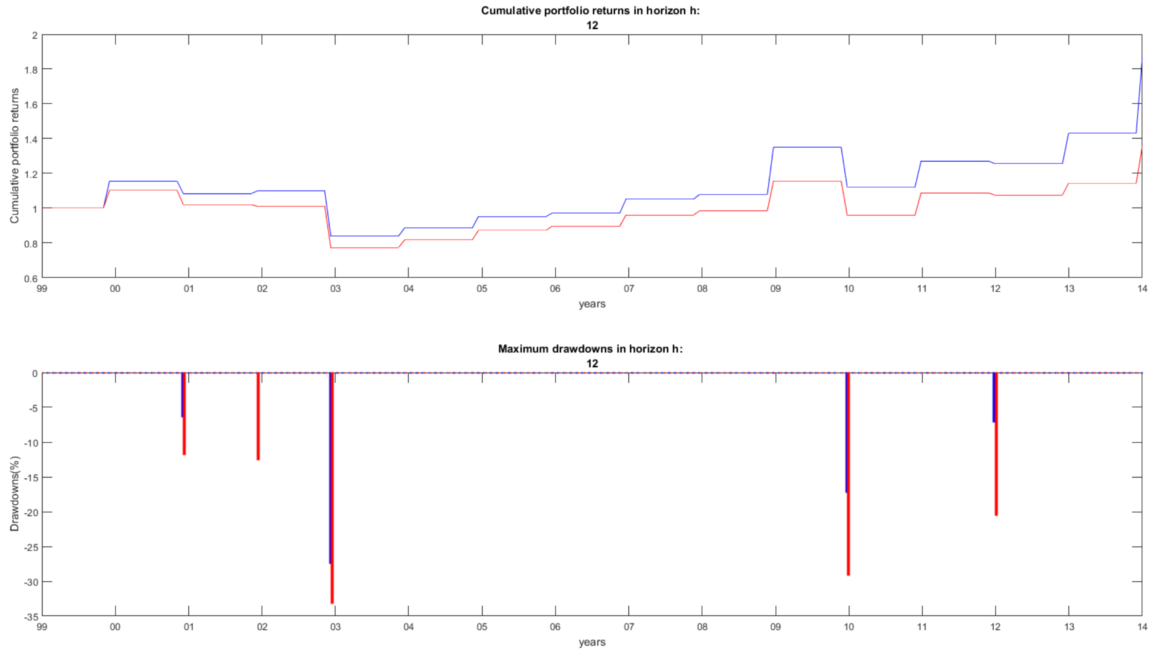


Figure 21: Cumulative portfolio returns with Ridge prior and h=12. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions twelve months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

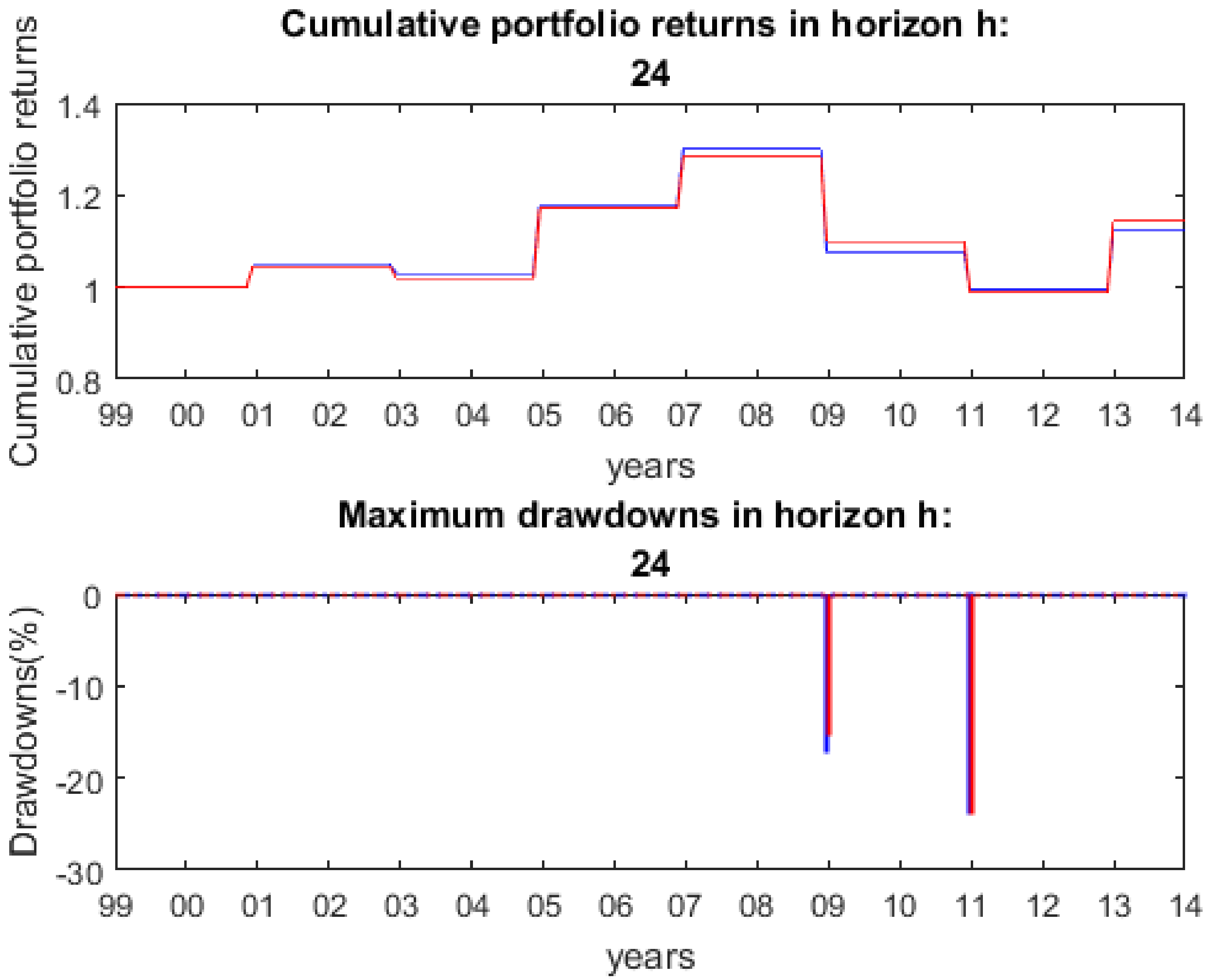


Figure 22: Cumulative portfolio returns with Ridge prior and h=24. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions twenty four months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

6.5.2 Cumulative portfolio returns and maximum drawdowns for BVARM and BVARMVS models and for $S\&W$ factors

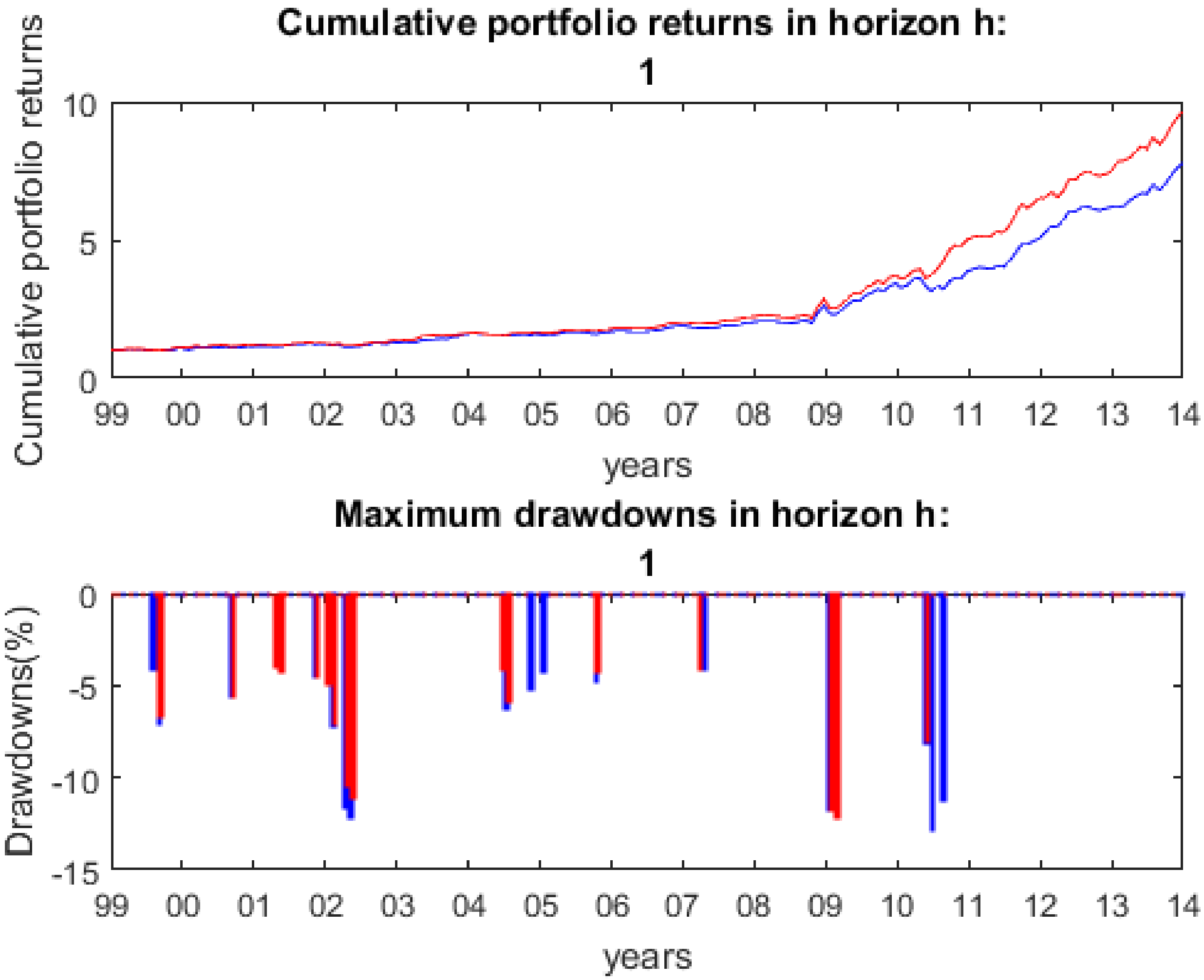


Figure 23: Cumulative portfolio returns with Minnesota prior and h=1. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions one month ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

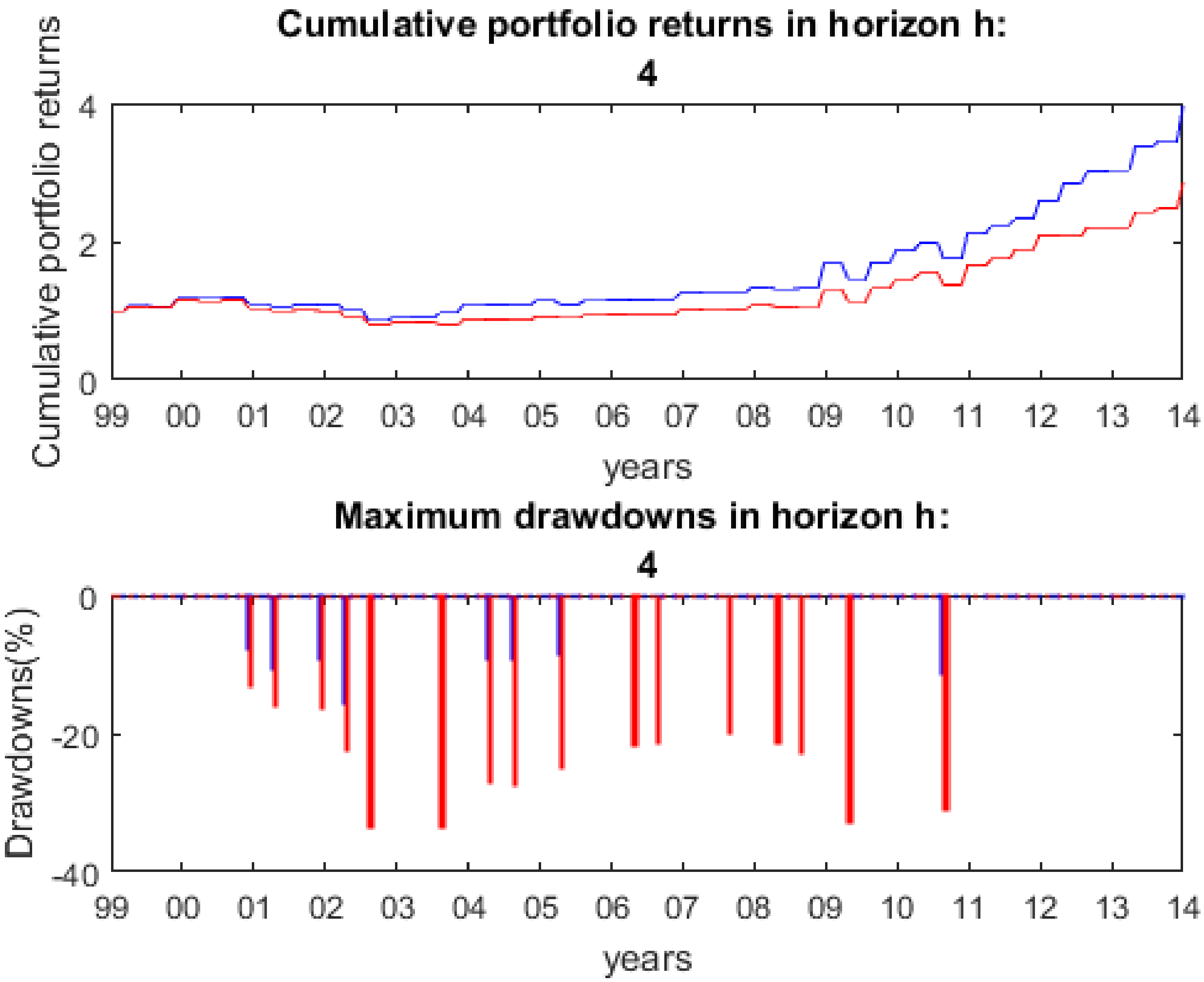


Figure 24: Cumulative portfolio returns with Minnesota prior and h=4. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions four months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

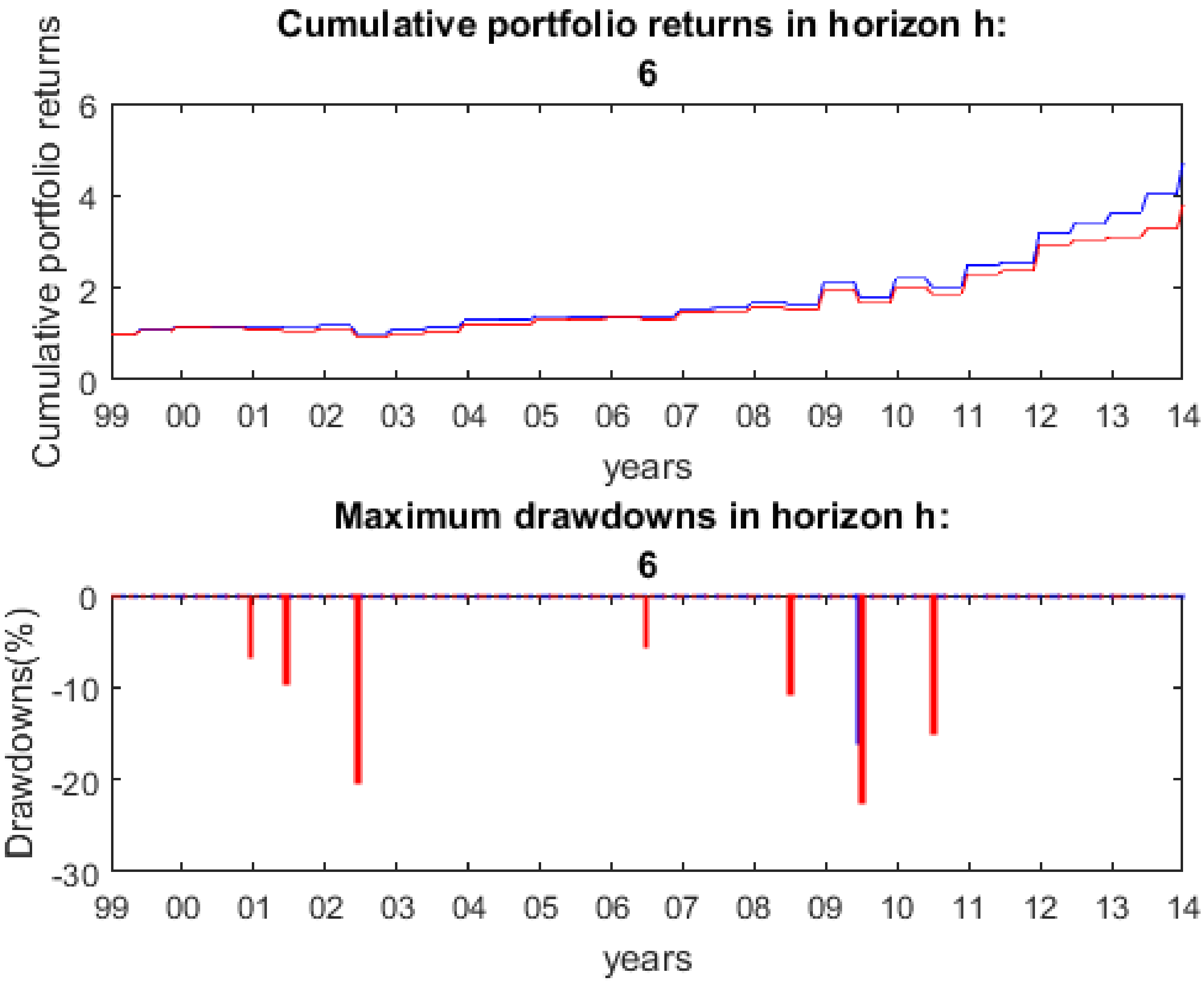


Figure 25: Cumulative portfolio returns with Minnesota prior and h=6. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions six months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

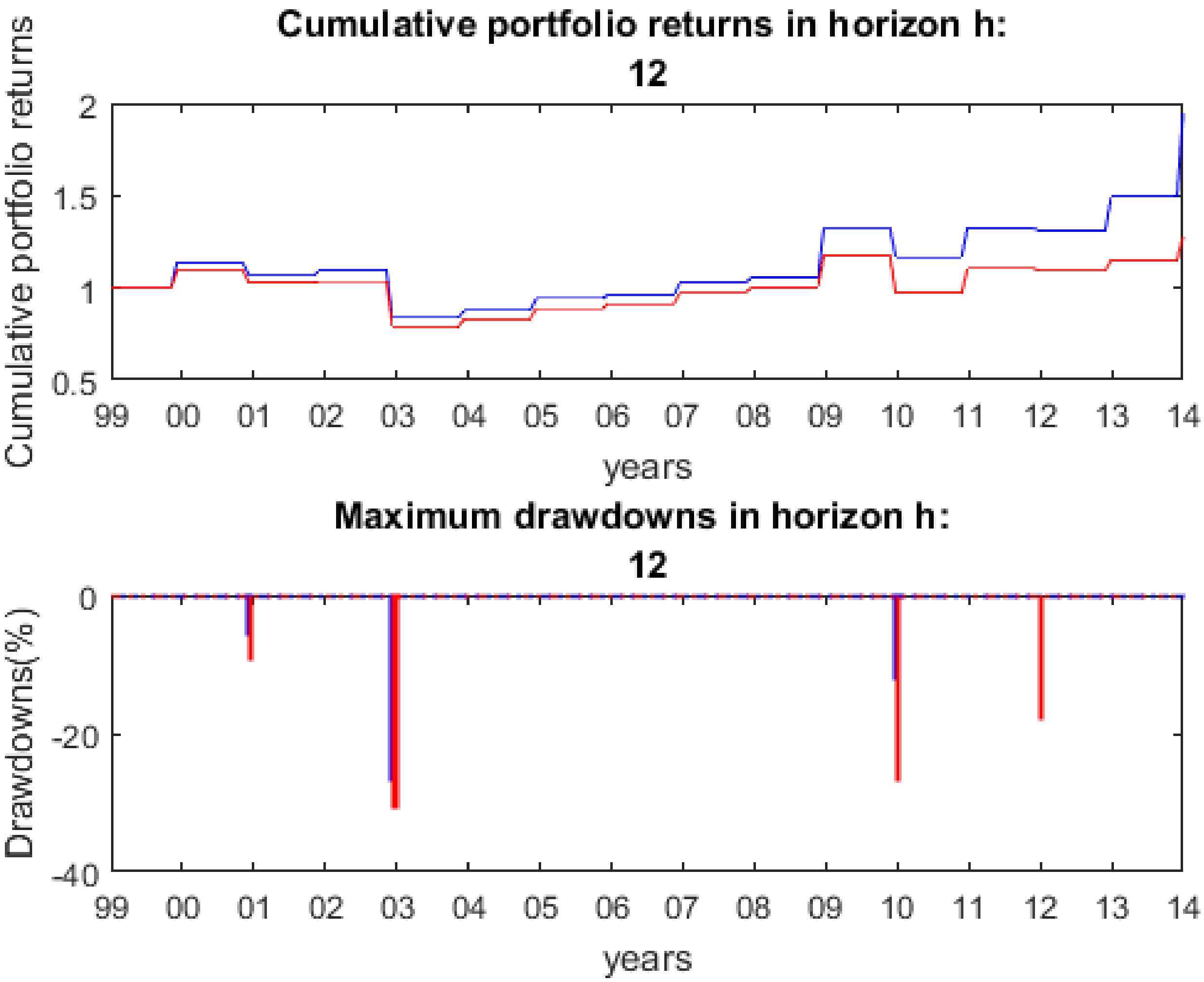


Figure 26: Cumulative portfolio returns with Minnesota prior and h=12. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also, see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions twelve months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

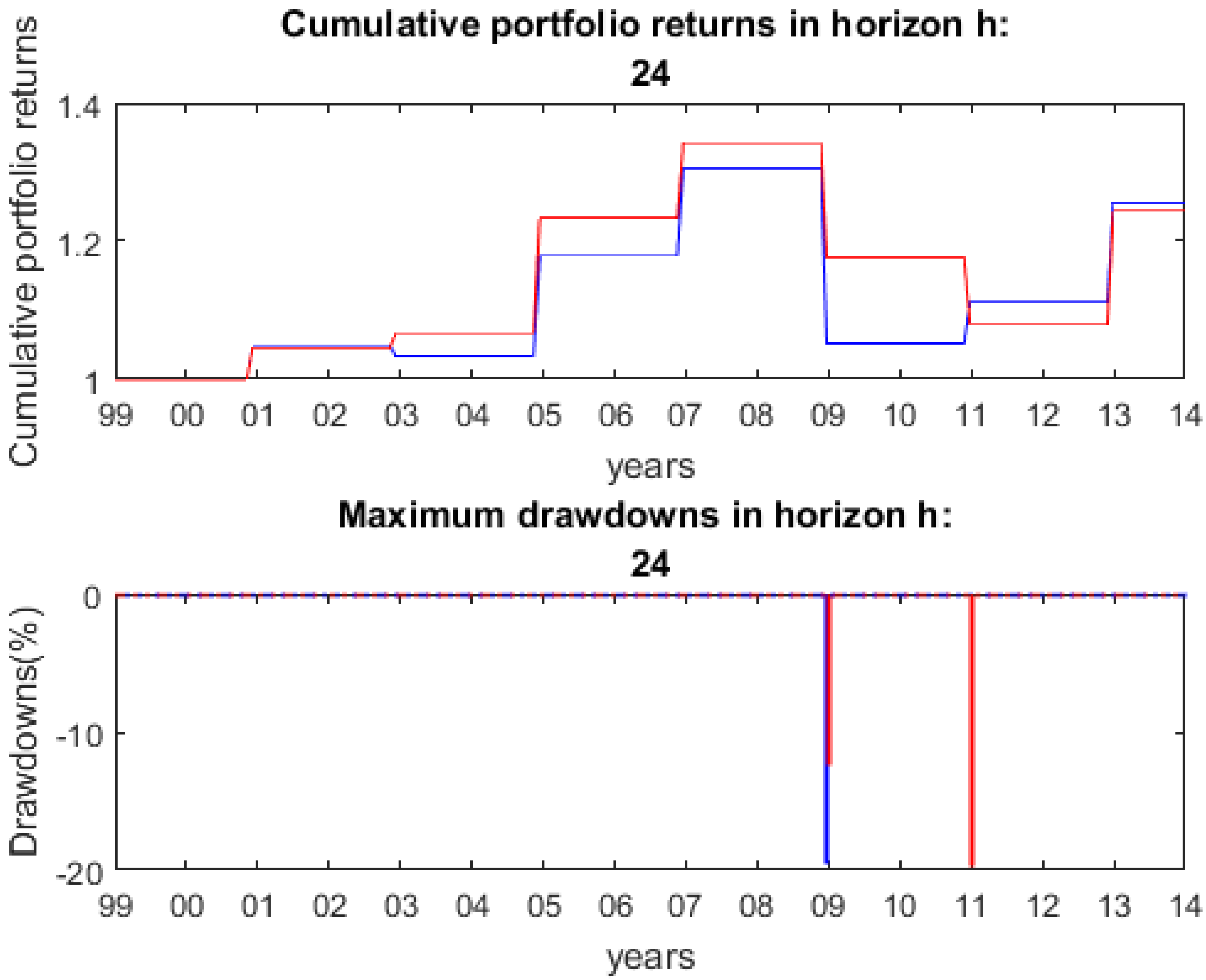


Figure 27: Cumulative portfolio returns with Minnesota prior and h=24. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also see Table (13) for definition of each model, for $S\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions twenty four months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

6.5.3 Cumulative portfolio returns and maximum drawdowns for BVARR and BVARRVS models and for $G\&W$ factors

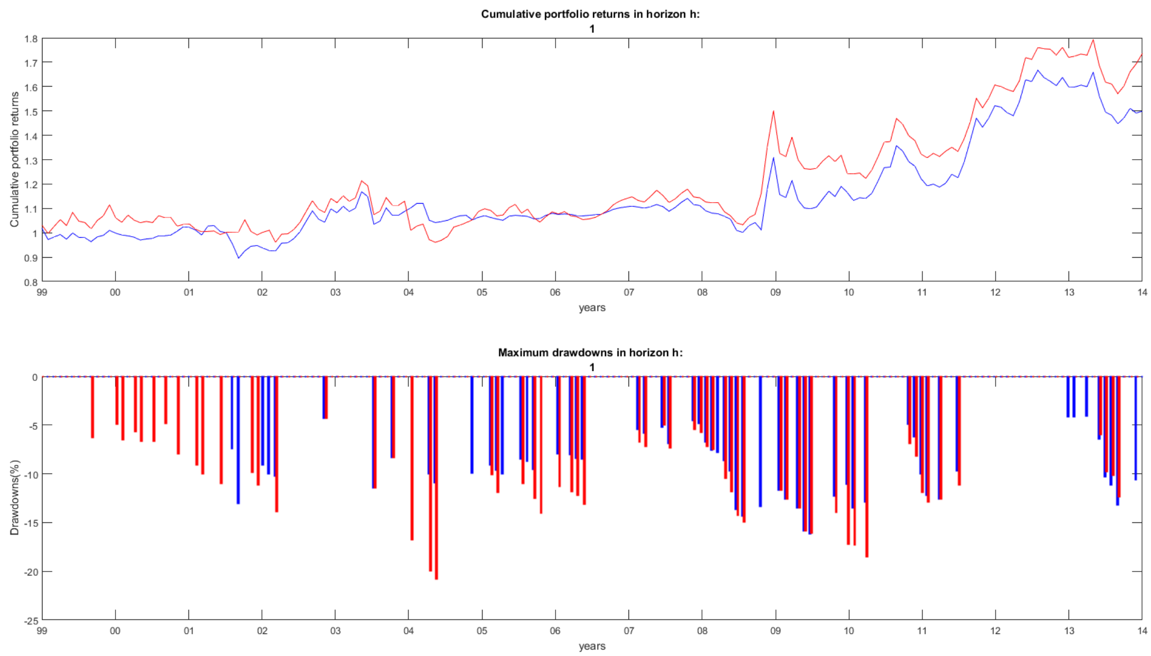


Figure 28: Cumulative portfolio returns with Ridge prior and h=1. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also, see Table (13) for definition of each model, for $G\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions one month ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

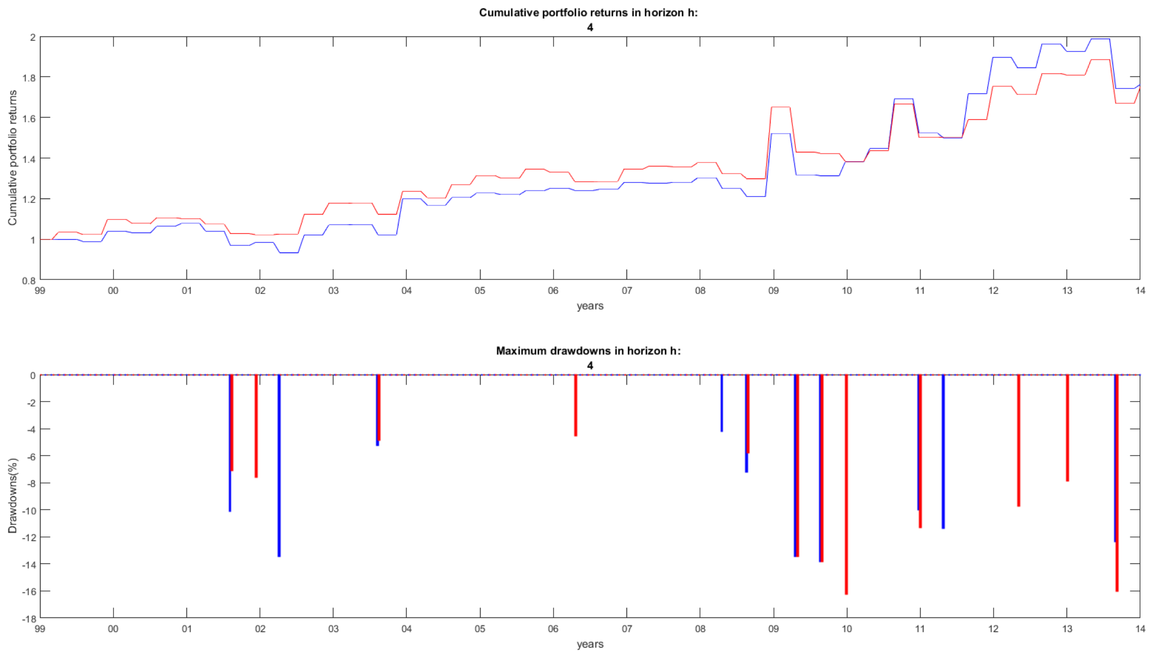


Figure 29: Cumulative portfolio returns with Ridge prior and h=4. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also, see Table (13) for definition of each model, for $G\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions four months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

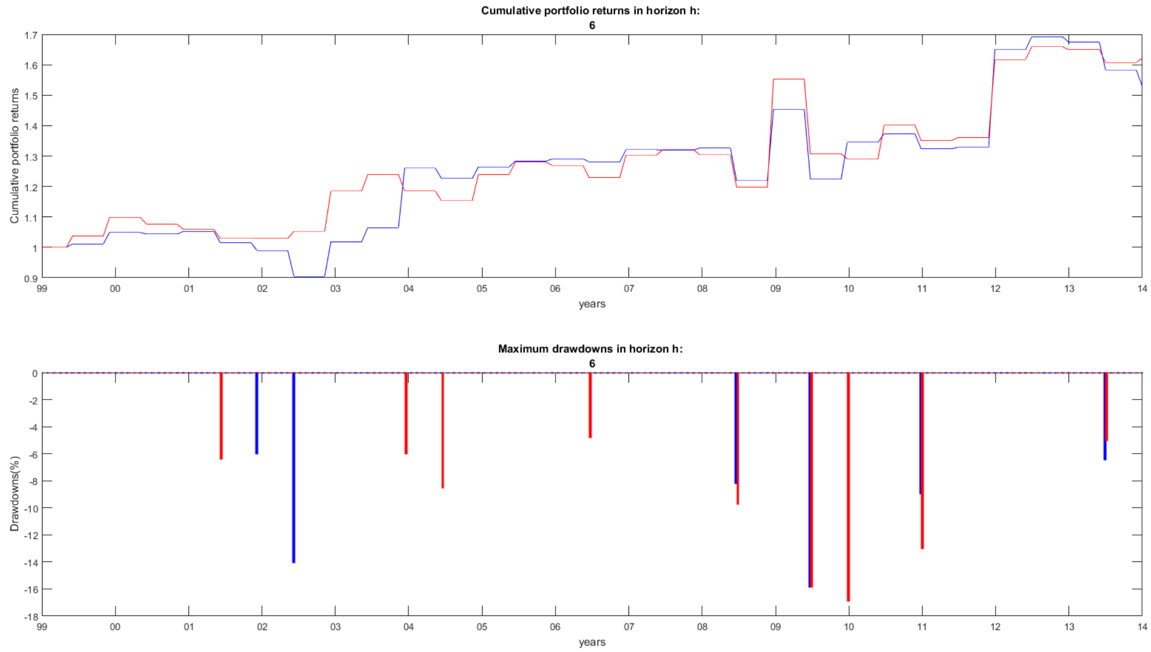


Figure 30: Cumulative portfolio returns with Ridge prior and h=6. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also, see Table (13) for definition of each model, for *G&W* factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions six months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

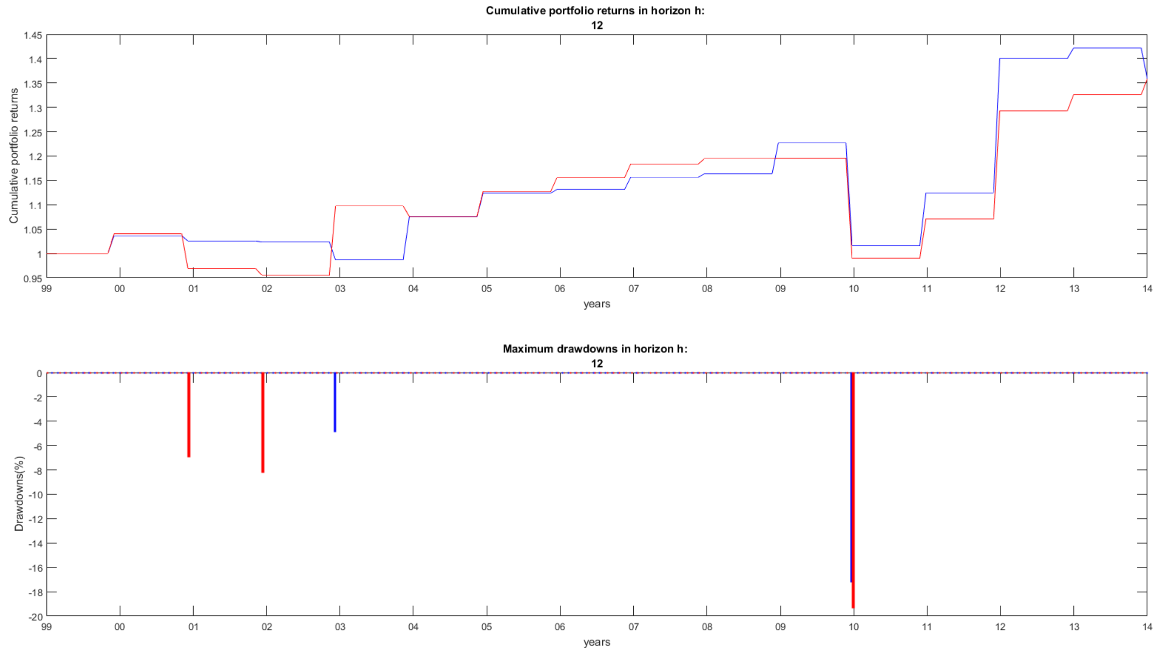


Figure 31: Cumulative portfolio returns with Ridge prior and h=12. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also, see Table (13) for definition of each model, for $G\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions twelve months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

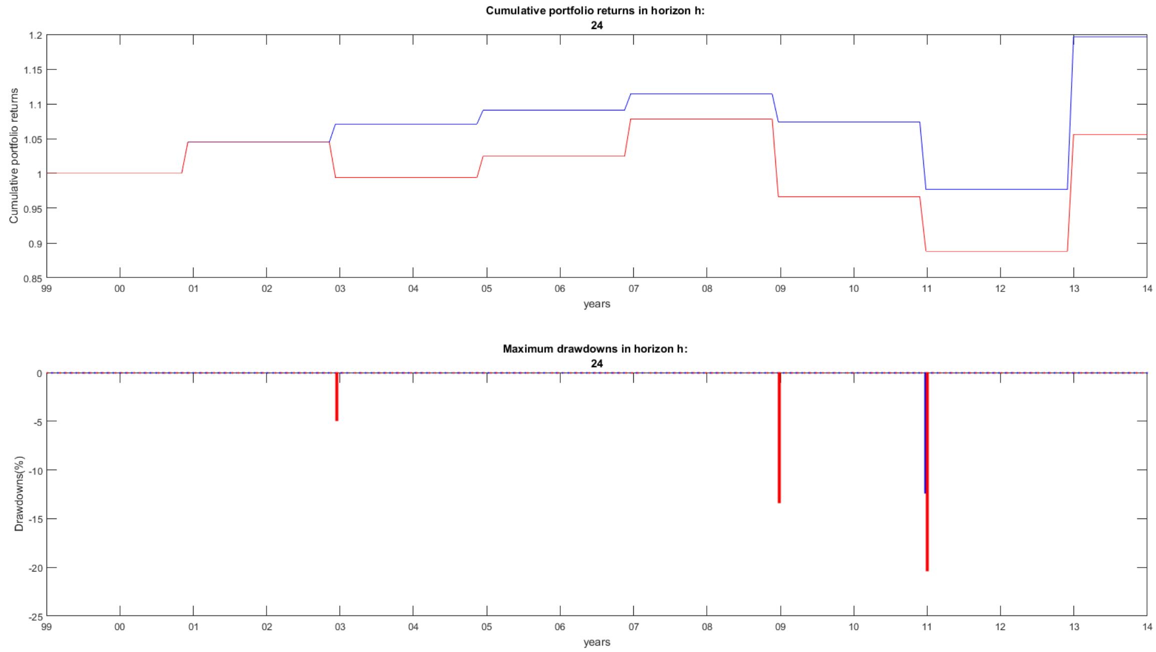


Figure 32: Cumulative portfolio returns with Ridge prior and h=24. This figure plots the cumulative portfolio returns of our two models BVARR and BVARRVS (also see Table (13) for definition of each model, for $G\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions twenty four months ahead. Each point in time for this portfolio construction belong to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

6.5.4 Cumulative portfolio returns and maximum drawdowns for BVARM and BVARMVS models and for $G\&W$ factors

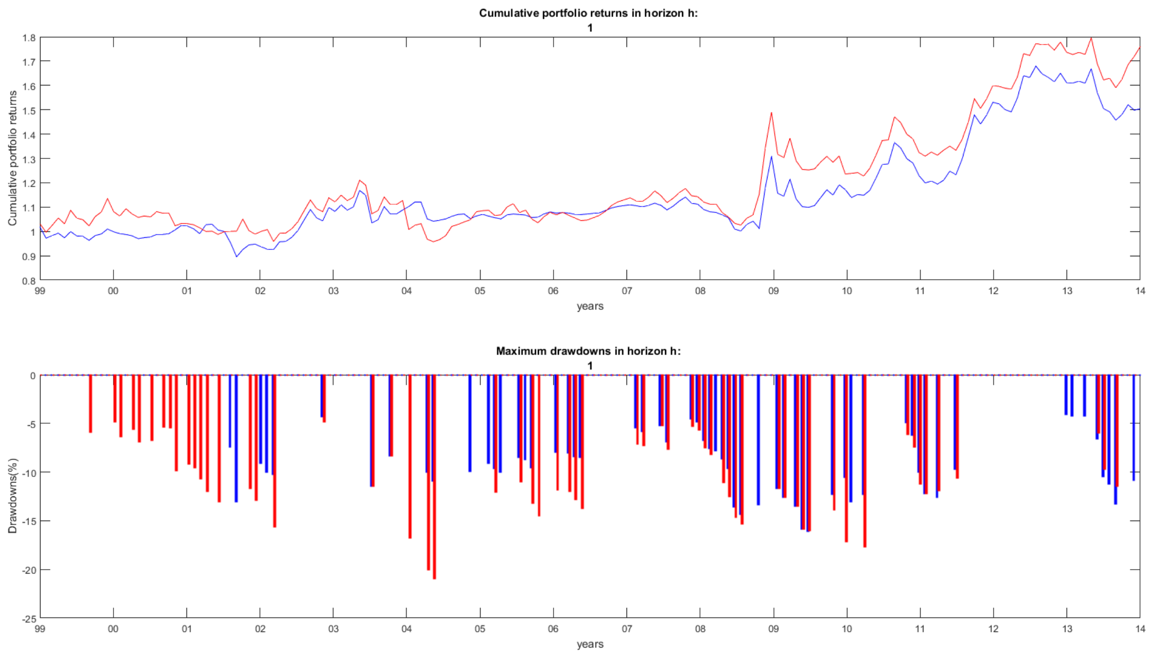


Figure 33: Cumulative portfolio returns with Minnesota prior and h=1. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also, see Table (13) for definition of each model, for $G\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions one month ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

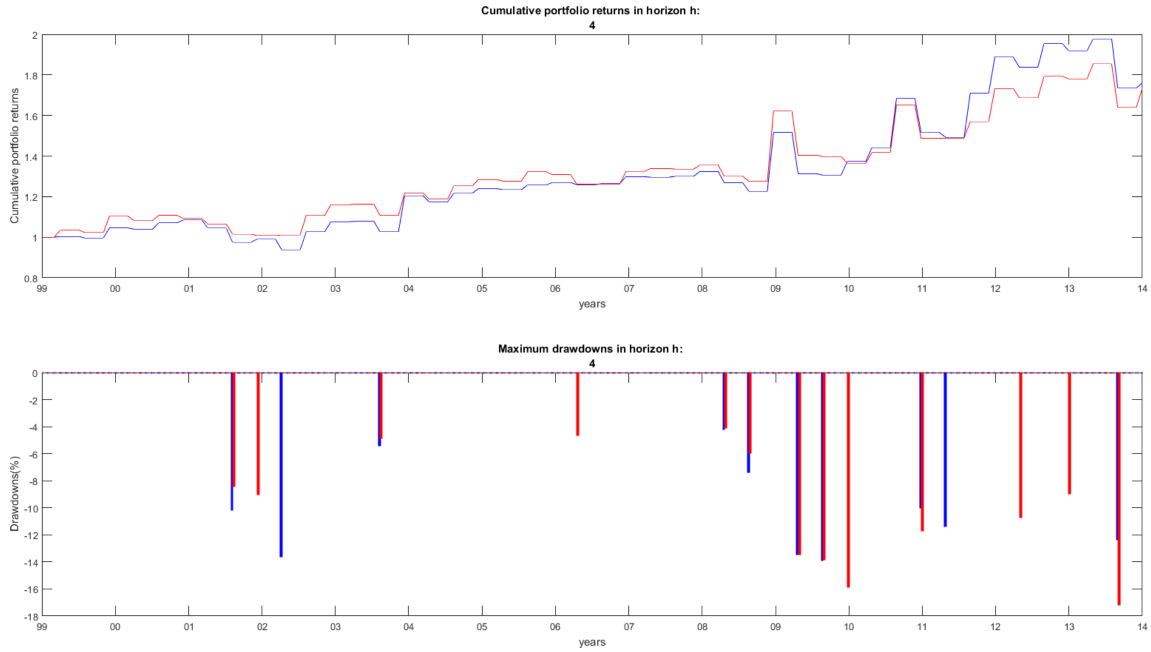


Figure 34: Cumulative portfolio returns with Minnesota prior and h=4. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also, see Table (13) for definition of each model, for $G\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions four months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

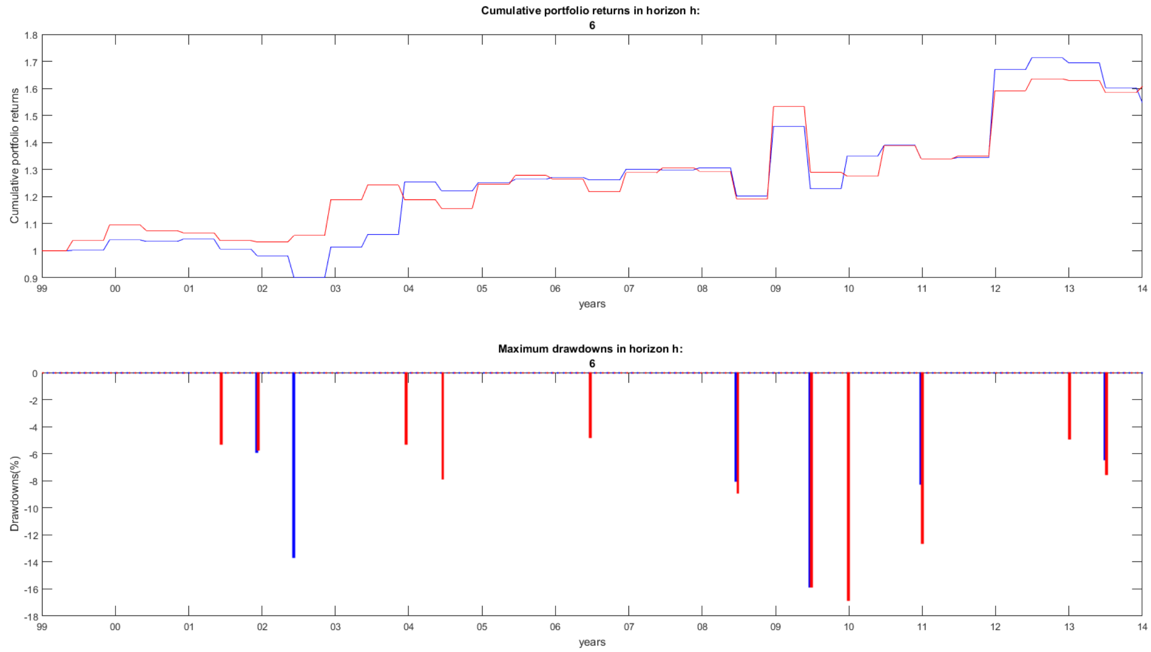


Figure 35: Cumulative portfolio returns with Minnesota prior and h=6. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also, see Table (13) for definition of each model, for $G\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions six months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

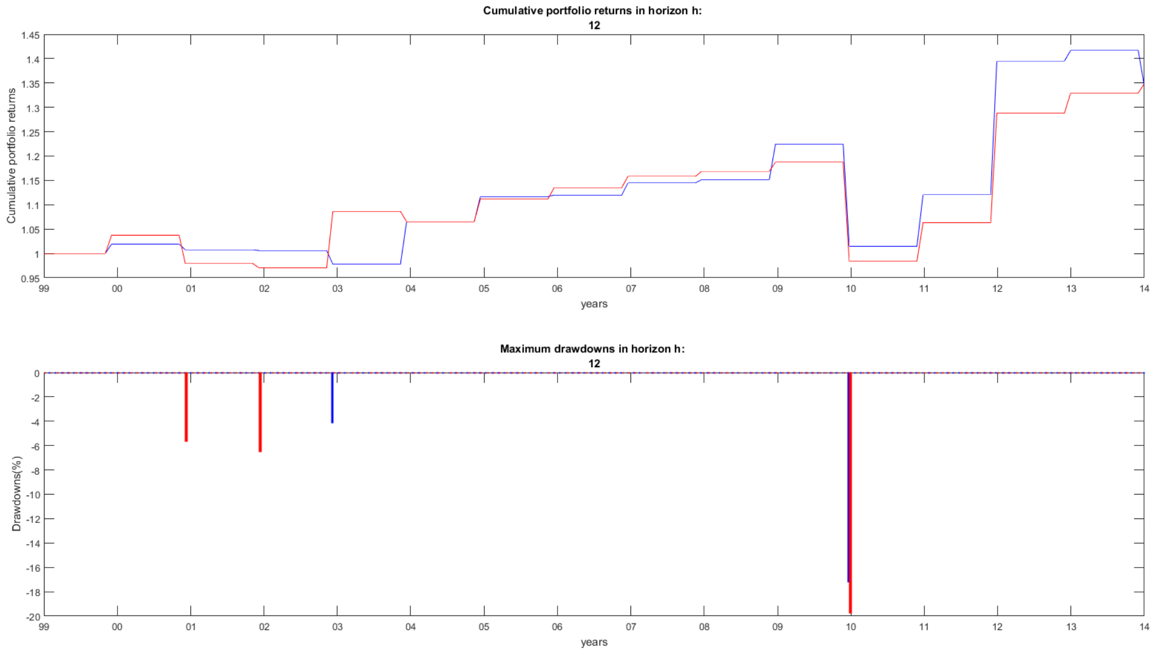


Figure 36: Cumulative portfolio returns with Minnesota prior and h=12. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also, see Table (13) for definition of each model, for *G&W* factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions twelve months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.

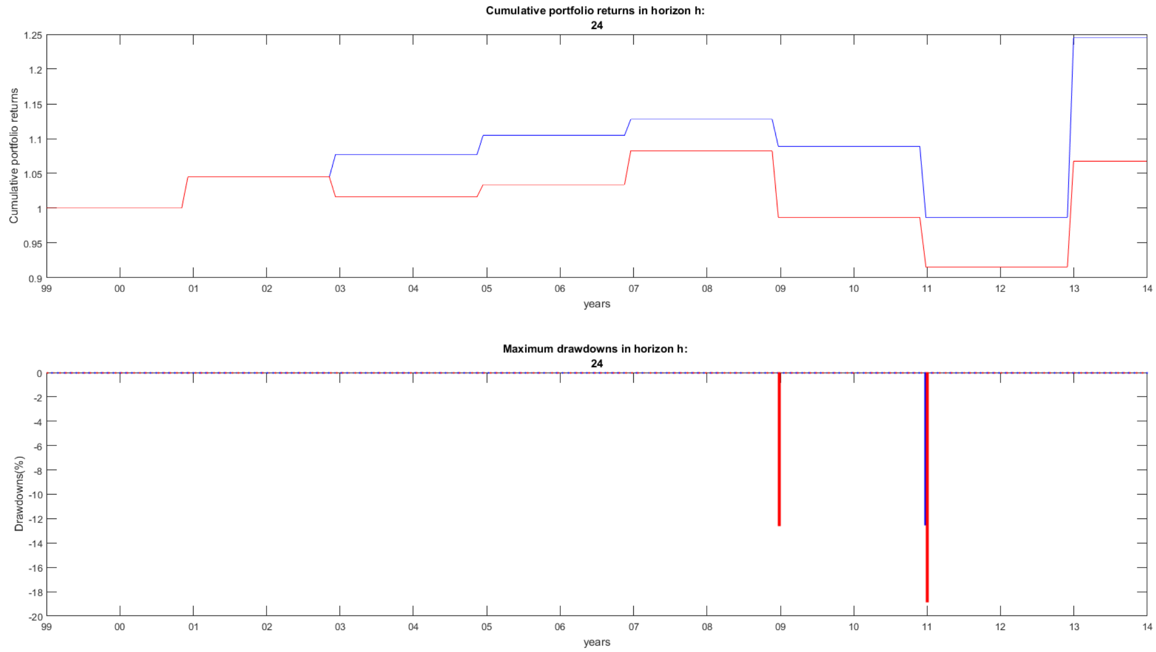


Figure 37: Cumulative portfolio returns with Minnesota prior and h=24. This figure plots the cumulative portfolio returns of our two models BVARM and BVARMVS (also, see Table (13) for definition of each model, for $G\&W$ factors. The investment strategy is a buy-and-hold strategy for an investor that optimizes her power utility functions twenty four months ahead. Each point in time for this portfolio construction belongs to the Test set. In order to predict each point in time of the Test set, we use 20,000 draws from the posterior distribution.