Model uncertainty for long-term investors*

Bart Diris[a,b†]

July, 2011

Abstract

We develop a method to incorporate model uncertainty with respect to restricted VAR(1) models using Bayesian Model Averaging (BMA) and apply our method to analyze the long-run predictability (forecast horizons up to 30 years) of asset returns. We find that only the dividend yield and credit spread are important predictors of stock returns in the short-run, but that almost all considered predictors are important for long-run predictability. Despite clear evidence of mean-reversion in stock returns, we show that stocks are in general at least as risky in the long-run as in the short-run if model uncertainty is incorporated and that stocks are even riskier in the long-run in case of an economic crisis such as the recent sub-prime mortgage crisis. Single models however underestimate the long-run riskiness of stock returns considerably. Finally, the strategic asset allocations for long-term investors using BMA are substantially different from investors that use the highest posterior probability model. Our analysis relates this finding to a lower mean, higher variance, more negative skewness and a higher kurtosis of the predictive distributions of excess stock returns when incorporating model uncertainty. Differences are especially large when the economy deviates substantially from its steady state value.

Affiliations: Erasmus University Rotterdam[a], Netspar[b]
Keywords: Bayesian Model Averaging, Predictability of asset returns, Strategic asset allocation

*I am grateful to Jules van Binsbergen, John Campbell, Gary Chamberlain, Magnus Dahlquist, Guido Imbens, Franz Palm, Peter Schotman, Ralph Stevens, Jean-Pierre Urbain and seminar participants at the 4th annual SoFiE Conference in Chicago, 16th International Conference on Computing in Economics and Finance in London, Nake Research Day in Utrecht, Netspar International Pension Workshop in Zürich, Erasmus University, Maastricht University and Harvard University. Part of this paper was written while the author was a visiting fellow at Harvard University.

†Corresponding author: Bart Diris, Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands (email: diris@ese.eur.nl)

1 Introduction

Merton (1969, 1971) showed that under changing investment opportunities, the optimal portfolios of long-term investors differ from the ones of short-term investors. Long-term investors hold hedge portfolios that anticipate future changes in the investment opportunities. Empirically, the main driving force in these hedge portfolios is the mean reversion of stock returns, which implies that equity is less risky for long-term investors than other types of assets. The main implication is that long-term investors in general hold more equity than short-term investors. The standard methodology in the strategic asset allocation literature is to select one model and to use its predictions to calculate the optimal portfolios of long-term investors. Prominent examples are Barberis (2000) - who considers a small asset menu, but includes parameter uncertainty - and Campbell, Chan, and Viceira (2003) - who consider a larger asset menu, but ignore parameter uncertainty.[1]

Long-term investors face substantial uncertainty about how to model the predictive distribution of future asset returns. The reason is that they face uncertainty about whether stock returns are predictable in the long-run and if so, which predictor variables they should include. Firstly, there is no consensus in the literature on whether stock returns are predictable in the short-run.[2] On one hand, Goyal and Welch (2008) show that none of the predictor variables they consider consistently outperforms the historical average of stock returns, but on the other hand Campbell and Thompson (2008) show that some of the predictor variables outperform the sample mean if one adds extra information to the regressions. Secondly, there is hardly any research on the long-run predictability of asset returns despite the fact that its presence or absence is very important for long-horizon investors. Taken together this means that long-term investors face substantial uncertainty about how to model the predictive distribution of future asset returns.

We take the perspective of such a long-term investor who explicitly acknowledges that she is uncertain about the econometric model she should use to model the distribution of future asset returns and about the values of the parameters in the econometric model. She considers models that differ in the predictor variables that are included. As a first step we develop an estimation framework that is able to include model uncertainty over long-term predictions using Bayesian Model Averaging (BMA). Secondly, we analyze empirically whether model uncertainty is present and relevant for investors with long investment horizons (up to 30 years). To the best of our knowledge we are the first ones to consider (the implications of) model uncertainty for long-horizon investors. We find that model uncertainty is not important at short horizons, but that it is very important at long horizons.

The models we consider all specify how to predict a set of twelve dependent variables. We include three asset returns - the real T-bill rate, excess stock returns and excess bond returns

[1]Campbell and Viceira (2002) and Brandt (2010) provide an extensive survey of the strategic asset allocation literature.

[2]Refer for example to the special issue (number 21) on stock return predictability of The Review of Financial Studies in 2008.

- and nine predictor variables - the default premium, the dividend-to-price ratio, the book-to-market ratio, the price-earnings ratio, the nominal yield on the 90-day T-bill, the yieldspread, the credit spread, a measure for newly issued stocks and a proxy for the stock return variance as dependent variables. Every model specifies the set of right-hand-side variables to include in the twelve different equations. This set is a subset of the first lag of the dependent variables and can differ across the equations. The model that nests all other models is a VAR(1) model. The other models are restricted VAR(1) models, where some (or all) of the lagged dependent variables are excluded in some (or all) of the equations. In order to obtain long-horizon predictions, we iterate the model forward.

The distinguishing feature of our specification is the fact that it allows us to incorporate model uncertainty in both the prediction of asset returns (which is usually done in the literature) and the prediction of the predictor variables (which is usually ignored in the literature). Since it is essential at long horizons to accurately predict the predictors of asset returns, the latter effect is at least as important as the former effect and therefore needs to be taken into account when considering the impact of model uncertainty on the long-run predictability of asset returns. Ignoring this effect will significantly underestimate the importance of model uncertainty at long horizons. To give an example, a long-term investor cannot use the information that the dividend-to-price ratio predicts next period's stock returns if she does not know how to predict the dividend-to-price ratio itself.

An essential difference between our specification and the set-ups in Avramov (2002) and Cremers (2002) is that their set-ups are not suited for analyzing the impact of model uncertainty on long-horizon predictability. Avramov (2002) briefly considers long-run predictions, but the posterior model probabilities in his analysis are only based on the equation for stock returns and are not based on the prediction equations for the predictor variables. Therefore, his set-up considerably underestimates the true impact of model uncertainty at long horizons.[3]

The estimation results show that the credit spread and the dividend-to-price ratio are the most important predictors of stock returns at short horizons. However, at long horizons all variables are important predictors of long-horizon stock returns either by predicting stock returns directly or by predicting the predictors of stock returns. Model uncertainty is clearly present, since none of the models receives a high posterior model probability.

We find that model uncertainty is very relevant for long-term investors, but relatively unimportant at short horizons. Firstly, our results show that the incorporation of model uncertainty increases the risk of stock returns - measured by the variance of the predictive distribution - considerably at long horizons. The reason is that the BMA specification averages over models that predict very different future trajectories of stock returns. Therefore, the variance of the (mean) forecasts of all these different models becomes an important component of the total predictive variance of stock returns. The impact is especially large when the predictor variables

[3]A related paper is Wright (2008). He applies a similar methodology as Avramov (2002) to forecast exchange rates at longer horizons using a BMA specification. Both use an auxiliary model for the prediction of predictor variables. However, in their setting the fit of the auxiliary model does not impact the model probabilities although it is essential for long-horizon predictability.

deviate substantially from their historical average, since the differences in predictions across models is especially large in that case. The predictive variance of stock returns changes over time, because predictors change over time, even though the considered models are homoscedastic. However, the incorporation of model uncertainty only has a minor effect at short horizons, since the different models predict very similar future stock returns for short horizons.

These results are partially related to recent findings in Pastor and Stambaugh (2010). They find that the annualized predictive variance of stock returns is much higher at long horizons than at short horizons. Our results only partially confirm Pastor and Stambaugh (2010). We find that the (annualized) predictive variance at a 30-year horizon was much higher than at a 1-quarter horizon during financial crises such as the great depression in 1929 and the subprime mortgage crisis in 2008. However, during the 1960s and 1970s, stocks turned out to be much safer at long horizons. Our results contradict some of the results in Avramov (2002). He finds that model uncertainty has the biggest impact at short horizon. On the contrary we show that model uncertainty is mainly important at long horizons, since both the prediction of asset returns and the prediction of the predictors of asset returns is important at longer horizons.

Secondly, we find that the incorporation of model uncertainty leads to very different strategic asset allocations at long horizons using buy-and-hold strategies. The incorporation of model uncertainty decreases stock allocations by up to 35% for investment horizons up to 20 years relative to a setting where we select the model that obtains the highest posterior probability. Empirically, we find that the stock allocation is so much lower, because the inclusion of model uncertainty lowers the mean, increases the variance, leads to (more) negative skewness and increases the kurtosis of the predictive distribution of stock returns. Note that all four moments of the predictive distribution depend on the values of predictor variables and therefore change substantially over time. We also find that the incorporation of model uncertainty has an important impact on the utility that long-term investors expect to receive from the strategic asset allocations. At the longest horizons, the certainty equivalent can be lower than 1 even when an investor follows the optimal strategy. This implies (since all returns are in real terms) that an investor is willing to pay a lot for an inflation-indexed bond to avoid following (optimal) buy-and-hold strategies. Furthermore, the results again show that model uncertainty only has a minor impact on short horizons. Its incorporation hardly changes the predictive moments, asset allocations and expected utility at short horizons.

These results partially contradict results in Barberis (2000) and Campbell, Chan, and Viceira (2003). They find that long-term investors should invest more in equity than short-term investors even when parameter uncertainty is included. However, we find that this is not the case when model uncertainty is also incorporated. Our results show that investors with a horizon of 20 years should invest as much in the stock market as short-term investors.

Thirdly, we show that our results are robust to changes in our specification. Firstly, we find that a different prior distribution leads to very similar posterior results. Secondly, we show that the estimation results are very stable over time by estimating our specification on smaller subsamples.

A final important contribution of our paper is that we succeed in efficiently estimating the BMA specification. Although we cannot calculate the posterior moments analytically, our refined Markov Chain Monte Carlo (MCMC) technique converges quickly to the stationary distribution of the Markov chain. The Monte Carlo technique is based on results in Godsill (2001) and only requires us to know the marginal likelihood conditional on the covariance matrix of the error term. Standard techniques to estimate BMA specifications instead crucially depend on the marginal likelihood.[4] The latter is however not analytically available for the VAR(1) models with arbitrary restrictions that we consider.

The remainder of this paper is organized as follows. Section 2 introduces the data we use. Next, Section 3 describes the methodology. It provides details on the model, the prior assumptions, the posterior distributions, the Markov Chain Monte Carlo (MCMC) techniques we use to obtain estimation results, and some alternative specifications. Subsequently, section 4 reports the estimation results of the weighted Bayesian Model Averaging (BMA) specification. Sections 5 - 6 analyze the impact of the incorporation of model uncertainty on the term structure of risk of stock returns and on strategic asset allocations. Next, section 7 provides some robustness checks. Finally, section 8 concludes. The appendix contains additional technical details on the posterior distributions and the simulation techniques that are used in this paper.

2 Data

We use a quarterly data-set for the US stock and bond market. It consists of three asset returns and nine predictor variables and is based on Goyal and Welch (2008). The data set starts in the fourth quarter of 1926 and ends in the fourth quarter of 2008. We use all predictor variables that are available in the quarterly Goyal and Welch (2008) data-set for this sample period.

[Table 1 about here.]

The first asset return is the ex post real T-bill rate (R_{tbill}) defined as the difference between the log return (or lagged yield) on the 3-month T-bill and log inflation. The second asset return is the excess log stock return (X_s) which is the difference between the log return (including dividends) on the $S\&P$ 500 and the log return on the (nominal) 3-month T-bill. The final asset return is the excess log return on a long-term government bond with a maturity of 20 years (X_b) and is defined similarly.

Next, we consider nine predictor variables. Several papers show that these variables predict excess stock returns and/or excess bond returns. Please refer to Goyal and Welch (2008) for references and more details on data construction. The first predictor variable is the default risk premium (Def_{pr}) formed as the return difference between a long-term corporate bond and a long-term government bond. The second predictor is the log dividend-to-price ratio (DP) defined as the log difference between the dividends over the past four quarters and the current log index level. The next predictor variable is the log book-to-market ratio (BM) given as the

[4] An example is the popular MC^3 technique used in e.g. Fernandez, Ley, and Steel (2001).

log of the ratio of the book-to-market value for the Dow Jones Industrial Average. The fourth predictor is the log price-earnings ratio (PE) formed as the difference between the current log index level and the log of average earnings over the past 10 years. The next predictor is the smoothed log nominal yield (Y_{nom}). It is defined as the log nominal yield on the 90-day T-bill minus the average log nominal yield over the past four quarters. Sixthly, we include the yield spread (Y_{spr}) which is defined as the difference between the log yield on a long-term government bond and the log yield on the 90-day T-bill. The seventh predictor is the credit spread (Cr_{spr}) formed as the difference in log yields of Moody's BAA and AAA rated bonds. The eighth predictor is the ratio of 12-month moving sums of net issues by NYSE stocks and the total end-of-year market capitalization of these stocks ($ntis$). The final predictor is the log of the stock return variance (Var) and is proxied by summing the squared daily returns on the $S\&P$ 500 over every quarter.

Table 1 provides summary statistics of our data. The equity risk premium of 1.32 % per quarter is in line with most recent papers that use historical data. The results in the table show that the last eight predictor variables are very persistent.[5]

3 Methodology

Define the $n \times 1$ vector y_t as follows ($n = 12$ in our set-up)

$$y_t = \begin{pmatrix} r_{tbill,t} \\ x_t \\ s_t \end{pmatrix}, \tag{1}$$

where $r_{tbill,t}$ is the real return on the T-bill, x_t is a vector of excess returns on stocks and bonds over the T-bill and s_t is a vector of state variables. Furthermore, define Y as the $T \times n$ matrix containing observations on y_t and define Y_{-1} as the $T \times n$ matrix containing observations on y_{t-1}. Finally, let Y_i be the i^{th} column of Y.

Our aim is to model the dynamics of all twelve variables in y_t. The common methodology in the strategic asset allocation literature is to select one model using a model selection criteria and to base inference on this model under the assumption that the selected model is the correct model, refer for example to Campbell, Chan, and Viceira (2003) and Barberis (2000). However, this method ignores the uncertainty in the model selection step and therefore substantially underestimates the uncertainty an investor truly faces. For example, suppose that the second-best model is almost as likely as the best model, but leads to very different implications for long-term investors. If an investor wants to obtain an accurate picture of the distribution of

[5] Since the frequentist sampling theory of (for example) the OLS estimator depends strongly on the presence of a unit root, a frequentist econometrician might wonder whether these variables actually contain unit roots. We cannot reject a unit root in DP, BM, PE and the Cr_{spr} using the Augmented Dickey Fuller test at the 5% significance level. However, it is important to note that this will not affect inference in our Bayesian setting, since posterior distributions do not condition on unit roots. Besides, if we interpret the results of the frequentist unit root tests in a Bayesian way as suggested by Sims and Uhlig (1991), we only find very limited evidence for the presence of unit roots with a largest "p-value" of only 2.80% for BM.

future asset returns and the uncertainty she faces, she should also include the implications of this second-best model. The Bayesian methodology allows us to incorporate model uncertainty in the decisions a long-term investor faces.

We use Bayesian Model Averaging (BMA) to average model predictions across all considered models. It assigns a posterior probability to all individual models and uses these probabilities as weights on the forecasts of the individual models to come up with the composite forecast. In this way, models that receive positive posterior probability are taken into account in the composite forecasts, but only the plausible models get a large weight. Several papers such as Avramov (2002) and Cremers (2002) have also shown that the use of the BMA technique leads to better out-of-sample forecasts.[6]

If more and more information gets available (if the sample size T goes to infinity) we would hope that the posterior probability of the "best" model goes to 1. Gelfand and Dey (1994) show that asymptotically the posterior probability of the true model goes to 1 if it is included in the model set. Fernandez-Villaverde and Rubio-Ramirez (2004) extend this result to a setting where the true model is not included in the model set. They find that the posterior probability of the model that minimizes the Kullback-Leibler distance to the true model goes to 1 asymptotically.

The latter result can be seen as a justification for using posterior model probabilities when the true model is not included in the model set. In fact, we believe - as do Hoeting, Madigan, Raftery, and Volinsky (1999) - that using Bayesian Model Averaging might even be more important in such a setting, since we expect that none of the considered models has a high posterior probability in a moderately sized sample. If instead the true model would have been included, we would expect its posterior probability already to be large in a moderately sized sample. Model uncertainty would be less important with such a dominant model.

The next subsections introduce the model (and likelihood function), the prior, the posterior and the MCMC techniques we use. The last subsection compares our specification to alternative specifications.

3.1 Model

The models we consider consist of twelve equations. Every model specifies the set of right-hand-side variables that is included to predict the individual elements in y_t. This set always includes a constant and a subset of Y_{-1}. For simplicity, we do not consider lags beyond one.[7] We allow the set of included right-hand-side variables to differ across equations.

First, we consider the model that includes all variables in y_{t-1} as right-hand-side variables

[6]We choose the Bayesian perspective, because it is conceptually straightforward to include Model Averaging by just treating models as random themselves and applying Bayes rule in the standard way. It would also have been possible to use model averaging from the frequentist perspective. A frequentist could for example use (functions of) information criteria to give weight to the different models, refer for example to Hjort and Claeskens (2003). The frequentist methods do not seem to be able to deal with such a large model class as we consider.

[7]Although the models only predict next period's asset returns and predictor variables by using current values of the asset returns and predictor variables, we use the models to make long-horizon forecasts of stock returns. We do this by iterating our model forward as in e.g. Barberis (2000).

in every equation

$$Y_i = X\beta_i + \epsilon_i, \; i = 1, ..., n, \tag{2}$$

where X is a $T \times (n+1)$ matrix $[\iota, Y_{-1}]$ and β_i and ϵ_i are respectively the $(n+1) \times 1$ vector of regression coefficients and the $T \times 1$ vector of error terms for equation i. This model nests all the other models we consider. Equivalently,

$$Y = XB' + E, \tag{3}$$

where B is a $n \times (n+1)$ matrix of regression coefficients and E is a $T \times n$ matrix of error terms. This all-encompassing model is a VAR(1) model. The other models are obtained by removing some (or all) of the right-hand-side variables from some (or all) of the n equations. We assume throughout that the $n \times 1$ vector ϵ_t is i.i.d. normally distributed

$$\epsilon_t \sim N(0, \Sigma). \tag{4}$$

Model j (M_j) is defined by specifying the set of included right-hand-side variables for every equation i. We denote the set of right-hand-side variables for equation i in model j as $Z_i^{(j)}$ ($Z_i^{(j)}$ always contains a constant). The regression model j is then given as

$$Y_i = Z_i^{(j)}\beta_i^{(j)} + \epsilon_i^{(j)}, \; i = 1, ..., n, \tag{5}$$

where $\beta_i^{(j)}$ is a $k_i^{(j)} \times 1$ vector of slope coefficients for equation i in model j.

The regression models we consider are restricted VAR(1) models.[8] Since $\Sigma^{(j)}$ - the covariance matrix of the error term for model j - is not diagonal, the error terms $\epsilon_i^{(j)}$ are correlated across equations i. This implies that the n regression equations form a Seemingly Unrelated Regression (SUR) model. Therefore, we need to estimate the n regression equations simultaneously to obtain efficient estimates.[9] In total we consider 2^{n^2} models.

In order to calculate the posterior results, we need to obtain an expression for the likelihood functions of the different models. Therefore, we first introduce some alternative notation. Instead of equation (5), we can express model j as follows

$$y = Z^{(j)}\beta^{(j)} + \epsilon^{(j)}, \tag{6}$$

where y is the $Tn \times 1$ vector vec(Y), $\beta^{(j)}$ is the $k^{(j)} \times 1$ vector $[\beta_1^{(j)\prime}, ...\beta_n^{(j)\prime}]'$ with $k^{(j)} = \sum_{i=1}^{n} k_i^{(j)}$,

[8] It is common in the strategic asset allocation literature to use small VAR(1) models to model the dynamics of asset returns and predictor variables, refer for example to Campbell and Viceira (2002) and Campbell, Chan, and Viceira (2003).

[9] In some restrictive cases, it would also be efficient to estimate the different equations separately, for example when the same set of right-hand-side variables is included in all equations.

$\epsilon^{(j)}$ is a $Tn \times 1$ vector of vertically stacked $\epsilon_i^{(j)}$ and $Z^{(j)}$ is a $Tn \times k^{(j)}$ matrix

$$Z^{(j)} = \begin{pmatrix} Z_1^{(j)} & 0 & \dots & 0 \\ 0 & Z_2^{(j)} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \vdots & Z_n^{(j)} \end{pmatrix}.$$

The likelihood function of model M_j is (conditional on the first observation)

$$P(Y|\beta^{(j)}, \Sigma^{(j)}, M_j) = (2\pi)^{-\frac{Tn}{2}} |\Sigma^{(j)}|^{-\frac{T}{2}} \exp\left\{ -\frac{1}{2}(y - Z^{(j)}\beta^{(j)})'(\Sigma^{(j)} \otimes I_T)^{-1}(y - Z^{(j)}\beta^{(j)}) \right\}, \tag{7}$$

where I_T is an identity matrix of dimension T.

3.2 Prior

For every model M_j we define a model prior $p(M_j)$, a prior distribution for $\Sigma^{(j)}$ given M_j $p(\Sigma^{(j)}|M_j)$ and a prior distribution for the slope parameters $\beta^{(j)}$ given M_j and $\Sigma^{(j)}$ $p(\beta^{(j)}|M_j, \Sigma^{(j)})$.

Firstly, we consider the model prior. The model prior probability is given as follows

$$p(M_j) \propto q^{|M_j|}(1-q)^{n^2 - |M_j|}, \tag{8}$$

where q is the prior probability that a variable is included and $|M_j|$ is the total number of included variables in all equations. We set q = 0.50, which implies that every model has the same prior probability

$$p(M_j) = \frac{1}{n^2}. \tag{9}$$

This is the standard choice in the literature, see for example Avramov (2002). As a robustness check, we also consider putting a beta-prior on q as in Ley and Steel (2009). We postpone further details on the robustness check until section 7.1.

Secondly, we consider the prior on the error covariance matrix. In general we cannot use improper priors for the coefficients within a model if we use BMA techniques. However, since all elements of $\Sigma^{(j)}$ are common to every model M_j, we can choose a standard improper prior for $\Sigma^{(j)}$. Therefore, we choose the standard improper Jeffrey's prior for $\Sigma^{(j)}$

$$p(\Sigma^{(j)}|M_j) \propto |\Sigma^{(j)}|^{-\frac{n+1}{2}}. \tag{10}$$

Finally, we look at the prior on the slope coefficients. The different elements of $\beta^{(j)}$ are not common to all considered models. Choosing an uninformative prior for $\beta^{(j)}|M_j\Sigma^{(j)}$ would therefore lead to an ill-defined posterior odds ratio $\frac{P(M_j|Y)}{P(M_i|Y)}$. Hence, we choose a proper prior distribution instead

$$p(\beta^{(j)}|\Sigma^{(j)}, M_j) \sim N(m^{(j)}, gV^{(j)}), \tag{11}$$

i.e. a multivariate normal distribution with $k^{(j)} \times 1$ mean vector $m^{(j)}$ and covariance matrix $gV^{(j)}$ with scalar g and $k^{(j)} \times k^{(j)}$ matrix $V^{(j)}$.

The prior mean $m^{(j)}$ consists of 1s and 0s. Most elements are equal to 0. Only the prior means on the slope coefficient of the most persistent lagged predictor variables (the dividend-to-price ratio, the book-to-market ratio, the price-earnings ratio, the smoothed nominal yield, the yield spread, the credit spread, net stock issues and the stock return variance) are set equal to 1 in their own equations. This prior mean reflects the prior belief that excess stock and bond returns are unpredictable and that the persistent predictor variables follow a random walk. The prior choice expresses a degree of skepticism about the predictability of stock returns as in Kandel and Stambaugh (1996) and Wachter and Warusawitharana (2009).

The choice of the prior covariance matrix $gV^{(j)}$ is less straightforward. We let $V^{(j)}$ depend on data $Z^{(j)}$ and on covariance matrix $\Sigma^{(j)}$

$$V^{(j)} = \left(Z^{(j)\prime} (\Sigma^{(j)} \otimes I_T)^{-1} Z^{(j)} \right)^{-1}. \tag{12}$$

In the special case where the explanatory variables $Z_i^{(j)}$ for equation i are equal to $Z^{(j*)}$ for all i, $V^{(j)} = \Sigma^{(j)} \otimes \left(Z_1^{(j*)\prime} Z_1^{(j*)} \right)^{-1}$. Our prior for $\beta^{(j)} | M_j, \Sigma^{(j)}$ is an empirical Bayes prior. We choose this prior because of two important reasons. Firstly, the prior is not sensitive to linear transformations of the data. Secondly, the prior covariance matrix (accurately) reflects the belief that the slope coefficients are strongly correlated. If instead we would specify a diagonal prior covariance matrix, we would set the prior correlation between slope coefficients equal to zero while in fact the slope coefficients are strongly correlated in the data. This conflict of information between the data and the prior would lead to a distorted posterior distribution with unintended consequences. The main drawback of letting $V^{(j)}$ depend on $\Sigma^{(j)}$ is that it complicates the MCMC algorithm that we use to calculate the results.

Scalar g determines the strength of the prior information within a model. A high g means that we are relatively uninformative about the parameters within a model. Therefore, it would seem natural to set parameter g to a very high number, e.g. $g = 10^6$. However, Fernandez, Ley, and Steel (2001) show that setting g equal to such a high value in order to be uninformative about the coefficients within a model implies that we are in fact very informative about the models that receive high posterior probability. It would mean that we put a lot of posterior probability on models with a small number of explanatory variables.[10] Fernandez, Ley, and Steel (2001) suggest to set $g = T$ when the square of the maximum number of considered explanatory variables per equation is smaller than T. Therefore, we set $g = T$, which means that the prior contains as much information as one observation. As a robustness check, we also consider putting a prior on g. Further details on the robustness check are postponed until section 7.1.

[10] By looking at equation (16) in the next section (the expression for the marginal likelihood conditional on $\Sigma^{(j)}$), we see that g acts as a penalty factor for larger models. The larger g, the more large models are penalized.

3.3 Posterior

We want to estimate the posterior model probabilities $p(M_j|Y)$, the posterior distributions for the slope coefficients $p(\beta^{(j)}|Y, M_j)$ and the posterior distributions for the error covariance matrix $p(\Sigma^{(j)}|Y, M_j)$. Unfortunately, none are analytically available in our setting. Therefore, we have to obtain the posterior distributions through MCMC simulation techniques. In order to be able to use the MCMC techniques that are introduced in the next section, we have to derive several expressions. The derivations themselves are shown in the appendix.

Let us firstly consider the posterior distributions of $\beta^{(j)}$. The properties of the posterior distributions of $p(\beta^{(j)}, \Sigma^{(j)}|Y, M_j)$ or $p(\beta^{(j)}|Y, M_j)$ are unknown since the integrating constant is unknown. However, the conditional posterior distribution $p(\beta^{(j)}|Y, M_j, \Sigma^{(j)})$ is analytically available

$$p(\beta^{(j)}|Y, M_j, \Sigma^{(j)}) = N(\beta^{*(j)}, M^{*(j)}) \tag{13}$$

where

$$\begin{aligned}
M^{*(j)} &= \left[\left(\frac{1+g}{g}\right) Z^{(j)\prime}(\Sigma^{(j)} \otimes I_T)^{-1} Z^{(j)}\right]^{-1} \\
\beta^{*(j)} &= \frac{g}{1+g}\hat{\beta}^{(j)} + \frac{1}{1+g} m^{(j)} \\
\hat{\beta}^{(j)} &= \left(Z^{(j)\prime}(\Sigma^{(j)} \otimes I_T)^{-1} Z^{(j)}\right)^{-1} Z^{(j)\prime}(\Sigma^{(j)} \otimes I_T)^{-1} y^{(j)}.
\end{aligned}$$

In the next section, we will use $p(\beta^{(j)}|M_j, \Sigma^{(j)}, Y)$ in a Gibbs step.

Secondly, let us look at the posterior distribution for $\Sigma^{(j)}$. We do not know the integrating constant for the conditional posterior $p(\Sigma^{(j)}|Y, M_j, \beta^{(j)})$ or for $p(\Sigma^{(j)}|Y, M_j)$, but we know that the conditional posterior is proportional to the following expression

$$p(\Sigma^{(j)}|Y, M_j, \beta^{(j)}) \propto |Z^{(j)\prime}(\Sigma^{(j)} \otimes I_T)^{-1} Z^{(j)}|^{\frac{1}{2}} |\Sigma^{(j)}|^{-\frac{T+n+1}{2}} \exp\left\{-\frac{1}{2} tr[(\Sigma^{(j)})^{-1}(E^{(j)\prime}E^{(j)} + \frac{1}{g}H^{(j)})]\right\}, \tag{14}$$

where

$$\begin{aligned}
E^{(j)} &= (Y - W^{(j)}B_c^{(j)})'(Y - W^{(j)}B_c^{(j)}) \\
W^{(j)} &= (Z_1^{(j)}, Z_2^{(j)}, ..., Z_n^{(j)}) \\
H^{(j)} &= (B_f^{(j)\prime} - M_f^{(j)\prime})' X' X (B_f^{(j)\prime} - M_f^{(j)\prime}),
\end{aligned}$$

$$B_c^{(j)} = \begin{pmatrix} \beta_1^{(j)} & 0 & \dots & 0 \\ 0 & \beta_2^{(j)} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \vdots & \beta_n^{(j)} \end{pmatrix}.$$

and where matrix $B_f^{(j)}$ is an $n \times (n+1)$ matrix. Elements of $B_f^{(j)}$ are either equal to 0 when a variable is not included in the model or equal to its corresponding element in $\beta^{(j)}$.[11] The $n\times(n+1)$ matrix $M_f^{(j)'}$ is similarly defined with respect to the prior means in $m^{(j)}$. Expression (14) allows us to approximate the conditional posterior distribution using a Metropolis-Hastings step in the next section.

Finally, let us consider the posterior model probability $p(M_j|Y)$. If we would know the marginal likelihood $p(Y|M_j)$, the posterior probability that model M_j is the true model could be calculated as

$$p(M_j|Y) = \frac{P(Y|M_j)p(M_j)}{\sum_k P(Y|M_k)P(M_k)}. \tag{15}$$

BMA papers that rely on simulation techniques to explore the model space (e.g. Fernandez, Ley and Steel, 2001) use an MC^3 algorithm to explore the model space. However, this is not applicable in our setting, since $p(Y|M_j)$ is not analytically available for a Seemingly Unrelated Regression (SUR) model with arbitrary restrictions. We are only able to analytically calculate $p(Y|M_j)$ in a couple of limited cases, e.g. when the set of right-hand-side variables is equal in all equations. Therefore, in the next section we develop an algorithm to explore the model space that only requires us to know the "conditional marginal likelihood" $p(Y|M_j,\Sigma^{(j)})$. The conditional marginal likelihood is analytically available

$$\begin{aligned} p(Y|M_j,\Sigma^{(j)}) = (2\pi)^{-\frac{Tn}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}}(1+g)^{-\frac{k^{(j)}}{2}} \exp\left\{-\frac{1}{2}(y'(\Sigma^{(j)}\otimes I_T)^{-1}y)\right\} \\ \exp\left\{-\frac{1}{2}\left(m^{(j)'}\frac{1}{g}V^{(j)-1}m^{(j)} - \beta^{(j)*'}M^{*(j)-1}\beta^{(j)*}\right)\right\}. \end{aligned} \tag{16}$$

3.4 Markov Chain Monte Carlo algorithm

We apply Markov Chain Monte Carlo techniques to estimate the posterior distributions $p(M_j|Y)$, $p(\beta^{(j)}|Y,M_j)$ an $p(\Sigma^{(j)}|Y,M_j)$. The expressions we derive in the previous section allow us to implement the algorithm. Suppose we are currently in iteration m with model M_j and suppose that the current coefficients for model M_j in iteration m are $\beta^{(j)}$ and $\Sigma^{(j)}$.

The first step is to draw a new model M_l in iteration $m{+}1$. The modeling step is a Metropolis-Hastings step and is based on a result in Godsill (2001) who shows that we can condition on parameters that are shared between different models - $\Sigma^{(j)}$ in our setting - when drawing a new model. Our method only requires that $\Sigma^{(j)}$ has a common interpretation and common dimensions across models, but does not require that $\Sigma^{(j)}$ has the same posterior distribution across different models. We randomly draw a proposal model from the neighbourhood of models around M_j and then accept the model with acceptance probability α.

Assume that the current model has $k^{(j)}$ included variables. The algorithm proceeds as follows.

1. Randomly select a proposal model M_l^* from all models with $k^{(j)}+1$ or $k^{(j)}-1$ variables.

[11]For example, if element (x_s, DP) is included, element (2,6) of $B_f^{(j)}$ is equal to its corresponding value in $\beta^{(j)}$, otherwise it is 0.

2. Set $M_l = M_l^*$ with acceptance probability

$$\alpha = \min\left\{1, \frac{p(M_l^*)p(Y|M_l^*, \Sigma^{(j)})}{p(M_j)p(Y|M_j, \Sigma^{(j)})}\right\}. \tag{17}$$

Otherwise set $M_l = M_j$.

The appendix provides details on the calculation of the acceptance probability.

The step for drawing slope coefficients $\beta^{(l)}$ is more straightforward. It is a standard Gibbs steps and is based on the conditional posterior we derived in the previous section. Note that we condition on new model M_l and old draw $\Sigma^{(j)}$.

1. Draw $\beta^{(l)}$ using the conditional posterior in equation (13).

We do not know the integrating constant of the conditional posterior of $\Sigma^{(l)}$ and therefore cannot use a standard Gibbs step. Instead, we draw $\Sigma^{(l)}$ using a Metropolis-Hastings step. As proposal density, we use an inverted Wishart distribution that approximates the conditional posterior density for $\Sigma^{(l)}$ as close as possible. Note that we condition on the new model M_l and the new draw $\beta^{(l)}$.

1. Draw $\Sigma^{(l)*}$ according to iWishart$(E^{(l)\prime}E^{(l)} + H^{(l)}, T + n + 1)$.
2. Set $\Sigma^{(l)} = \Sigma^{(l)*}$ with acceptance probability

$$\alpha = \min\left\{1, \frac{|Z^{(l)\prime}(\Sigma^{(l)*} \otimes I_T)^{-1}Z^{(l)}|^{\frac{1}{2}}|\Sigma^{(l)*}|^{\frac{n+1}{2}}}{|Z^{(l)\prime}(\Sigma^{(j)} \otimes I_T)^{-1}Z^{(l)}|^{\frac{1}{2}}|\Sigma^{(j)}|^{\frac{n+1}{2}}}\right\}. \tag{18}$$

Otherwise set $\Sigma^{(l)} = \Sigma^{(j)}$.

The appendix provides details on the calculation of the acceptance probability. Note that we use the parameterization of Bauwens, Lubrano, and Richard (1999) for the inverted Wishart distribution.

3.5 Comparison to alternative specifications

In this section, we compare our weighted Bayesian Model Averaging specification to alternative specifications that are proposed in the literature.

The first alternative specification is proposed in Avramov (2002). The models he considers are VAR(1) models. Since his focus is mainly on short-horizon forecasting, he bases the posterior model probabilities only on the stock return equation within the VAR(1) model. He mentions "(..) the weighted predictive distribution makes use of posterior probabilities computed based on the return generating process in Eq.(1)" (Avramov, 2002, page 432). The implication is that model uncertainty regarding the prediction of the predictor variables is not taken into account. Therefore, given our long-run focus, we cannot use Avramov's (2002) setting since it would substantially underestimate the impact of model uncertainty for long horizons.

A second alternative specification is used in George, Ni, and Sun (2008). They suggest a Bayesian stochastic search method to select restrictions in VAR models. They put a tight prior centered around 0 on the slope coefficients of irrelevant right-hand-side variables instead of considering models where irrelevant right-hand-side variable are excluded as in our setting. The idea is that the use of such a prior sets the slope coefficient arbitrarily close to 0 such that the effect of these variables is negligible. An issue with George, Ni and Sun's (2008) approach is the calibration of the tightness of the prior for each slope coefficient. They need to specify a large number of additional prior hyperparameters and in our experience the results turn out to be very sensitive to this calibration. If the prior is too tight the stochastic search algorithm does not converge, while if the prior is too loose the effect of these variables could still be substantial. Our method avoids the calibration of these parameters and therefore avoids these issues.

A third alternative specification is analyzed in Andersson and Karlsson (2008). They base the posterior model probabilities on the actual forecasting performance of the variable of interest at a certain horizon. Andersson and Karlsson's (2008) method uses simulation techniques to calculate the predictive performance of each individual model at each horizons. We do not use their method, because of several reasons. Firstly, their approach is not feasible in our setting, since we consider a very large number of models and it is impossible to calculate the predictive performance of each model individually. Secondly, our objective is to find one combination of models that describes the predictive distribution at different horizons, while their method leads to different posterior model probabilities at different horizons. Thirdly, at long horizons we cannot reliable calculate the predictive performance of individual models due to data limitations. For example, if we consider a horizon of 30 years, the predictive performance of each model is based on only two non-overlapping 30-year windows. This is clearly undesirable.

A fourth alternative method is proposed in Wright (2008). He considers models that regress the variable of interest at time point $t+k$ on right-hand-side variables at time point t for horizon k. He uses these models to directly forecast variables k-period ahead (so-called direct forecasts). We instead consider models that regress next period's variable of interest on current right-hand-side variables and iterate these models k-periods forward in order to get predictions at a k-period horizon. We consider iterated forecasts because of several reasons. Firstly, Marcellino, Stock, and Watson (2006) show that iterated forecasts outperform direct forecasts at longer forecast horizons in an empirical analysis using macro-economic time-series. The iterated forecasts are more efficient. Secondly, when $k > 1$ the overlapping nature of observations on the variable of interest leads to serial correlation in the error term when one uses the models in Wright (2008). This is an econometric issue one needs to take into account when using the direct forecasting method. Thirdly, Wright's (2008) method gives different forecasting models for different horizons. We instead want to find a parsimonious model that describes the predictive distribution of asset returns at different horizons. Finally, when the forecast horizon is large we cannot reliably estimate the model in Wright (2008). There are not enough non-overlapping observations when we consider forecasting horizons of for example 30 years.

4 Estimation Results

As a first step we estimate the unrestricted VAR(1) - the model in equation (3) - on the full data-set. Table 2 gives the OLS parameter estimates, its standard errors and the correlations and standard deviations of the residuals.

[Table 2 about here.]

We focus on the most important results. Firstly, since the maximum eigenvalue of 0.9902 is less than 1, the system is strictly-speaking stationary. However, since the value is close to 1, we need to be careful in interpreting the frequentist t-statistics below. As mentioned above, this does not have an impact on (the interpretation of) the posterior distributions. Secondly, the R^2's in the equations for stock and bond returns are respectively 9.0% and 10.1%. This implies that a large part of the return variation still remains unexplained. Thirdly, the t-statistics suggest that the price-earnings ratio is the most important predictor of excess stock returns followed by the book-to-market ratio, net stock issuance, the credit spread and the default premium. Surprisingly, the popular dividend-to-price ratio is less important, but this can be due to the fact that the dividend-to-price ratio, the book-to-market ratio and the price-earnings ratio are highly collinear. Finally, the table shows that the dividend-to-price ratio, the book-to-market ratio, the price-earnings ratio, the smoothed nominal yield, the yieldspread, the credit spread, net stock issues and the stock variance are very persistent processes.

Next, let us consider the Bayesian Model Averaging (BMA) specification. Table 3 reports the posterior probability that a right-hand-side variable is included in a particular equation in the weighted BMA specification. The posterior probability of variable Y in equation X is obtained by summing the posterior model probabilities of all models in which variable Y occurs in equation X. The different equations are given in the different rows.

[Table 3 about here.]

Firstly, the table shows that the posterior probabilities vary a lot across equations. The most extreme example is the stock variance. In 10 out of 12 equations, it is only rarely included in the BMA model (see the last column), but in 2 out of 12 equations - namely the credit spread equation and the stock variance equation itself - the stock variance is always included. Hence, if we would exclude the stock variance from the set of predictors, the equation for the credit spread would be misspecified. However, if we would include the stock variance in all equations we would obtain inefficient estimates, since the stock variance has no predictive power in most equations. Therefore it is clearly important that we allow the set of right-hand-side variables to differ across equations.[12] Secondly, we conclude that the posterior probabilities are only roughly related to the frequentist t-statistics that are reported in the previous table. In some cases, there is a correspondence. For example, in most predictor equations the lag of the dependent variable both has a high t-statistic and a high posterior probability. However, let us also consider the dividend-to-price ratio, the book-to-market ratio and the price-earnings ratio in the equation for stock returns. Although 2 out of 3 (the book-to-market ratio and the price-earnings ratio) have high t-statistics, only 1 of them (the dividend-to-price ratio) is frequently included in the weighted BMA model. The reason is that the dividend-to-price ratio, the book-to-market ratio and the price-earnings ratio are highly collinear and therefore contain almost the same information. The BMA specification therefore only includes one of them - the dividend-to-price ratio - on average. Thirdly, the posterior probability that the system is non-stationary is 6.2%. Although Barsky and De Long (1993) show that under certain assumptions the dividend-to-price ratio is an I(1) process, our analysis shows that there is little evidence for this claim. Fourthly, the table shows that between 2 - 4 variables out of 12 are on average included in most equations. Hence, the full model clearly overfits the data. Fifthly, we see that the dividend-to-price ratio and the credit spread are the most important direct predictors of excess log stock returns. The marginal probabilities that these variables are included are larger than 84% for both. Furthermore, the most important predictor of the T-bill rate is its own lag and the most important predictor of excess bond returns is the yield spread.

Since the dividend-to-price ratio and the credit spread are by far the most important predictors of excess stock returns, it seems natural to conclude that the other 10 variables are not important to model future stock returns. However, this is incorrect. To see this, let us again consider the stock variance. The current value of the stock variance only predicts next period's stock returns with a posterior probability of 9.00%. However, note that if we want to predict excess stock returns two periods in the future, we first have to be able to predict next period's credit spread and note that equation 10 in the table shows that the stock variance is the most important predictor of next period's credit spread. Hence, if we want to predict stock returns two periods ahead, the stock variance is one of the most essential predictors of future stock returns. If we would have only based model probabilities on the stock return equation, we would have wrongly concluded that the stock variance is not an important predictor of stock returns. Therefore, it is essential that model probabilities not only reflect which predictors accurately predict next period's stock returns, but also which predictors predict predictors itself.

There are more and more predictors that become important at longer horizons. Therefore, we develop a measure to assess the importance of predictors of excess stock returns at different horizons. We explain this measure using an example. Firstly, suppose that the considered model is an unrestricted VAR(1) model with known slope coefficients A and suppress constants

[12] As an example, if we restrict the set of right-hand-side variables to be equal across all equations, the most persistent predictor variables are always included. The reason is that the lags of the persistent predictors need to be included to model the persistent predictor variables themselves. Since the right-hand-side variables are the same for all equations in this example, the lags are included in all equations. This is clearly undesirable.

for notational convenience. The following holds

$$\begin{aligned} E_t y_{t+1} &= A y_t \\ E_t y_{t+2} &= A^2 y_t \\ &\vdots \\ E_t y_{t+k} &= A^k y_t. \end{aligned} \tag{19}$$

We can use this result to calculate the importance of the right-hand-side variables for predicting excess stock returns at different horizons in the following way. Suppose we want to analyze whether the lagged book-to-market ratio is an important predictor of stock returns k periods in the future. In order to do so, we firstly draw a model and a parameter draw from our posterior distribution. Then, we rewrite our model as a restricted VAR(1) model and iterate the model forward by k periods.[13] Next, we check whether element (X_s, BM) of matrix A^k is different from 0. If it is, the book-to-market ratio is a predictor of excess stock returns at horizon k for this parameter draw in this particular model. We repeat this process N times for different model and parameter draws. Finally, we calculate the fraction of draws for which the book-to-market ratio is a predictor of excess stock returns. We use this fraction as a measure for the importance of the book-to-market ratio for predicting excess stock returns at horizon k.

[Table 4 about here.]

Table 4 shows the measure for horizons up to 40 quarters. For a horizon of 1 quarter, the measure is equal to the posterior probabilities that are given in table 3. When horizon k increases, the measure substantially deviates from these probabilities. As argued before we see that the stock variance becomes an important predictor of excess stock returns when horizon $k \geq 2$. All 12 right-hand-side variables become important at a horizon of 10 years. At these long investment horizons variables either predict stock returns directly or predict the predictors of stock returns or the predictors of the predictors of stock returns etcetera.

Table 5 reports the posterior moments of the coefficients of the weighted BMA model. The table shows the posterior means and standard deviations of the slope coefficients in panel A and the posterior mean of the covariance matrix of the residuals in panel B.[14]

[Table 5 about here.]

Firstly, the results in the table are roughly in line with table 3. The posterior means of variables that are hardly ever included in the model are close to 0. Secondly, the table shows that a higher credit spread implies a lower future excess stock return, while a higher dividend-to-price ratio leads to a higher future excess stock return. Both signs are in line with the OLS

[13]Note that the different models we consider can easily be rewritten as restricted VAR(1) models with 0s in place of excluded variables.

[14]The reported numbers are the unconditional moments, i.e. the posterior means of the coefficients are not conditional on the inclusion of the variables in the model. When a variable is not included in a particular model, the posterior mean is equal to 0 for that model.

results. Next, the maximum eigenvalue of 0.9823 indicates that the system is stationary (using the means of the parameters as estimates). Finally, the R^2 of the equations for stock and bond returns are respectively 2.33% and 5.45%. This is a lot lower than the R^2's we obtain when we estimate the unrestricted VAR(1) model by OLS. This finding is not surprising, since the BMA model only includes 3 variables on average per equation, while the full model includes 12 variables per equation. The full model clearly overfits the data.

The posterior model probabilities are conditional on the considered model space. The used methodology could easily be extended to a larger model space. For example, if one believes that the true model might be a VAR(2) model or might not be homoscedastic, we could in principle include such models in the (extended) model space and apply the same methodology. In fact, we can use exactly the same MCMC algorithm to incorporate the VAR(2) model. The incorporation of heteroscedastic models is more difficult (we cannot condition on a constant error covariance matrix anymore) but can be done.

How does the posterior distribution of the coefficients look like in the BMA specification? Figure 1 shows the posterior distribution of the coefficient on the credit spread in the equation for stock returns, (x_s, Cr_{spr}).

The figure shows that the posterior distribution is clearly non-normal. Firstly, there is a spike at 0. This spike corresponds to the probability that the credit spread is not included in the stock return equation. Secondly, even when the credit spread is included in the model we see that the posterior distribution is bi-modal. The bi-modality is caused by the presence or absence of other variables in the model. In case the credit spread is included in the equations for stock returns and the dividend-to-price ratio, the posterior distribution of (x_s, Cr_{spr}) is equal to the right bell-shaped curve in the figure. If the spread is only included in the stock return equation, the posterior distribution is given by the left bell-shaped curve.

Do most variables occur independently from each other in the composite model? In order to analyze this question we analyze the variables that occur jointly in the weighted BMA model (these variables are complements) and the variables that occur disjointly in the weighted model (these variables are substitutes). We use the measure developed in Ley and Steel (2007) to quantify whether pairs of variables are complements or substitutes.

[Figure 1 about here.]

[Table 6 about here.]

Consider the prediction equation for excess stock returns. Let $P(i)$ be the marginal probability that variable i is included in this equation, $P(j)$ be the marginal probability that variable j is included and $P(i \bigcap j)$ the probability that both variables are included. Ley and Steel's (2007) measure equals

$$P_{ij} = \frac{P(i \bigcap j)}{P(i) + P(j) - 2P(i \bigcap j)}, \tag{20}$$

where a high value (larger than 3) indicates that variables i,j occur jointly in the weighted model and where small values (smaller than 1/3) indicate that variables i,j occur disjointly in the model.

[Table 7 about here.]

Table 6 reports P_{ij} for pairs of variables i and j in the prediction equation for excess stock returns. It reports the (posterior) jointness measure for a horizon of 1. The table shows that there are hardly any pairs of variables that are complements except the dividend-to-price ratio /credit spread pair with a P_{ij} relatively close to 3. For all other pairs the P_{ij} values are much smaller than 1. Consider for example the dividend-to-price ratio/book-to-market ratio pair. A P_{ij} value of 0.0905 suggests that there is strong evidence that the dividend-to-price ratio and the book-to-market ratio are substitutes in the stock return equation. This is not surprising, since the correlation between DP and BM is very high.

The method we outline in the previous section to calculate the posterior distribution of the composite BMA model can also be used to select one best model, i.e. the model with the highest posterior probability. In the next sections, we compare the highest posterior probability model with the overall weighted BMA model. Table 7 shows the variables that are included in the highest posterior probability model and their posterior means and standard deviations.

The incorporated variables are roughly in line with the results in table 3. The posterior probability of the highest probability model is not large, i.e. less than 0.001. This suggests that the posterior probability is widely spread over many models and suggests that model uncertainty is prevalent even after using more than 80 years of data.

5 The term structure of risk

In this section we consider the variance of the predictive distribution of future cumulative excess stock returns. Since we are working with log stock returns, we can easily calculate cumulative excess stock returns as follows

$$x_{s,t\to t+K} = x_{s,t+1} + x_{s,t+2} \ldots . x_{s,t+K},$$

where $x_{s,t\to t+K}$ is the cumulative excess stock return from period t to period $t+k$.

The predictive distribution of cumulative excess stock returns is given in the following equation

$$p(x_{s,t\to t+K}|Y) = \int p(x_{s,t\to t+K}|Y, M_j, \beta^{(j)}, \Sigma^{(j)}) p(\beta^{(j)}, \Sigma^{(j)}|Y, M_j) p(M_j|Y) d(\beta^{(j)}) d(\Sigma^{(j)}) d(M_j).$$

This distribution incorporates both parameter and model uncertainty. The first component is the distribution of future stock returns conditional on a model and parameter values. This distribution is normal, ignores both parameter and model uncertainty and is used in Campbell, Chan, and Viceira (2003) and Campbell and Viceira (2005) among others. The second component is the posterior distribution of the parameters conditional on a particular model. We use it to include parameter uncertainty by integrating over the parameter space. Barberis

(2000) and Hoevenaars, Molenaar, Schotman, and Steenkamp (2007) also consider settings that incorporate parameter uncertainty. The last component is the posterior model probability. We use this component to sum over the model space and take model uncertainty into account. This is unique to our setting.

We present the results using the term structure of risk. It plots the annualized standard deviation of the predictive distribution of cumulative excess stock returns versus the investment horizon. If annualized volatility at long horizons is smaller than at short horizons, stocks are safer in the long-run. Campbell and Viceira (2005) show in a setting without parameter uncertainty that this term structure is downward sloping due to the mean reversion in stock returns. Hoevenaars, Molenaar, Schotman, and Steenkamp (2007) show that the incorporation of parameter uncertainty increases the volatility, that it leads to an upward sloping term structure for long investment horizons, but that the annualized volatilities at horizons up to 50 years are still lower than the annualized volatility at investment horizons around one year. Both papers therefore conclude that stocks are less volatile in the long-run than in the short-run.

Pastor and Stambaugh (2010) consider a different framework. They explicitly take into account that predictor variables are imperfect and that linear functions of a small number of predictor variables are likely to be less than perfectly correlated with the (unknown) true expected future stock return. Pastor and Stambaugh (2010) show that stocks are much riskier in the long-run than in the short-run in that setting.

In this paper, we consider a different setting than in the previously mentioned papers. We make the basic assumption that the expected stock return can be captured by a linear combination (of a subset of) 12 predictor variables. Hence, instead of assuming that the expected stock return is imperfectly correlated with a small number of predictor variables, we assume that it is perfectly correlated with a subset of a much larger number of predictor variables. We explicitly take into account that we are uncertain about the correct model and that we also face parameter uncertainty.[15]

[Figure 2 about here.]

Firstly, let us analyze whether stocks mean-revert in our setting. An indication for mean reversion is that there is a negative correlation between a shock to current stock returns and a shock to the expectation of future stock returns. Our objective is to find the posterior distribution of this correlation. We explain this measure using an example. Suppose for a moment that the model we consider is an unrestricted VAR(1) model and suppress constants for notational convenience

$$y_{t+1} = Ay_t + \epsilon_t. \tag{21}$$

Define A_2 as the second row of A. Obviously, $\epsilon_{t,2}$ is the shock to excess stock returns. Furthermore, since $A_2 y_t$ is the predicted value of next period's stock return, $A_2 \epsilon_t$ is the shock to the

[15] In this section, we plot the term-structure of risk for *excess* stock returns. The term structure of risk for *real* stock returns lies strictly above the term structure of risk for excess stock returns and is steeper due to the strong mean-aversion in the real T-bill rate.

expectation of next period's stock return. Therefore, the correlation between $A_2\epsilon_t$ and $\epsilon_{t,2}$ is our measure for mean-reversion. Our aim is to find the posterior distribution of this correlation.

Figure 2 plots this posterior distribution for our weighted BMA model. The figure shows that the posterior probability that the correlation is negative is 98.4% which means that it is very likely that stocks exhibit mean reversion. Furthermore, the probability that the correlation is smaller than -0.5 is still 75.7%. Hence, there is a lot of evidence that there is very strong mean reversion in stock returns.

Next, we analyze the term structure of risk. We decompose the (annualized) total predictive variance of excess log stock returns in the (annualized) mean of the conditional variance and the (annualized) variance of the conditional mean. Note that the conditional mean and conditional variance are conditional on a particular model and set of parameters.

$$\frac{1}{K}V(x_{s,t\to t+K}|Y) = \frac{1}{K}E(V(x_{s,t\to t+K}|Y, M_j, \beta^{(j)}, \Sigma^{(j)})) + \frac{1}{K}V(E(x_{s,t\to t+K}|Y, M_j, \beta^{(j)}, \Sigma^{(j)}))$$

Figure 3 plots the term structure of risk using the weighted BMA model when variables are set equal to their historical average. It shows total volatility, the square root of the mean of the conditional variance and the volatility of the conditional mean. In order to obtain the posterior distribution, we estimate the specification on the full data-set.

[Figure 3 about here.]

[Figure 4 about here.]

The picture shows that stocks are almost as risky in the long-run as in the short-run. The annualized total volatility is 22% at short horizons and around 20% at horizons up to 30 years. What effects play a role? Firstly, the mean-reversion effect is important. If a long-term investor sees a bad stock return, he knows that on average the bad stock return will be followed by a better stock return due to mean-reversion. This leads to a negative correlation in stock returns and makes stock returns safer in the long-run. The second effect is the parameter and model uncertainty effect. If a long-term investor sees a bad stock return, it could also be the case that the true model and the true parameter set (the ones he does not know) are relatively unfavourable for him. Since this will persist in the future as well, this will create a positive autocorrelation in stock returns in the eyes of an investor who does not know the true model and does not know the true parameter values. Therefore, the model and parameter uncertainty effects make stock returns riskier in the long-run. Hence, whether stocks are riskier or safer in the long-run depends on the magnitude of both opposing effects and is an empirical matter. In the figure, mean-reversion still dominates parameter and model uncertainty.

The line that depicts the volatility of the conditional mean in the figure (the red line) shows the parameter and model uncertainty effect. Different models (and parameters) give different predictions of future stock returns, i.e. they have different conditional means. The volatility of these different conditional means therefore depicts the parameter and model uncertainty effect. At short horizons the parameter and model uncertainty effect has a negligible impact, because at short horizons the different models predict very similar future stock returns. However, at longer horizons the parameter and model uncertainty effect becomes really important, since the different models predict very different trajectories of future stock returns. At a horizon of 30 years this effect plays a relatively large role, but not enough to make stock returns riskier in the long-run. The other component of total volatility - (the square root of) the mean of the conditional variance (green line) - depicts the residual component of total volatility. It is largely downward sloping due to mean reversion.

The figure shows the term structure of risk when the variables are set equal to their historical average. In that case, the different models agree on future stock returns and therefore the parameter and model uncertainty effects are not large. If variables deviate substantially from their historical average, the different models predict very different future stock returns. In that case parameter and model uncertainty becomes really important. Figure 4 plots the term structure of risk using the weighted BMA model when variables are set equal to their Q4 2008. At this date, we were in the middle of the sub-prime mortgage crisis and all predictors deviated substantially from their historical average.

If we compare the two figures, we see that the mean of the conditional variance turns out to be exactly equal to its values in the previous figure. However, the variance of the conditional mean changes substantially. As expected we see that the variance of the conditional mean becomes much larger when the variables deviate significantly from their historical average. Figure 4 also shows that the variance of the conditional mean dominates the mean of the conditional variance as the most important component at a 30 year horizon. Total volatility increases due to the larger variance of the conditional mean. It is again 22% for short horizons, decreases slightly for medium horizons and increases up to 28% at a horizon of 30 years. Hence, the volatility at a 30-year horizon is larger than at short horizons. Hence, stocks are not safer in the long-run when we incorporate parameter and model uncertainty and when predictors deviate a lot from their historical average. The figure also shows that parameter and model uncertainty is not important at short horizons.

How does total volatility of stock returns change over time when predictor variables change over time? In order to answer this question, we plot a time-series of total volatility for different investment horizons in figure 5. The dates on the x-axis indicate the value of the predictor variables we use to calculate the term structure of risk. The posterior distribution is obtained by estimating the composite specification on the full data-set.

[Figure 5 about here.]

Firstly, annualized total volatility at an investment horizon of 1 quarter is not very sensitive to values of the predictor variables, i.e. it is relatively constant over time at a value of around 22%. Parameter and model uncertainty is not important, since volatility hardly changes over time. Secondly, we conclude that the annualized volatility at an investment horizon of 15 years is always lower than the volatility at shorter horizons. Its average value is 18%. Parameter and model uncertainty turns out to be important at this horizon, since volatility changes quite a lot over time. The mean-reversion effect however still dominates the parameter and model uncertainty effect at this investment horizon. Finally and most interestingly, the annualized volatility at an investment horizon of 30 years indicates that parameter and model uncertainty is really important, because it shows a lot of variation over time. The figure also shows that the parameter and model uncertainty effect dominates the mean-reversion effect when predictor variables are rather extreme, such as in the recent crisis or in the crisis of 1929. Annualized volatility can be as high as 30%. However, in stable times such as the 1960s, annualized volatility at a 30-year horizon is very close to annualized volatility at a 1-quarter horizon. We clearly see that the riskiness of stocks (as measured by the variance of the predictive distribution) changes a lot over time even though we only consider homoscedastic models!

These results only partially confirm results in Pastor and Stambaugh (2010). We also find that stocks can be riskier in the long-run than in the short-run. However, we find that there are also periods in which stocks are safer in the long-run.

The previous three figures consider a setting where we include both parameter and model uncertainty. How do these figures change if we ignore model uncertainty? In order to answer that question, figure 6 plots the term structure of risk for the highest posterior probability model if variables are set equal to their historical average. In this setting, only parameter uncertainty plays a role.

[Figure 6 about here.]

[Figure 7 about here.]

The decomposition shows that the mean of the conditional variance is the biggest component of the total variance at all horizons. The total volatility is around 22% for short investment horizons, decreases to 15.5% for moderate investment horizons and slightly increases again towards 17% for investment horizons up to 30 years. Results suggest that stocks are safer assets in the long-run.

Figure 7 shows the term structure of risk when variables are set equal to Q4 2008 values. The figure shows that the mean of the conditional variance is again exactly equal to the values in the previous picture. It also shows that the variance of the conditional mean increases slightly. Therefore, the total volatility is slightly higher than in the previous figure. The annualized volatility at the 30 year horizon is around 18%. Hence, if we ignore model uncertainty we (incorrectly) find that stocks are safer in the long-run than in the short-run no matter how we set the predictor variables. This confirms the results in Hoevenaars, Molenaar, Schotman, and Steenkamp (2007): parameter uncertainty increases risk at longer horizons but does not change the fact that stocks are safer in the long-run.

In the previous section we conclude that the full model - the unrestricted VAR(1) that includes all variables - overfits the data considerably. Since this model contains too many parameters, we expect that this model significantly overestimates the uncertainty of future stock returns. Do the term structures of risk for the full model confirm this suspicion? Figures 8 and

9 show the term structures of risk when variables are set equal to respectively their historical average and the end-of-sample values.

[Figure 8 about here.]

[Figure 9 about here.]

If we compare the results in the figures with the other figures in this section, we clearly see that the full model - as expected - overestimates the true uncertainty considerably, especially at the end of the sample. Figure 9 shows that the parameter uncertainty effect - depicted as the volatility of the conditional mean in the figure - is extremely high at this point in time.

Note that the full model is one of the models that we consider in calculating the BMA specification. However, it receives an extremely low weight in the composite model, because it contains way too many parameters. Therefore, the uncertainty in the full model plays a negligible role in the composite specification. Hence, if one only considers the full model without excluding irrelevant variables, one considerably overestimates the true uncertainty an investor faces when predicting stock returns.

We conclude that the incorporation of model uncertainty has an important effect on the variance of the conditional mean. This component dominates at longer horizons when the economy deviates from its steady state, i.e. when predictor variables are not equal to their historical average. When one recognizes model uncertainty as an important risk factor, the total volatility at long horizons could be substantially larger than at short horizons. We also find that parameter and model uncertainty has a negligible impact on short horizons.

6 Optimal portfolio choice

In this section we investigate the impact of model uncertainty on the asset allocations of long-term investors. Therefore, we start with analyzing the mean, variance, skewness and kurtosis of the predictive distribution in a setting without parameter and model uncertainty, in a setting with only parameter uncertainty and finally in a setting with both parameter and model uncertainty. Next, we consider the asset allocations in these settings.

We consider a risk-averse investor who chooses a buy-and-hold portfolio such that her expected utility is maximized and is allowed to invest in the real T-bill rate, in stock returns and in bond returns. As a utility function, we choose power utility with constant relative risk aversion parameter $\gamma > 1$

$$\max_{w_t} E_t \left(\frac{W_{t+K}^{1-\gamma}}{1-\gamma} \right) \tag{22}$$

subject to the budget constraint

$$W_{t+K} = \sum_{i=1}^{3} w_i R_{i,t+K},$$

where w is the vector of three portfolio weights, W_{t+K} is the terminal wealth at time point $t+K$, K is the investment horizon and $R_{i,t+K}$ is the cumulative gross return for asset i over K periods. We assume that short-selling is not allowed such that $0 \leq w_i \leq 1 \ \forall i$.

We solve the maximization problem in equation (22) by using simulations. As a first step, we set-up a grid of portfolio weights. Secondly, we draw N scenarios from the predictive distribution of asset returns. Next, we calculate the average realized utility for all scenarios and for all grid points. Finally, we choose the asset allocation that maximizes average realized utility. It turns out that it is troublesome to accurately calculate the kurtosis of the predictive distribution and the asset allocations for investment horizons beyond 20 years. Therefore we limit the maximum investment horizon to 20 years to guarantee the accuracy of our results.

The predictive distribution of asset returns has fat tails if we either include parameter or model uncertainty. Therefore, the expected utility of all portfolios is $-\infty$ unless we make additional assumptions, because the simple returns for all three assets can get arbitrarily close to -100%. In order to take this issue into account, we make the additional assumption that the quarterly real T-bill rate is not lower than -10%. This implies that the investor cannot go bankrupt if she invests a positive amount in the T-bill and makes sure that expected utility is finite for at least some portfolios.

[Figure 10 about here.]

Firstly, we consider the predictive distribution of future stock returns. Figure 10 plots the mean, standard deviation, skewness and kurtosis of the predictive distributions of excess stock returns for three different settings. The first setting - Plug-in - is based on the highest posterior probability model and ignores parameter uncertainty. The second setting - Parameter uncertainty HPD - is also based on the highest posterior probability model, but includes parameter uncertainty. The third setting - BMA - is based on the weighted BMA model and includes both parameter and model uncertainty. The figure plots these moments versus the investment horizon when the variables are set equal to their historical average.[16]

The figure shows that the four moments are very similar for short-investment horizons. However, if the investment horizon increases we see that the moments for the BMA predictive distribution deviate quite a bit from the moments for the other two distributions. The incorporation of model uncertainty leads to a slightly smaller mean, a higher standard deviation, slightly more negative skewness and a much higher kurtosis at long investment horizons, compared to the other two specifications. The figure shows that especially the even moments - the standard deviation and the kurtosis - are quite a bit different at long horizons. The distribution for the BMA specification clearly deviates from the normal distribution. The third and fourth moment for the plug-in method are respectively 0 and 3 for all investment horizons, since this distribution is normal.

[16] The predictive distribution of all asset returns, including the correlations between asset returns, determines the asset allocations. For the sake of brevity, our main focus in this section is on the predictive distribution of excess stock returns.

Figure 11 plots the moments of the predictive distributions when the variables are equal to their values at the end of 2008. Again, the figure shows that the four moments are almost exactly equal at short-investment horizons. However, at longer investment horizons, the moments differ a lot. The difference is much larger than in the previous figure. Firstly, the means of the specifications are very different. The reason is that the different models give very different predictions when predictor variables deviate from their historical average. The weighted BMA model takes all of them into account, whereas the highest probability model ignores the information in other models. At the end of 2008, the highest probability model is more optimistic about future stock returns than the weighted BMA model.[17] The means also deviate substantially from the means in figure 10. Secondly, the figure shows that the variance of the predictive distribution for the weighted BMA model is a lot higher than for the highest probability model. Whereas the predictive variance for the highest probability specification that includes parameter uncertainty hardly increases compared to figure 10, the predictive variance for the weighted BMA model increases substantially. The variance for the plug-in specification on the other hand is exactly the same as in the previous figure. This is consistent with the results in the previous section. Thirdly, the predictive distribution for the weighted BMA specification is negatively skewed while the specifications based on the highest probability model are (close to being) symmetric. Note that in the previous figure the distribution based on the BMA specification is also very close to symmetric. Finally, the kurtosis for the weighted BMA specification increases substantially compared to the previous figure. Hence, the figure shows that both the even and odd moments are quite a bit different at long horizons. Especially the distribution for the BMA specification is very different from the normal distribution.

[Figure 11 about here.]

[Figure 12 about here.]

How do the moments for the predictive distribution using the BMA specification change over time when variables change over time? Figure 12 plots a time-series of these moments at investment horizons of 1 quarter, 10 years and 20 years. The figure shows that all four moments change considerably over time, especially at longer investment horizons. It is remarkable how much the kurtosis changes over time. Also note that the distribution can be both negatively and positively skewed.

The previous figures show that there are substantial differences in the predictive distributions between the three different specifications. The BMA specifications leads on average to a lower mean, higher standard deviation, more negative skewness and higher kurtosis than the other two specifications. Since a risk-averse investor dislikes all of this, we expect that such an investor invests less of her money in the stock market when model uncertainty is included. We also expect that the differences between specifications are larger when variables deviate from their historical average.

[17] Note that the figures show the predictive distributions of excess stock returns. In other words, a mean of zero means that the stock returns increase on average as much as the real T-bill rate.

[Figure 13 about here.]

Figure 13 plots the asset allocations versus the investment horizon for an investor with risk aversion parameter $\gamma = 5$, where predictor variables are set equal to their historical average. The figure shows that allocations are remarkably similar at short investment horizons no matter what method we use, i.e. neither parameter nor model uncertainty significantly changes asset allocations. However, if we consider longer horizons we see that both parameter and model uncertainty play an important role. Firstly, consider the differences in allocations between the highest posterior probability model with and without parameter uncertainty. The plot shows that the asset allocation to stocks is up to 12% lower when we take parameter uncertainty into account. Secondly, consider the difference in allocations between the highest posterior probability model incorporating parameter uncertainty and the weighted BMA model. The allocation to stocks is 20% lower when we incorporate model uncertainty. Furthermore, the graph shows that the allocation to stocks decreases when the investment horizon approaches 20 years. Clearly, an investor with a longer horizon should not always invest more in the stock market than an investor with a shorter horizon, since the stock allocation of a long-term investor is remarkable similar to the allocation of a short-term investor. A long-term investor replaces the bonds in her portfolio with investments in the T-bill rate compared to the investment of a short-term investor.

[Figure 14 about here.]

[Figure 15 about here.]

[Figure 16 about here.]

What happens with asset allocations when we consider the optimal buy-and-hold asset allocations at the end of the sample? Figure 14 plots these allocations at the end of 2008. Firstly, the figure indicates that allocations to stocks are on average a lot lower compared to figure 13. Whereas the allocations range between 35% and 70% when predictor variables are equal to their historical average, these allocations vary between 0% and 45% when predictor variables are equal to their end-of-sample values. Secondly, the three methods lead to different asset allocations even at short horizons. For short investment horizons up to 2 years, the specification based on the weighted BMA model implies a 0% allocation to stocks, whereas the specifications based on the highest probability model imply allocations up to 30%. This is mainly due to the difference in means across specifications. Thirdly, at long investment horizons the impact of model uncertainty is again larger than the impact of parameter uncertainty. Incorporating parameter uncertainty lowers the allocation to stocks by 10% for the highest probability model, while the incorporation of model uncertainty decreases the stock allocation by 35%. Finally, the allocation to stocks again decreases when the investment horizon becomes very long. If we include model uncertainty, a long-term investor should not allocate more to the stock market than a short term investor.

How much expected utility do long-term investors obtain from investing in the stock market and how do these expected utilities change across specifications? In order to answer these questions, figure 15 and 16 plot the certainty equivalence (a monotonic transformation of expected utility) against the investment horizon for the three specifications.[18] We calculate the portfolio weights using the HPD model without parameter uncertainty, the HPD model with parameter uncertainty and the BMA specification. All strategies are evaluated using the BMA specification. The portfolio weights based on the BMA specification therefore give by definition the best performance.

Both figures show that the three different strategies give the same expected utility (and certainty equivalent) at short horizons. At longer horizons, there are differences. The strategy that only includes parameter uncertainty approximates the optimal strategy based on the BMA specification the best. However, the loss in certainty equivalence is still considerably: 15% when predictor variables are at their historical average and 20% when they are equal to their end-of-sample values. The losses when ignoring both parameter and model uncertainty are respectively 23% and 38%. These are quite big losses if we compare them to the magnitude of the certainty equivalences themselves.

The most remarkable result is that the certainty equivalent is smaller than 1 for all three specifications if the investment horizon is sufficiently large and if the variables are set equal to their end-of-sample values. The reason is that risk-averse investors dislike the extreme fat tail of the predictive distribution at the longest horizons. It is very important to recall that all asset returns are in real terms. Hence, a certainty equivalent of 1 means that an investor is indifferent between either following the strategy or getting 100% of its starting wealth at the end of the horizon in *real* terms. Certainty equivalents less than 1 therefore mean that an investor who follows the optimal strategy would rather pay a lot for an inflation-indexed bond than that he would follow this strategy. An important implication is that it would be very valuable for an investor who faces both parameter and model uncertainty to have the ability to invest in these inflation-indexed bonds.

7 Robustness tests

In this section, we consider several robustness checks. In the first section we recalculate the posterior distribution using a different prior distribution. In the second section we recalculate the posterior distribution at different points in time.

7.1 Different prior distribution

In this section, we check the prior robustness of our results by comparing the posterior distribution in section 4 with a posterior distribution based on a different prior. We use a different model prior and we use a different prior on the slope coefficients.

[18] Note that the certainly equivalents are not annualized.

Firstly, let us consider a different prior for the model prior probability. In the previous sections, we use the following prior

$$p(M_j|q) \propto q^{|M_j|}(1-q)^{n^2-|M_j|},$$

where q is set equal to 0.50. This choice implies that all models are given the same prior probability.

The choice of q is arbitrary and therefore we consider an alternative choice for q. Instead of setting q equal to a constant, we put a prior on q as in Ley and Steel (2009). We follow their recommendation and use the following prior for q

$$p(q) = Beta(1, b), \tag{23}$$

where we choose b in such a way that the prior mean of the model size is equal to $\frac{n^2}{2}$, i.e. half the size of the full model. This prior lets the data determine a value for q instead of fixing a value for q a priori. Ley and Steel (2009) show that this prior leads to a posterior that is very robust to the choice of the prior parameter, i.e. the choice of b. Note that this prior choice does not lead to a more complicated MCMC algorithm. The only aspect in the MCMC algorithm that changes is that we use a different marginal model prior in the model selection step of the MCMC algorithm, refer to section 3.4 for details on the MCMC algorithm.

Secondly, let us consider a different prior for the slope coefficients. In the main part of the paper, we use the following prior

$$p(\beta^{(j)}|\Sigma^{(j)}, M_j) = N(m^{(j)}, gV^{(j)}),$$

where $m^{(j)}$ and $V^{(j)}$ are given in section 3.2 and where g is set equal to T.

[Table 8 about here.]

[Table 9 about here.]

The choice for g is an important choice. This parameter determines on one hand how much information the prior contains within a model, but on the other hand it also determines the penalty factor for larger models. Therefore, we cannot choose g too large or too small. We put a prior on g to let the data determine a value for g instead of specifying it a priori. We choose the following proper rather flat inverse gamma prior for g (using the parameterization of Bauwens, Lubrano, and Richard (1999) for the inverse-gamma distribution)

$$p(g) = iG(0.01, 0.01). \tag{24}$$

Liang, Paulo, Molina, Clyde, and Berger (2008) also consider several priors on the g parameter in Zellner's g-prior in standard linear regression models.

In order to draw g we need to introduce an extra step in our MCMC algorithm. It is easy to show that we can draw g using the inverse gamma distribution in a Gibbs-step

$$p(g|Y, M_j, \beta^{(j)}) = iG((\beta^{(j)} - m^{(j)})' V^{(j)-1} (\beta^{(j)} - m^{(j)}) + 0.01, k^{(j)} + 0.01). \tag{25}$$

Since parameter g is common to all models, we condition on parameter g in the model selection step in the MCMC algorithm. The marginal likelihood in equation (16) now conditions on both $\Sigma^{(j)}$ and g, but the expression for the marginal likelihood itself stays exactly the same.

Table 8 shows the posterior probability that a variable is included in an equation for the alternative prior specification. If we compare this table to table 3, we see that most probabilities are very similar for both prior specifications. For example, the dividend-to-price ratio and the credit spread are still the most important direct predictors of stock returns. However, the posterior probability for the credit spread is lower and the posterior probability for the dividend-to-price ratio is higher than in table 3. Table 9 shows the posterior means and standard deviations for the alternative specification. If we compare the posterior moments with the posterior moments in table 5, we again see that these posterior moments are in general very similar. Since the posterior distribution of the parameters is not very sensitive to the exact prior choice, we conclude that our results are robust to the prior choice.

7.2 Posterior moments over time

[Figure 17 about here.]

[Figure 18 about here.]

[Figure 19 about here.]

[Figure 20 about here.]

In this section, we calculate the posterior distribution at different points in time in order to check the stability of our results. We do this as follows. We calculate the posterior distribution in the same way as in section 4 but by considering an expanding window of observations. In other words, we re-estimate our BMA specification on data-sets that contain more and more observations.

Figures 17 and 18 contain respectively the posterior probability that the dividend-to-price ratio is included in the equation for excess stock returns and the probability that the credit-spread is included in the equation for excess stock returns. The date on the horizontal axis indicates the last included observation in the subsample. Firstly, both figures show that the probabilities are quite stable after 1960. Furthermore, in both cases the probabilities decrease in the last 5 years. This is not surprising since the last 5 years were rather turbulent. Figures 19 and 20 show similar figures for the posterior means of these coefficients. These figures largely confirm the results above. Firstly, the posterior means stabilize around 1960. Furthermore, we see that the means for both coefficients are shrunk towards 0 in the last years, because the probability that the variable is included in the BMA specification decreases in these years.

All together we conclude that the posterior results are relatively stable over time as long as we estimate the specification on 40 years of data or more.

8 Conclusion

We develop an estimation framework that is able to include model uncertainty over long-term predictions using Bayesian Model Averaging (BMA). Our methodological framework allows us to consider model uncertainty in the prediction equations for both asset returns and predictor variables. The latter is very important when considering the impact of model uncertainty on the long-run predictability of stock returns. A variable, that is a good predictor of asset returns but cannot be predicted itself, is useless for a long-term investor.

Our results show that the credit spread and the dividend-to-price ratio are the most important predictors of stock returns at short horizons. However, at longer horizons, all variables are important for predicting long-horizon stock returns, either by directly predicting stock returns or by indirectly predicting the predictors of stock returns. If one would instead only base the inclusion of predictor variables on the prediction equation for stock returns, one would wrongly conclude that most predictor variables are not important for modeling stock returns. Furthermore, our results clearly show that model uncertainty is substantial, since the posterior probability mass is widely spread across many models.

The incorporation of model uncertainty has important implications for the term structure of risk. At long horizons, model uncertainty increases the variance of the predictive distribution of stock returns substantially, especially when predictor variables deviate significantly from their average values. The fact, that different models give significantly different predictions of stock returns, increases the variance of the predictive distribution. In extreme events such as the great depression and the subprime mortgage crisis, long-run stock returns can be significantly riskier than short-run stock returns. However, in relatively stable periods such as the 1960s, stock returns are safer in the long-run than in the short-run. Furthermore, we find that model uncertainty only has a negligible impact in the short-run.

The incorporation of model uncertainty also leads to significantly different asset allocations (up to 35%) at longer horizons compared to specifications that ignore model uncertainty. The reason is that the inclusion of model uncertainty leads to a predictive distribution of excess stock returns with a lower mean, higher standard deviation, more negative skewness and higher kurtosis, especially when predictor variables deviate from their own average. Also, despite the homoscedasticity of our models, the mean, variance, skewness and kurtosis of the predictive distribution of asset returns change substantially over time. Furthermore, we find that model uncertainty hardly has an impact on asset allocations for short-horizon investors. Finally, our results show that the certainty equivalent of the optimal buy-and-hold strategy for a long-horizon investor can be lower than 1 for sufficiently long investment horizons. The incorporation of model uncertainty makes the asset market too risky at long horizons in the eyes of an investor who recognizes that she does not know the true model and true parameters. This implies that such an investor would be willing to pay a lot for an inflation-indexed bond.

In this paper we consider the impact of model uncertainty in a setting where models only differ in the variables that are included and where we estimate the specification on a relatively long data-set. Of course, we could also extend the model space by considering non-linear models, stochastic volatility models etcetera. We can also look at the impact of model uncertainty if we use shorter data sets. It is likely that the impact of model uncertainty is larger in such settings, because we would consider more models and less information (data). The results in this paper can therefore be interpreted as a lowerbound on the true impact of model uncertainty.

A Posterior distribution and MCMC algorithm

Firstly, we derive the posterior distributions that are used in the paper. The notation is consistent with the notation introduced in the main paper. Most terms are introduced in the main text.

The likelihood function for model M_j in equation (5) is (all three expressions are equivalent)

$$\begin{aligned}
P(Y|\beta^{(j)}, \Sigma^{(j)}, M_j) &= (2\pi)^{-\frac{Tn}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - Z^{(j)}\beta^{(j)})'(\Sigma^{(j)} \otimes I_T)^{-1}(y - Z^{(j)}\beta^{(j)})\right\} \\
P(Y|\beta^{(j)}, \Sigma^{(j)}, M_j) &= (2\pi)^{-\frac{Tn}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}} \exp\{-\frac{1}{2}s^{(j)}\} \\
&\quad \exp\left\{-\frac{1}{2}(\beta^{(j)} - \hat{\beta}^{(j)})'Z^{(j)\prime}(\Sigma^{(j)} \otimes I_T)^{-1}Z^{(j)}(\beta^{(j)} - \hat{\beta}^{(j)})\right\} \\
P(Y|\beta^{(j)}, \Sigma^{(j)}, M_j) &= (2\pi)^{-\frac{Tn}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}tr(\Sigma^{(j)-1}(Y - WB_c)'(Y - WB_c))\right\},
\end{aligned}$$

where

$$s^{(j)} = y'(\Sigma^{(j)} \otimes I_T)^{-1}y - \hat{\beta}^{(j)\prime}Z^{(j)\prime}(\Sigma^{(j)} \otimes I_T)^{-1}Z^{(j)}\hat{\beta}^{(j)}.$$

Please refer to Bauwens, Lubrano, and Richard (1999) for details. The prior $p(\beta^{(j)}, \Sigma^{(j)}|M_j)$ is proportional to (both expressions are equivalent)

$$\begin{aligned}
p(\beta^{(j)}, \Sigma^{(j)}|M_j) &\propto |\Sigma^{(j)}|^{-\frac{n+1}{2}}g^{-\frac{k^{(j)}}{2}}|V^{(j)}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\beta^{(j)} - m^{(j)})'\frac{1}{g}V^{(j)-1}(\beta^{(j)} - m^{(j)})\right) \\
p(\beta^{(j)}, \Sigma^{(j)}|M_j) &\propto |\Sigma^{(j)}|^{-\frac{n+1}{2}}g^{-\frac{k^{(j)}}{2}}|V^{(j)}|^{-\frac{1}{2}} \\
&\quad \exp\left(-\frac{1}{2}tr(\Sigma^{(j)-1}(B_f^{(j)\prime} - M_f^{(j)\prime})'\frac{1}{g}X'X(B_f^{(j)\prime} - M_f^{(j)\prime}))\right).
\end{aligned}$$

The posterior is proportional to the product of the prior and the likelihood. We do not know the analytical properties of the joint posterior $p(\beta^{(j)}, \Sigma^{(j)}|M_j, Y)$, since its integrating constant is unknown in general. Therefore, we derive expressions for the conditional posteriors.

The conditional posterior $p(\beta^{(j)}|M_j, \Sigma^{(j)}, Y)$ is analytically known. Use the second expression for the likelihood and the first expression for the prior to get

$$\begin{aligned}
p(\beta^{(j)}|M_j, Y, \Sigma^{(j)}) &\propto \exp\left\{-\frac{1}{2}(\beta^{(j)} - \hat{\beta}^{(j)})'V^{(j)-1}(\beta^{(j)} - \hat{\beta}^{(j)})\right\} \\
&\quad \exp\left\{-\frac{1}{2}(\beta^{(j)} - m^{(j)})'\frac{1}{g}V^{(j)-1}(\beta^{(j)} - m^{(j)})\right\}. \qquad (26)
\end{aligned}$$

Note the following:

$$
\begin{aligned}
C &= (\beta^{(j)} - \hat{\beta}^{(j)})' V^{(j)-1} (\beta^{(j)} - \hat{\beta}^{(j)}) + (\beta^{(j)} - m^{(j)})' \frac{1}{g} V^{(j)-1} (\beta^{(j)} - m^{(j)}) \\
C &= \beta^{(j)'} \left(\frac{1+g}{g} V^{(j)-1} \right) \beta^{(j)} - 2\beta^{(j)'} \left(\frac{1+g}{g} V^{(j)-1} \right) \left(\frac{1+g}{g} V^{(j)-1} \right)^{-1} (V^{(j)-1} \hat{\beta}(j) \\
&+ \frac{1}{g} V^{(j)-1} m^{(j)}) + \hat{\beta}^{(j)'} V^{(j)-1} \hat{\beta}^{(j)} + m^{(j)'} \frac{1}{g} V^{(j)-1} m^{(j)} \\
C &= \beta^{(j)'} M^{*(j)-1} \beta^{(j)} - 2\beta^{(j)'} M^{*(j)-1} \beta^{*(j)} + \hat{\beta}^{(j)'} V^{(j)-1} \hat{\beta}^{(j)} + m^{(j)'} \frac{1}{g} V^{(j)-1} m^{(j)} \\
C &= (\beta^{(j)'} - \beta^{*(j)})' M^{*(j)-1} (\beta^{(j)'} - \beta^{*(j)}) + \hat{\beta}^{(j)'} V^{(j)-1} \hat{\beta}^{(j)} + m^{(j)'} \frac{1}{g} V^{(j)-1} m^{(j)} - \beta^{*(j)'} M^{*(j)-1} \beta^{*(j)},
\end{aligned}
$$

where $M^{*(j)}$ is defined in equation (14) and $\beta^{*(j)}$ is

$$
\begin{aligned}
\beta^{*(j)} &= \left(\frac{1+g}{g} V^{(j)-1} \right)^{-1} (V^{(j)-1} \hat{\beta}(j) + \frac{1}{g} V^{(j)-1} m^{(j)}) \\
\beta^{*(j)} &= \frac{g}{1+g} \hat{\beta} + \frac{1}{1+g} m^{(j)}
\end{aligned}
$$

and hence equals equation (14). These results imply that

$$
p(\beta^{(j)} | M_j, Y, \Sigma^{(j)}) \propto \exp \left\{ -\frac{1}{2} (\beta^{(j)} - \beta^{*(j)})' M^{*(j)-1} (\beta^{(j)} - \beta^{*(j)}) \right\}.
$$

and therefore

$$
p(\beta^{(j)} | M_j, \Sigma^{(j)}, Y) = N(\beta^{*(j)}, M^{*(j)}).
$$

The conditional posterior $p(\Sigma^{(j)} | M_j, \beta^{(j)}, Y)$ is not analytically known. By combining the last expression for the likelihood and the last expression for the prior we can easily show that it is proportional to

$$
\begin{aligned}
p(\Sigma^{(j)} | M_j, \beta^{(j)}, Y) \propto |\Sigma^{(j)}|^{-\frac{T+n+1}{2}} |V^{(j)}|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} tr \left(\Sigma^{(j)-1} [(Y - W B_c)'(Y - W B_c)] \right) \right) \\
\exp \left(-\frac{1}{2} tr \left(\Sigma^{(j)-1} (B_f^{(j)'} - M_f^{(j)'})' \frac{1}{g} X'X (B_f^{(j)'} - M_f^{(j)'}) \right) \right)
\end{aligned}
$$

and hence equals equation (14).

Finally, we need to obtain an expression for the conditional marginal likelihood $p(Y | M_j, \Sigma^{(j)})$. By combining the second expression for the likelihood and the first expression for the prior we get

$$
\begin{aligned}
p(Y|M_j,\Sigma^{(j)}) &= \int p(Y|M_j,\beta^{(j)},\Sigma^{(j)})p(\beta^{(j)}|M_j,\Sigma^{(j)})d\beta^{(j)} \\
p(Y|M_j,\Sigma^{(j)}) &= \int (2\pi)^{-\frac{Tn}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}}\exp\left\{-\frac{1}{2}s^{(j)}\right\} \\
&\quad \exp\left\{-\frac{1}{2}(\beta^{(j)}-\hat{\beta}^{(j)})'Z^{(j)\prime}(\Sigma^{(j)}\otimes I_T)^{-1}Z^{(j)}(\beta^{(j)}-\hat{\beta}^{(j)})\right\} \\
&\quad (2\pi)^{-\frac{k^{(j)}}{2}}g^{-\frac{k^{(j)}}{2}}|V^{(j)}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\beta^{(j)}-m^{(j)})'\frac{1}{g}V^{(j)-1}(\beta^{(j)}-m^{(j)})\right\}d\beta^{(j)} \\
p(Y|M_j,\Sigma^{(j)}) &= (2\pi)^{-\frac{Tn+k^{(j)}}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}}g^{-\frac{k^{(j)}}{2}}|V^{(j)}|^{-\frac{1}{2}} \\
&\quad \exp\left\{-\frac{1}{2}\left(s^{(j)}+\hat{\beta}^{(j)\prime}V^{(j)-1}\hat{\beta}^{(j)}+m^{(j)\prime}\frac{1}{g}V^{(j)-1}m^{(j)}-\beta^{*(j)\prime}M^{*(j)-1}\beta^{*(j)}\right)\right\} \\
&\quad \int \exp\left\{-\frac{1}{2}(\beta^{(j)\prime}-\beta^{*(j)})'M^{*(j)-1}(\beta^{(j)\prime}-\beta^{*(j)})\right\}d\beta^{(j)} \\
p(Y|M_j,\Sigma^{(j)}) &= (2\pi)^{-\frac{Tn+k^{(j)}}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}}g^{-\frac{k^{(j)}}{2}}|V^{(j)}|^{-\frac{1}{2}} \\
&\quad \exp\left\{-\frac{1}{2}\left(s^{(j)}+\hat{\beta}^{(j)\prime}V^{(j)-1}\hat{\beta}^{(j)}+m^{(j)\prime}\frac{1}{g}V^{(j)-1}m^{(j)}-\beta^{*(j)\prime}M^{*(j)-1}\beta^{*(j)}\right)\right\} \\
&\quad (2\pi)^{\frac{k^{(j)}}{2}}|M^{*(j)}|^{\frac{1}{2}}
\end{aligned}
$$

$$
\begin{aligned}
p(Y|M_j,\Sigma^{(j)}) &= (2\pi)^{-\frac{Tn}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}}g^{-\frac{k^{(j)}}{2}}|V^{(j)}|^{-\frac{1}{2}}|M^{*(j)}|^{\frac{1}{2}} \\
&\quad \exp\left\{-\frac{1}{2}\left(y'(\Sigma^{(j)}\otimes I_T)^{-1}y+m^{(j)\prime}\frac{1}{g}V^{(j)-1}m^{(j)}-\beta^{*(j)\prime}M^{*(j)-1}\beta^{*(j)}\right)\right\} \\
p(Y|M_j,\Sigma^{(j)}) &= (2\pi)^{-\frac{Tn}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}}(1+g)^{-\frac{k^{(j)}}{2}} \\
&\quad \exp\left\{-\frac{1}{2}\left(y'(\Sigma^{(j)}\otimes I_T)^{-1}y+m^{(j)\prime}\frac{1}{g}V^{(j)-1}m^{(j)}-\beta^{*(j)\prime}M^{*(j)-1}\beta^{*(j)}\right)\right\},
\end{aligned}
$$

where the third equality follows from the last expression for C. The final expression for $p(Y|M_j,\Sigma^{(j)})$ is equal to expression (16).

Secondly, we explain the MCMC algorithm we use to estimate the posterior distributions. In every iteration we draw model M_l, coefficient vector $\beta^{(l)}$ and covariance matrix $\Sigma^{(l)}$. Current values are indexed by j. The discussion in this section is partially based on Godsill (2001) and Troughton and Godsill (1997) who explain how to apply model selection to univariate time-series. Godsill (2001), Troughton and Godsill (1997) and Han and Carlin (2001) show that the method works well when the common parameter (in our case $\Sigma^{(j)}$) has a common meaning across models. We generalize their method to model uncertainty over systems of equations, i.e. restricted VAR(1) models.

In the first step, we choose to draw a new model M_l and a new coefficient vector $\beta^{(l)}$. As in Godsill (2001), we draw $\Sigma^{(l)}$ in a second step and therefore condition on $\Sigma^{(j)}$ in the first step. Since the dimension of the parameter space of the new model is not necessarily equal to the dimension of the parameter space of the current model, we cannot use standard Metropolis-Hastings techniques. We have to rely on the generalization in Green (1995) that allows moves between parameter spaces of different dimensions.

Let the probability of proposing model M_l^* when we are currently in model M_j be $q(M_j \to M_l^*)$ and let $q(\beta^{(l)}|M_l^*, \beta^{(j)}, \Sigma^{(j)})$ be the proposal density for coefficient vector $\beta^{(l)}$ when we are in model M_l^*. Results in Green (1995) imply that the acceptance probability is

$$\alpha = \min\left\{1, \frac{p(M_l^*, \beta^{(l)}|Y, \Sigma^{(j)})q(M_l^* \to M_j)q(\beta^{(j)}|Y, M_j, \beta^{(l)}, \Sigma^{(j)})}{p(M_j, \beta^{(j)}|\Sigma^{(j)})q(M_j \to M_l^*)q(\beta^{(l)}|M_l^*, \beta^{(j)}, \Sigma^{(j)})}\right\}$$

We propose a new model M_l^* by randomly selecting a model with $k^{(j)}+1$ and $k^{(j)}-1$ variables and propose $\beta^{(l)}$ using the conditional posterior distribution in equation (13): $p(\beta^{(l)}|M_l^*, \Sigma^{(j)}, Y)$. The acceptance probability simplifies to

$$\alpha = \min\left\{1, \frac{p(M_l^*, \beta^{(l)}|Y, \Sigma^{(j)})p(\beta^{(j)}|Y, M_j, \Sigma^{(j)})}{p(M_j, \beta^{(j)}|Y, \Sigma^{(j)})p(\beta^{(l)}|Y, M_l^*, \Sigma^{(j)})}\right\}$$

Obviously, the following holds

$$p(M_l^*, \beta^{(l)}|Y, \Sigma^{(j)}) = p(M_l^*|Y, \Sigma^{(j)})p(\beta^{(l)}|Y, M_l^*, \Sigma^{(j)})$$

and therefore we can simplify the acceptance probability to

$$\alpha = \min\left\{1, \frac{p(M_l^*|Y, \Sigma^{(j)})}{p(M_j|Y, \Sigma^{(j)})}\right\}.$$

Finally, since

$$p(M_l^*|Y, \Sigma^{(j)}) = \frac{p(M_l^*)p(Y|M_l^*, \Sigma^{(j)})}{p(Y|\Sigma^{(j)})}$$

the acceptance probability becomes

$$\alpha = \min\left\{1, \frac{p(M_l^*)p(Y|M_l^*, \Sigma^{(j)})}{p(M_j)p(Y|M_j, \Sigma^{(j)})}\right\},$$

which is equivalent to the expression in equation (17).

Note that the acceptance probability does not depend on the value of $\beta^{(l)}$. Therefore, in practice we only draw $\beta^{(l)}$ when model M_l^* is accepted. In case model M_l^* is rejected and therefore $M_l = M_j$, we simply update $\beta^{(l)}$ using the conditional posterior distribution in equation (13): $p(\beta^{(l)}|M_l = M_j, \Sigma^{(j)}, Y)$.

Finally, we consider the updating step for $\Sigma^{(l)}$. We use a standard Metropolis-Hastings algorithm. Suppose we draw $\Sigma^{(l)*}$ according to proposal density $q(\Sigma^{(l)*}|Y, M_l)$ and let $h(\Sigma^{(l)*}|Y, M_l)$

be the kernel of the target density. The acceptance probability is

$$\alpha = \min \left\{ 1, \frac{h(\Sigma^{(l)*}|Y, M_l)q(\Sigma^{(j)}|Y, M_l)}{h(\Sigma^{(j)}|Y, M_l)q(\Sigma^{(l)*}|Y, M_l)} \right\}.$$

The kernel of the target distribution is given in equation (14). As a proposal density, we choose an iWishart$\left(E^{(l)\prime}E^{(l)} + \frac{1}{g}H^{(l)}, T + n + 1\right)$.[19] The acceptance probability becomes

$$\alpha = \min \left\{ 1, \frac{|Z^{(l)\prime}(\Sigma^{(l)*} \otimes I_T)^{-1}Z^{(l)}|^{\frac{1}{2}}|\Sigma^{(l)*}|^{\frac{n+1}{2}}}{|Z^{(l)\prime}(\Sigma^{(j)} \otimes I_T)^{-1}Z^{(l)}|^{\frac{1}{2}}|\Sigma^{(j)}|^{\frac{n+1}{2}}} \right\},$$

which is equivalent to the expression in equation (18).

In the empirical section, most results are based on 500,000 retained draws after an initialization phase of 100,000 draws. Increasing the burn-in phase or the number of simulations does not significantly change results. Visual inspection of the posterior draws suggests that the estimates converge.

[19]We use the parameterization of Bauwens, Lubrano, and Richard (1999) for the inverted Wishart distribution.

References

Andersson, Michael, and Sune Karlsson, 2008, Bayesian forecast combination for var models, Chapter in *Bayesian Econometrics*, page 501-524.

Avramov, Doran, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *The Journal of Finance* 55, 225–264.

Barsky, Robert, and Bradford De Long, 1993, Why does the stock market fluctuate?, *Quarterly Journal of Economics* 2, 291–311.

Bauwens, Luc, Michel Lubrano, and Jean-François Richard, 1999, *Bayesian Inference in Dynamic Econometric Models* (Oxford University Press).

Brandt, Michael, 2010, Portfolio choice problems, *Handbook of Financial Econometrics, Volume 1: Tools and Techniques* pp. 269–336.

Campbell, John, Yeung Chan, and Luis Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of Financial Economics* 67, 41–80.

Campbell, John, and Samuel Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *The Review of Financial Studies* 21, 1509–1531.

Campbell, John, and Luis Viceira, 2002, *Strategic Asset Allocation* (Oxford University Press).

——— , 2005, The term structure of the risk-return trade-off, *Financial Analysts Journal* 61, 34–44.

Cremers, Martijn, 2002, Stock return predictability: A bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.

Fernandez, Carmen, Eduardo Ley, and Mark Steel, 2001, Benchmark priors for bayesian model averaging, *Journal of Econometrics* 100, 381–427.

Fernandez-Villaverde, Jesus, and Juan Francisco Rubio-Ramirez, 2004, Comparing dynamic equilibrium models to data: a bayesian approach., *Journal of Econometrics* 123, 153–187.

Gelfand, Alan, and Dipak Dey, 1994, Bayesian model choice: asymptotics and exact calculations, *Journal of the Royal Statistical Society. Series B (Methodological)* 56, 501–514.

George, Edward, Shawn Ni, and Dongchu Sun, 2008, Bayesian stochastic search for var model restrictions, *Journal of Econometrics* 142, 553–580.

Godsill, 2001, On the relationship between markov chain monte carlo methods for model uncertainty, *Journal of Computational and Graphical Statistics* 10, 230–248.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.

Green, Peter, 1995, Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika* 82, 711–732.

Han, Cong, and Bradley Carlin, 2001, Markov chain monte carlo methods for computing bayes factors: A comparative review, *Journal of the American Statistical Assocation* 96, 1122–1132.

Hjort, Nils, and Gerda Claeskens, 2003, Frequentist model average estimators, *Journal of the American Statistical Association* 98, 879–899.

Hoeting, J.A., D. Madigan, A. Raftery, and C. Volinsky, 1999, Bayesian Model Averaging: A Tutorial, *Statistical Science* 14, 382–417.

Hoevenaars, Roy, Roderick Molenaar, Peter Schotman, and Tom Steenkamp, 2007, Strategic asset allocation for long-term investors: Parameter uncertainty and prior information, *Working paper.*

Kandel, Shmuel, and Robert Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *Journal of Finance* 51, 385.

Ley, Eduardo, and Mark Steel, 2007, Jointness in bayesian variable selection with applications to growth regression, *Journal of Macroeconomics* 29, 476–493.

——— , 2009, On the effect of prior assumptions in bayesian model averaging with applications to growth regression, *Journal of Applied Econometrics* Forthcoming.

Liang, Feng, Rui Paulo, German Molina, merlise Clyde, and Jim Berger, 2008, Mixtures of g priors for bayesian variable selection, *Journal of the American Statistical Association* 103, 410–423.

Marcellino, Massimiliano, James Stock, and Mark Watson, 2006, A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series, *Journal of Econometrics* 135, 499–526.

Merton, Robert, 1969, Lifetime portfolio selection under uncertainty: the continuous time case, *Review of Economics and Statistics* 51, 247–257.

——— , 1971, Optimal consumption and portfolio rules in a continous-time model, *Journal of Economic Theory* 3, 373–413.

Pastor, Lubos, and Robert Stambaugh, 2010, Are stocks really less volatile in the long run?, *Working paper.*

Sims, Christopher, and Harald Uhlig, 1991, Understanding unit rooters: A helicopter view, *Econometrica* 59, 1591–1599.

Troughton, Paul, and Simon Godsill, 1997, A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves, *Technical Report CUED/F-INFENG/TR.304* Cambridge University Engineering Department.

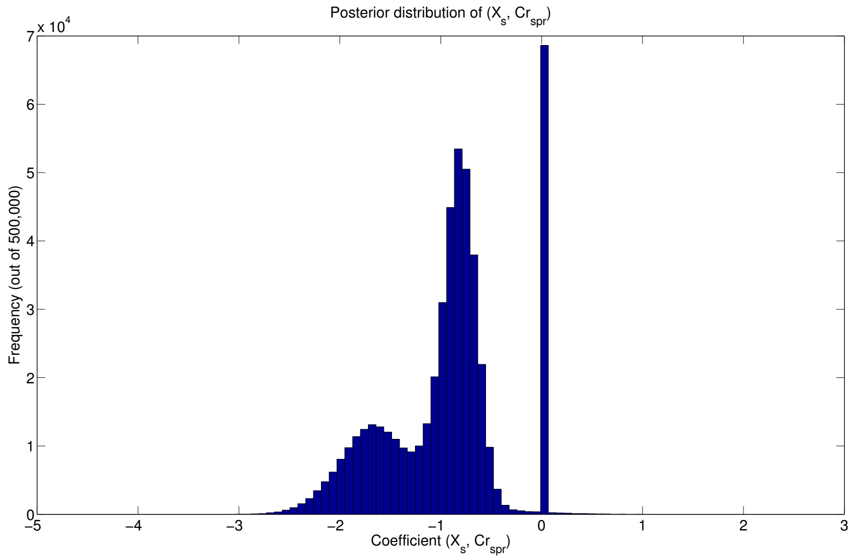
Wachter, Jessica, and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market?, *Journal of Econometrics* 148, 162–178.

Wright, Jonathan, 2008, Bayesian model averaging and exchange rate forecasts, *Journal of Econometrics* 146, 329–341.

List of Figures

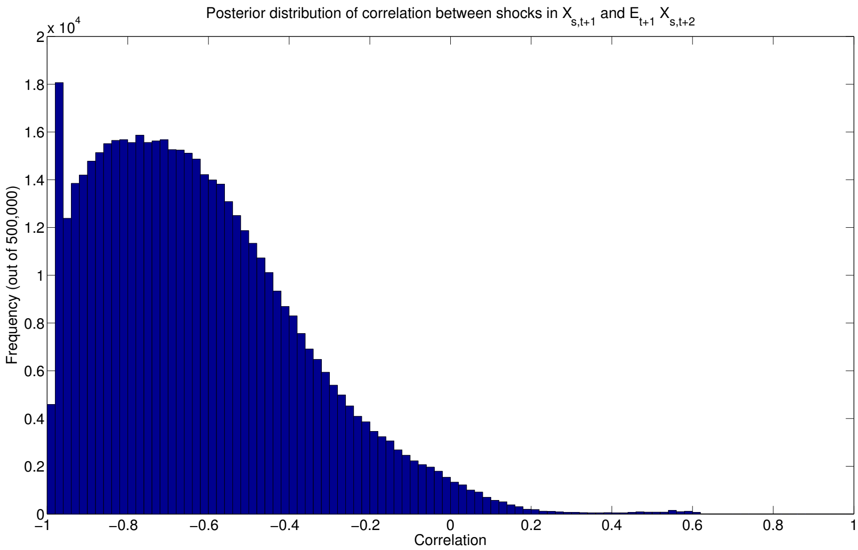
1 Posterior distribution of the coefficient on the credit spread in the equation for excess stock returns . . . 41
2 Posterior distribution of the correlation between a shock to current excess stock returns and a shock to future expected excess stock returns. . . . 42
3 Term structure of risk for the excess log stock return using the BMA model. Predictor variables are equal to their historical average. . . . 43
4 Term structure of risk for the excess log stock return using the BMA model. Predictor variables are equal to the values at the end of the sample. . . . 44
5 Time-series of annualized total volatility at different horizons. . . . 45
6 Term structure of risk for the excess log stock return using the highest posterior probability model. Predictor variables are equal to their historical average. . . . 46
7 Term structure of risk for the excess log stock return using the highest posterior probability model. Predictor variables are equal to the values at the end of the sample. . . . 47
8 Term structure of risk for the excess log stock return using the full model. Predictor variables are equal to the values at their historical average. . . . 48
9 Term structure of risk for the excess log stock return using the full model. Predictor variables are equal to the values at the end of the sample. . . . 49
10 Higher moments stock returns. Predictor variables are equal to their historical average. . . . 50
11 Higher moments of stock returns. Predictor variables are equal to the values at the end of the sample. . . . 51
12 Time-series of four moments of predictive distribution. . . . 52
13 Portfolio weights versus horizon for different estimation techniques. Predictor variables are equal to their historical average. . . . 53
14 Portfolio weights versus horizon for different estimation techniques. Predictor variables are equal to the values at the end of the sample. . . . 54
15 Plot of Certainty Equivalent versus horizon for different estimation techniques. Predictor variables are equal to their historical average. . . . 55
16 Plot of Certainty Equivalent versus horizon for different estimation techniques. Predictor variables are equal to the values at the end of the sample. . . . 56
17 Time-series of model probability (X_s, DP) . . . 57
18 Time-series of model probability (X_s, Cr_{spr}) . . . 58
19 Time-series of posterior mean of coefficient (X_s, DP) . . . 59
20 Time-series of posterior mean of coefficient (X_s, Cr_{spr}) . . . 60

Figure 1: **Posterior distribution of the coefficient on the credit spread in the equation for excess stock returns**



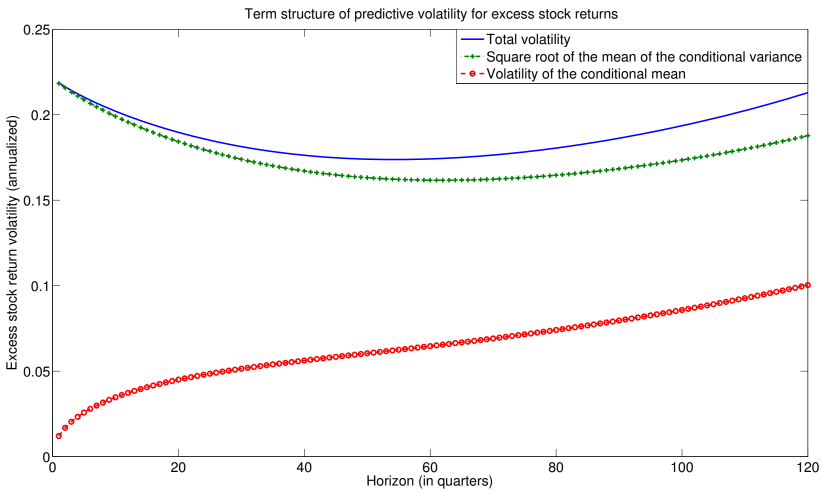
This figure shows the posterior distribution of the coefficient on the credit spread in the equation for excess stock returns, i.e. (x_s, Cr_{spr}). It is based on 500,000 retained draws from the posterior distribution of the weighted BMA model.

Figure 2: **Posterior distribution of the correlation between a shock to current excess stock returns and a shock to future expected excess stock return.**



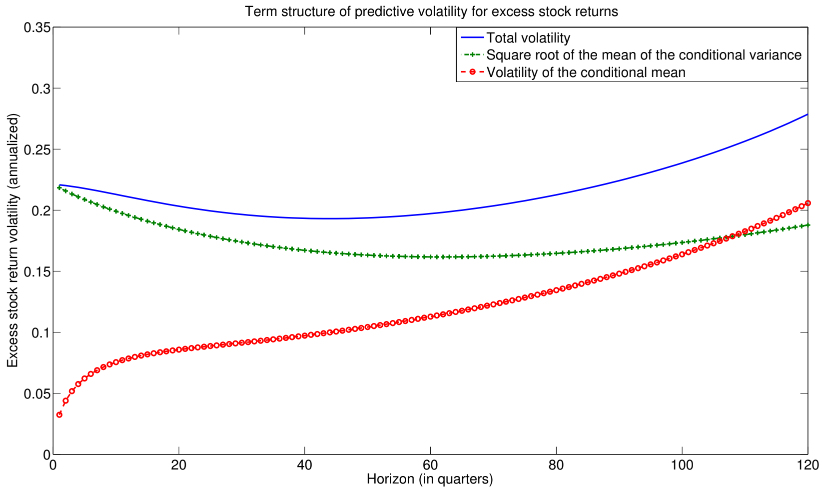
This figure shows the posterior distribution of the correlation between a shock in current excess stock returns and a shock in the expectation of future excess stock returns. The correlation is a measure for the mean reversion in stock returns. It is based on 500,000 retained draws from the posterior distribution of the weighted BMA model.

Figure 3: **Term structure of risk for the excess log stock return using the BMA model. Predictor variables are equal to their historical average.**



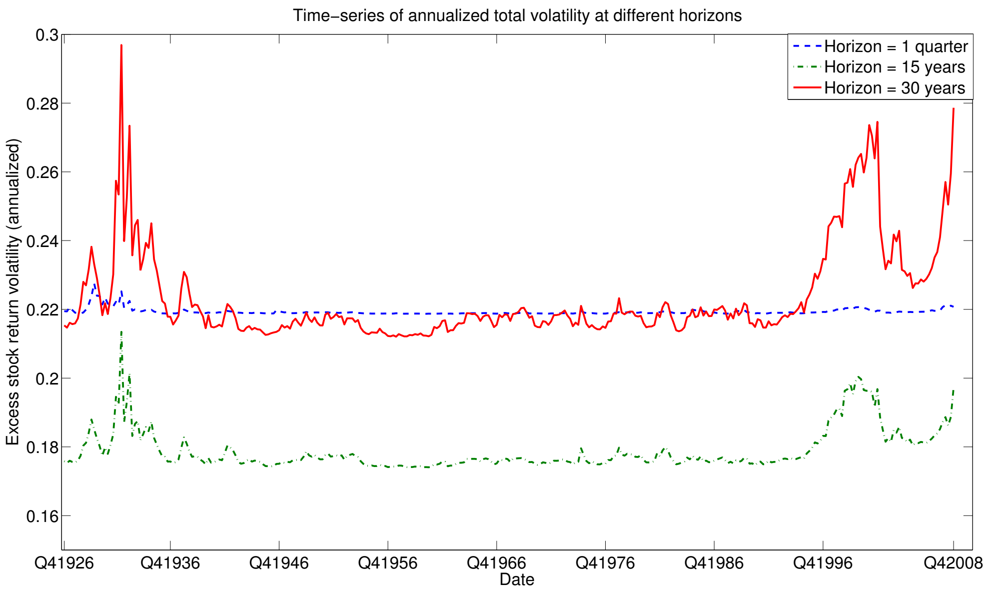
This figure shows the term structure of risk for excess log stock returns. It is based on 100,000 retained draws from the posterior distribution of the weighted BMA model. The predictor variables are equal to the sample means. Note that the total variance is equal to the mean of the conditional variance plus the variance of the conditional mean and note that all values are annualized.

Figure 4: **Term structure of risk for the excess log stock return using the BMA model. Predictor variables are equal to the values at the end of the sample.**



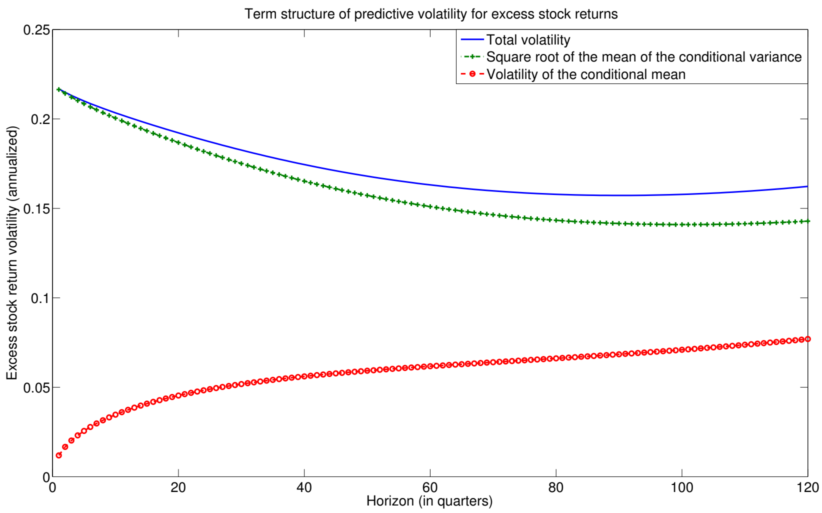
This figure shows the term structure of risk for excess log stock returns. It is based on 100,000 retained draws from the posterior distribution of the weighted BMA model. The predictor variables are equal to values in the fourth quarter of 2008. Note that the total variance is equal to the mean of the conditional variance plus the variance of the conditional mean and note that all values are annualized.

Figure 5: **Time-series of annualized total volatility at different horizons.**



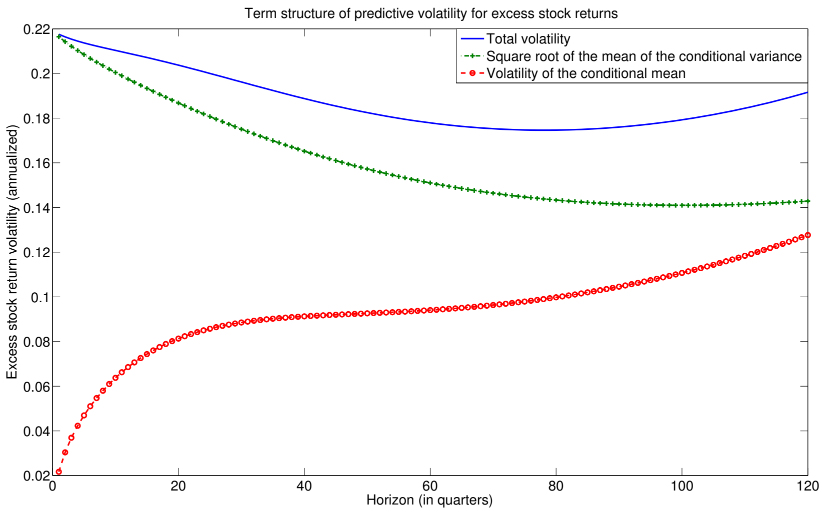
This figure shows a time-series of the square root of the annualized total variance of excess log stock returns for three different investment horizons. It is based on 50,000 retained draws from the posterior distribution of the weighted BMA model.

Figure 6: **Term structure of risk for the excess log stock return using the highest posterior probability model. Predictor variables are equal to their historical average.**



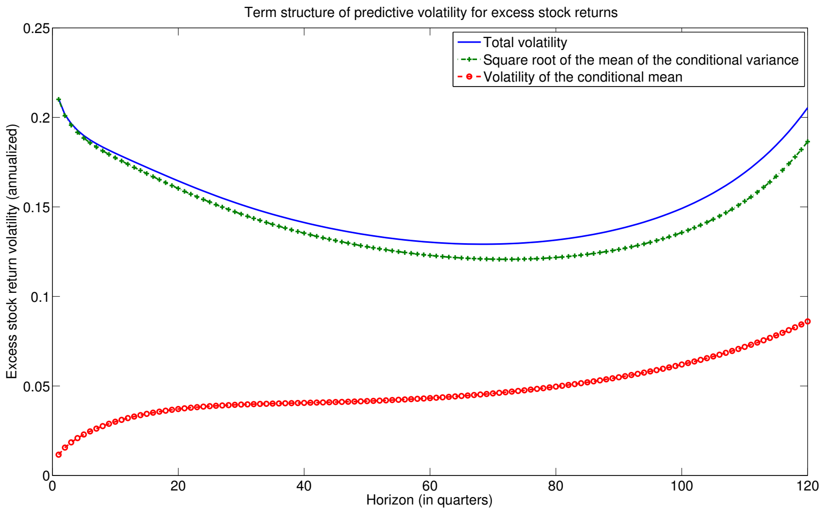
This figure shows the term structure of risk for excess log stock returns. It is based on 100,000 retained draws from the posterior distribution of the SUR model that receives the highest posterior probability. The predictor variables are equal to the sample means. Note that the total variance is equal to the mean of the conditional variance plus the variance of the conditional mean and note that all values are annualized.

Figure 7: **Term structure of risk for the excess log stock return using the highest posterior probability model. Predictor variables are equal to the values at the end of the sample.**



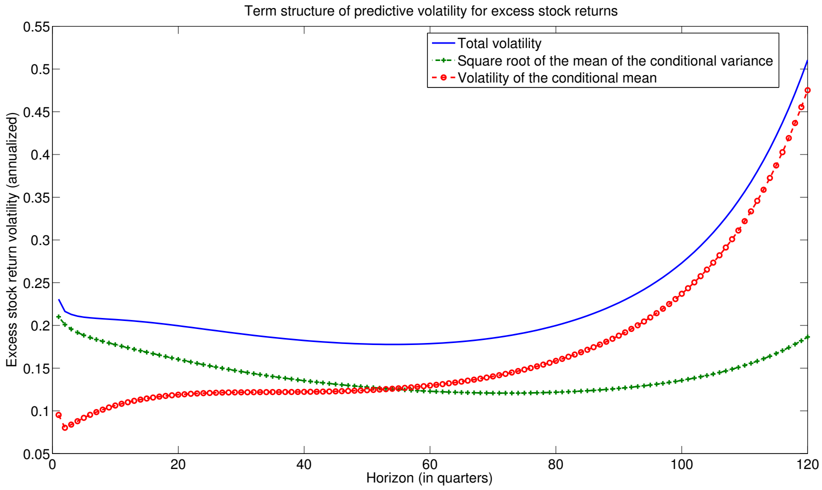
This figure shows the term structure of risk for excess log stock returns. It is based on 100,000 retained draws from the posterior distribution of the SUR model that receives the highest posterior probability. The predictor variables are equal to values in the fourth quarter of 2008. Note that the total variance is equal to the mean of the conditional variance plus the variance of the conditional mean and note that all values are annualized.

Figure 8: **Term structure of risk for the excess log stock return using the full model. Predictor variables are equal to the values at their historical average.**



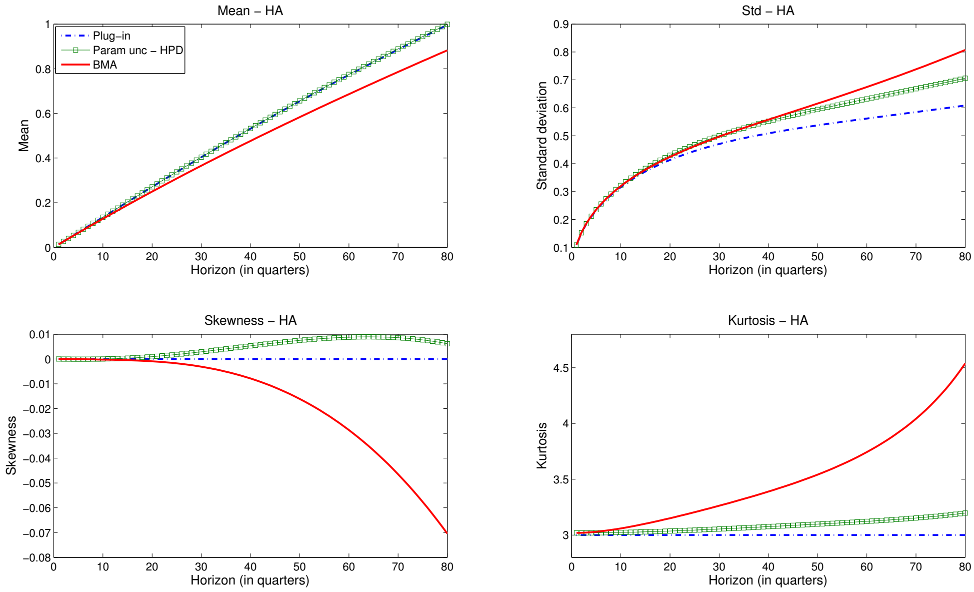
This figure shows the term structure of risk for excess log stock returns. It is based on 100,000 retained draws from the posterior distribution of the full model. The predictor variables are equal to their historical average. Note that the total variance is equal to the mean of the conditional variance plus the variance of the conditional mean and note that all values are annualized.

Figure 9: **Term structure of risk for the excess log stock return using the full model. Predictor variables are equal to the values at the end of the sample.**



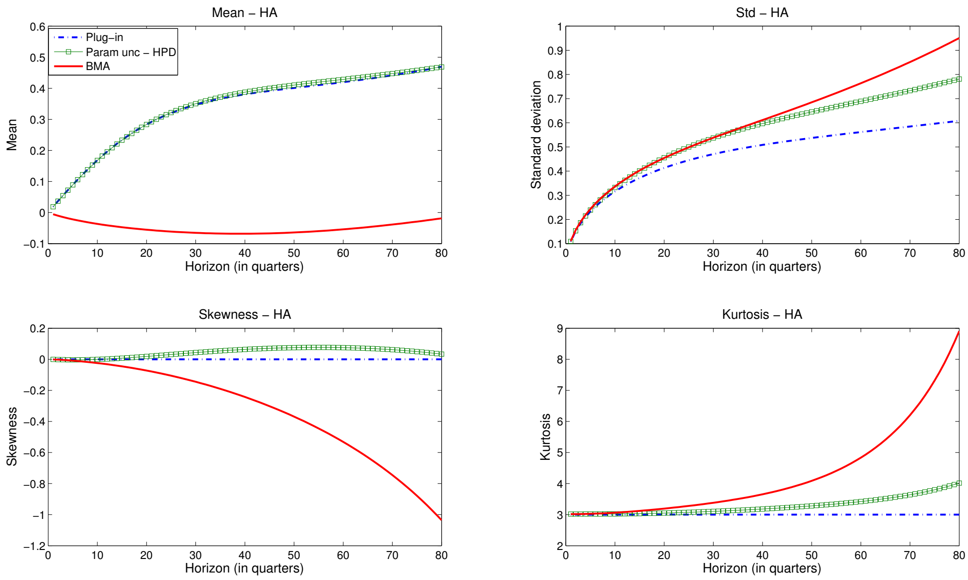
This figure shows the term structure of risk for excess log stock returns. It is based on 100,000 retained draws from the posterior distribution of the weighted BMA model. The predictor variables are equal to values in the fourth quarter of 2008. Note that the total variance is equal to the mean of the conditional variance plus the variance of the conditional mean and note that all values are annualized.

Figure 10: **Higher moments stock returns. Predictor variables are equal to their historical average.**



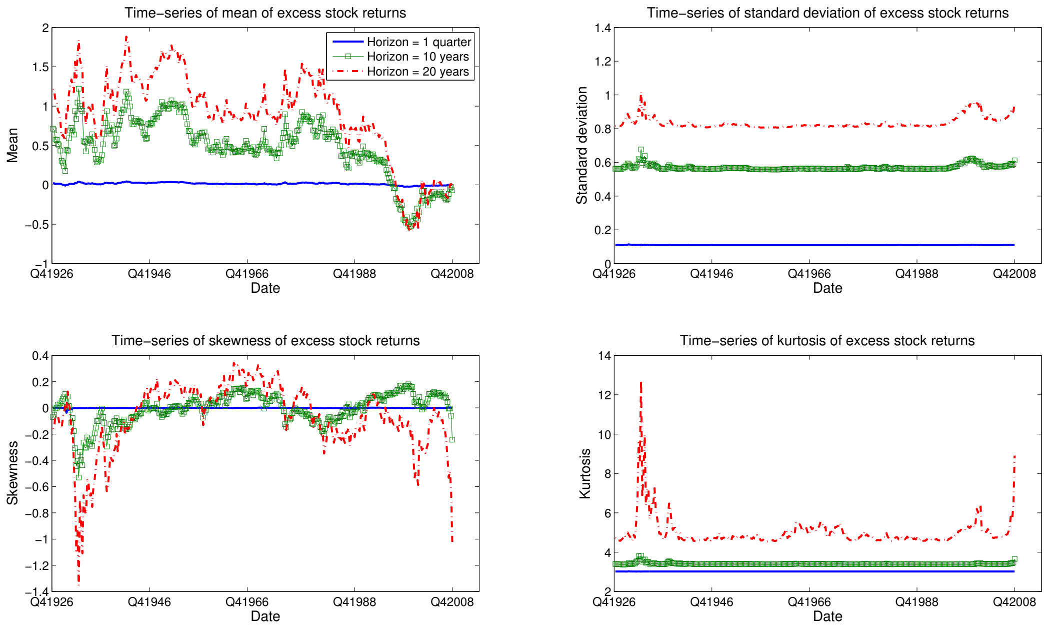
This figure plots the first four moments (mean, standard deviation, skewness and kurtosis) of the predictive distribution of excess log stock returns using either the highest posterior probability (HPD) model without parameter uncertainty (plug-in), the HPD with parameter uncertainty (param unc - HPD) or the BMA model with parameter and model uncertainty (BMA). Results are based on 100,000 draws from the predictive distribution. Predictor variables are set equal to their sample means. Note that the values in the graphs are not annualized.

Figure 11: **Higher moments of stock returns. Predictor variables are equal to the values at the end of the sample.**



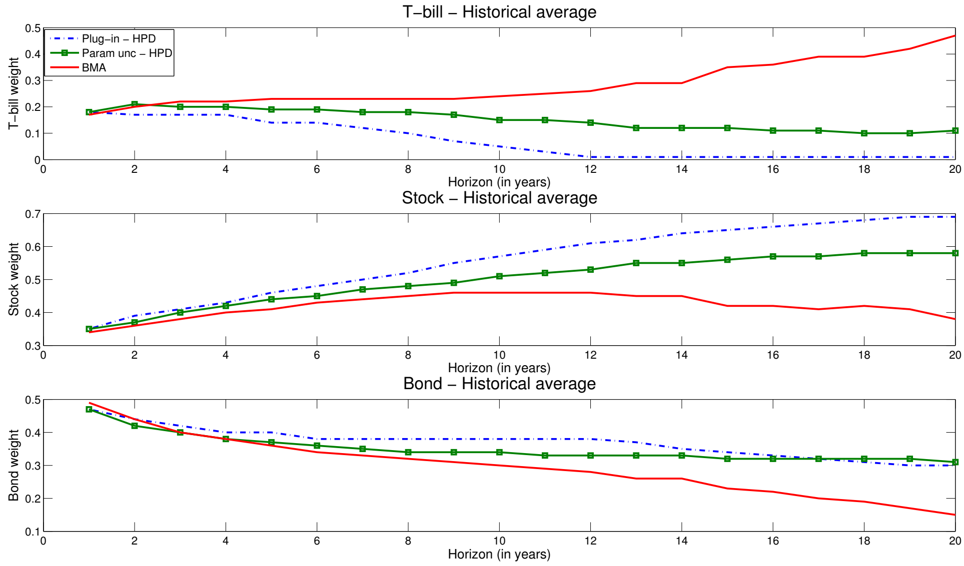
This figure plots the first four moments (mean, standard deviation, skewness and kurtosis) of the predictive distribution of excess log stock returns using either the highest posterior probability (HPD) model without parameter uncertainty (plug-in), the HPD with parameter uncertainty (param unc - HPD) and the BMA model with parameter and model uncertainty (BMA). Results are based on 100,000 draws from the predictive distribution. Predictor variables are equal to their end-of-sample values. Note that the values in the graphs are not annualized.

Figure 12: **Time-series of four moments of predictive distribution.**



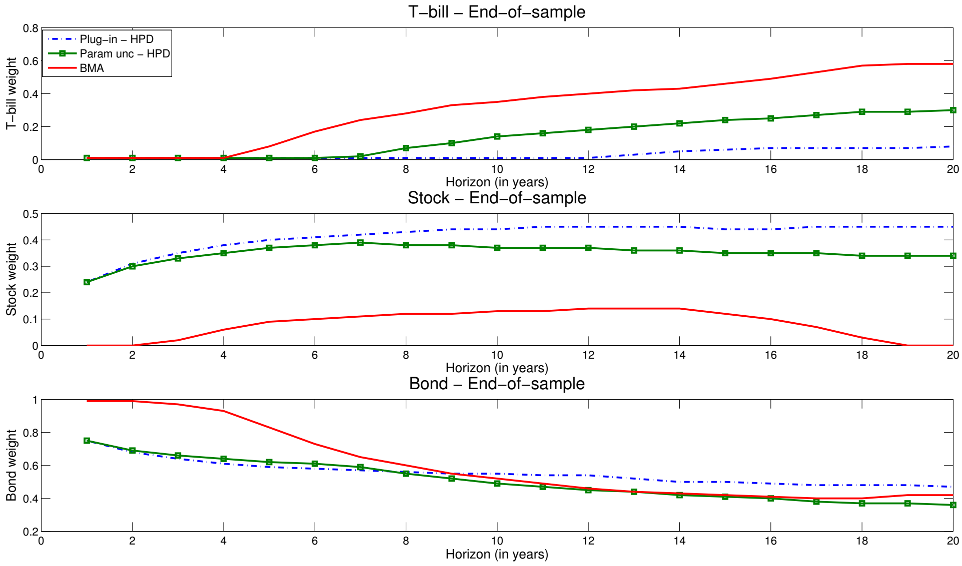
This figure shows a time-series of the four moments of the predictive distribution for three different investment horizons. It is based on 50,000 retained draws from the posterior distribution of the weighted BMA model.

Figure 13: **Portfolio weights versus horizon for different estimation techniques. Predictor variables are equal to their historical average.**



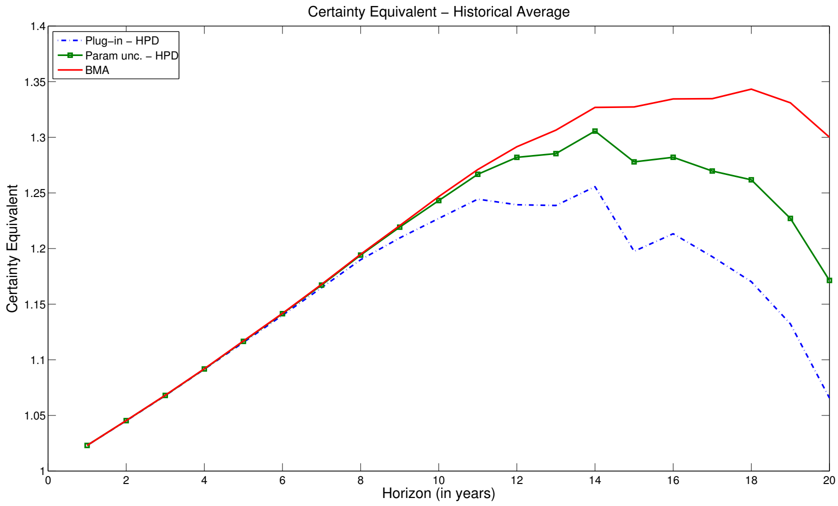
This figure plots T-bill weights (panel A), stock weights (panel B) and bond weights (panel C) against different investment horizons for different estimation techniques. The weights are the optimal weights for a buy-and-hold investor with risk aversion parameter $\gamma = 5$. Weights are either based on the highest posterior probability model without parameter uncertainty (plug-in), on the HPD model incorporating parameter uncertainty (param unc - HPD) and on the BMA specification that includes both parameter and model uncertainty (BMA). Results are based on 100,000 draws from the predictive distribution. Predictor variables are equal to their historical average values.

Figure 14: **Portfolio weights versus horizon for different estimation techniques. Predictor variables are equal to the values at the end of the sample.**



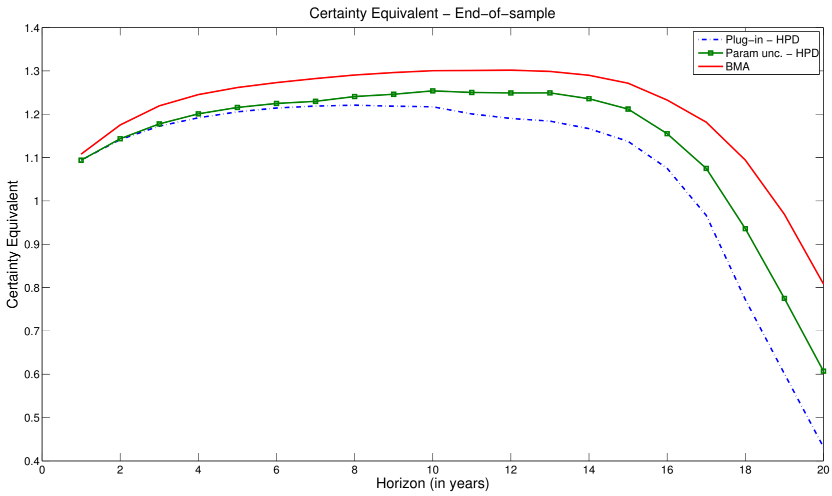
This figure plots T-bill weights (panel A), stock weights (panel B) and bond weights (panel C) against different investment horizons for different estimation techniques. The weights are the optimal weights for a buy-and-hold investor with risk aversion parameter $\gamma = 5$. Weights are either based on the highest posterior probability model without parameter uncertainty (plug-in), on the HPD model incorporating parameter uncertainty (param unc - HPD) and on the BMA specification that includes both parameter and model uncertainty (BMA). Results are based on 100,000 draws from the predictive distribution. Predictor variables are equal to the values in the fourth quarter of 2008.

Figure 15: **Plot of Certainty Equivalent versus horizon for different estimation techniques. Predictor variables are equal to their historical average.**



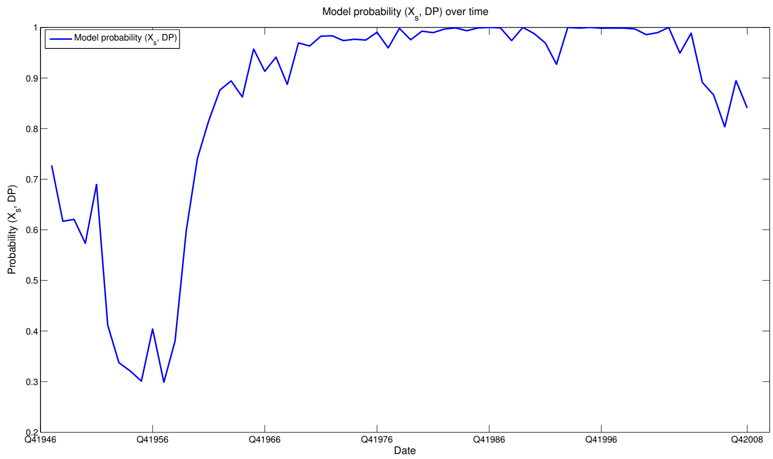
This figure plots the certainty equivalence of different strategies versus the investment horizon. The certainty equivalents are calculated using a specification based on the HPD without parameter uncertainty, a specification based on the HPD with parameter uncertainty and a specification based on the BMA model. All strategies are evaluated using the BMA specification. Results are based on 100,000 retained draws from the predictive distribution. Variables are set equal to their historical average.

Figure 16: **Plot of Certainty Equivalent versus horizon for different estimation techniques. Predictor variables are equal to the values at the end of the sample.**



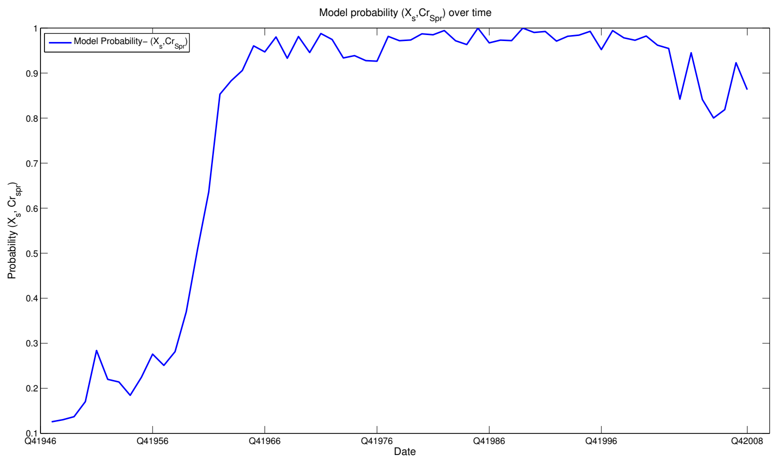
This figure plots the certainty equivalence of strategies versus the investment horizon. The certainty equivalents are calculated using a specification based on the HPD without parameter uncertainty, a specification based on the HPD with parameter uncertainty and a specification based on the BMA model. All strategies are evaluated using the BMA specification. Results are based on 100,000 retained draws from the predictive distribution. Variables are set equal to their end-of-sample values.

Figure 17: **Time-series of model probability** (X_s, DP)



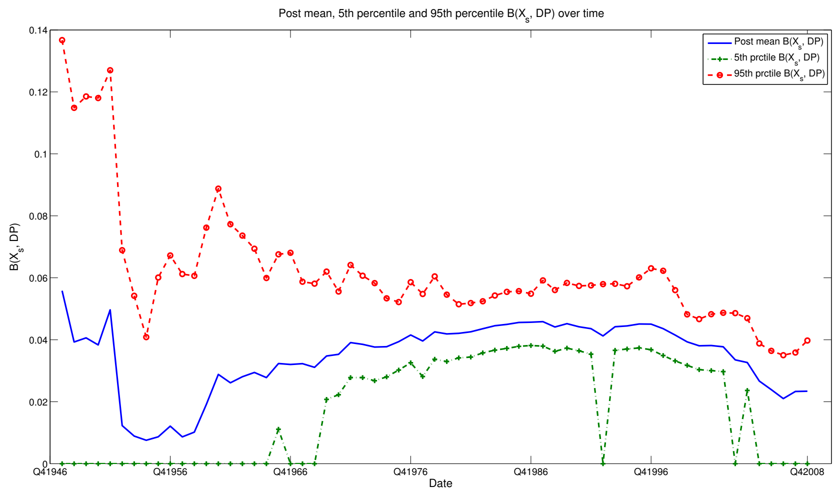
This figure shows the time-series of the model probability (X_s, DP). These are calculated by estimating the BMA specification on an expanding window. The x-axis indicates the value of the last included observation in the window. Results are based on 100,000 retained draws.

Figure 18: **Time-series of model probability** (X_s, Cr_{spr})



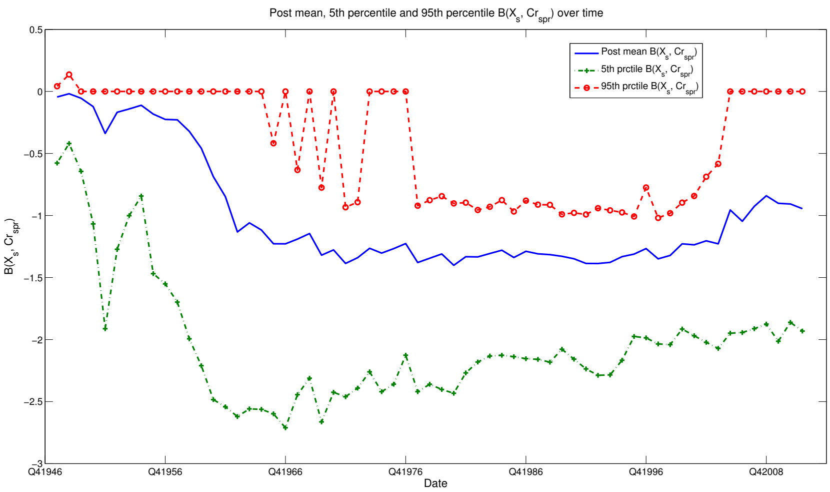
This figure shows the time-series of the model probability (X_s, Cr_{spr}). These are calculated by estimating the BMA specification on an expanding window. The x-axis indicates the value of the last included observation in the window. Results are based on 100,000 retained draws.

Figure 19: **Time-series of posterior mean of coefficient** (X_s, DP)



This figure shows the time-series of the posterior mean of (X_s, DP), its 5th percentile and its 95th percentile. These are calculated by estimating the BMA specification on an expanding window. The x-axis indicates the value of the last included observation in the window. Results are based on 100,000 retained draws.

Figure 20: **Time-series of posterior mean of coefficient** (X_s, Cr_{spr})



This figure shows the time-series of the posterior mean of (X_s, Cr_{spr}), its 5th percentile and its 95th percentile. These are calculated by estimating the BMA specification on an expanding window. The x-axis indicates the value of the last included observation in the window. Results are based on 100,000 retained draws.

List of Tables

1 Summary Statistics of the quarterly data-set 62
2 OLS Estimates and standard errors unrestricted model 63
3 Posterior probability of including a variable 64
4 Relevance of explanatory variables for predicting excess stock returns at different horizons 65
5 Posterior means and standard deviations 66
6 Jointness measure of explanatory variables in equation for excess stock returns . 67
7 Posterior means and standard deviations Best Model 68
8 Robustness: posterior probability of including a variable using a different prior distribution 69
9 Robustness: posterior means and standard deviations using a different prior distribution 70

Table 1: **Summary Statistics of the quarterly data-set**

This table reports the means, standard deviations, minima, maxima, AR(1) coefficients and Sharpe ratios for the ex post T-bill rate (R_{tbill}), the excess stock return (X_s), the excess bond return (X_b), the default risk premium (Def_{pr}), the dividend-to-price ratio (DP), the book-to-market ratio (BM), the price-earnings ratio (PE), the smoothed nominal yield (Y_{nom}), the yield spread (Y_{spr}), the credit spread (Cr_{spr}), net stock issues ($ntis$) and the stock return variance (Var). The data set is quarterly and starts in Q4 of 1926 and ends in Q4 of 2008. Percentages are given as fractions.

	Mean	Std	Min	Max	AR(1)	Sharpe
R_{tbill}	0.0018	0.0132	-0.0878	0.0449	0.5266	
X_s	0.0132	0.1079	-0.4992	0.6399	-0.0408	0.1226
X_b	0.0045	0.0424	-0.1859	0.1854	-0.0555	0.1063
Def_{pr}	0.0005	0.0180	-0.1217	0.0398	-0.0795	
DP	-3.3422	0.4674	-4.5054	-1.6747	0.9711	
BM	-0.6280	0.5089	-2.0778	0.7073	0.9769	
PE	2.9140	0.3834	1.4214	3.9258	0.9601	
Y_{nom}	-0.0002	0.0098	-0.0430	0.0443	0.5746	
Y_{spr}	0.0152	0.0122	-0.0308	0.0416	0.8484	
Cr_{spr}	0.0107	0.0068	0.0032	0.0517	0.9175	
$ntis$	0.0197	0.0247	-0.0530	0.1634	0.9196	
Var	-5.3703	0.9834	-7.9027	-2.1677	0.7591	

Table 2: **OLS Estimates and standard errors unrestricted model**

This table reports the OLS estimates and standard errors of the coefficients in the unrestricted VAR(1) model where the different equations are given in different rows. Panel A reports the OLS estimates and standard deviations. Panel B reports the OLS estimates of the covariance matrix of the error term. The elements on the diagonal are standard deviations x100, the off-diagonal elements are correlations. Note that constants are suppressed in the table.

Panel A: OLS estimates and standard errors of slope coefficients

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	0.4360	-0.0105	-0.0219	-0.0556	0.0024	-0.0061	-0.0002	-0.1278	-0.1149	0.4131	0.0267	-0.0010
	0.0535	0.0063	0.0188	0.0410	0.0043	0.0037	0.0075	0.0832	0.0693	0.1728	0.0291	0.0010
X_s	0.2323	-0.1085	0.3134	0.7852	-0.0315	-0.0884	-0.2228	-0.4919	-0.6622	-3.5008	-0.6495	-0.0080
	0.5049	0.0597	0.1773	0.3871	0.0406	0.0353	0.0710	0.7859	0.6543	1.6315	0.2751	0.0090
X_b	0.2793	-0.0121	-0.1501	-0.2173	0.0052	-0.0210	-0.0278	0.3182	0.8938	-1.0045	-0.1987	0.0047
	0.1971	0.0233	0.0692	0.1511	0.0159	0.0138	0.0277	0.3068	0.2554	0.6369	0.1074	0.0035
Def_{pr}	0.0203	0.0114	0.0479	-0.0801	-0.0138	0.0015	-0.0207	-0.0998	-0.0606	-0.1700	0.1376	0.0013
	0.0851	0.0101	0.0299	0.0652	0.0068	0.0059	0.0120	0.1324	0.1102	0.2748	0.0463	0.0015
DP	-0.6584	0.0880	-0.2978	-0.8914	1.0360	0.0755	0.2029	0.4034	0.6222	1.9710	0.7770	0.0048
	0.5188	0.0613	0.1822	0.3977	0.0418	0.0363	0.0729	0.8074	0.6722	1.6762	0.2826	0.0092
BM	0.0589	0.1230	-0.3403	-0.7438	0.0460	1.0366	0.1485	0.5279	0.9587	1.6938	0.8609	0.0052
	0.5221	0.0617	0.1833	0.4002	0.0420	0.0365	0.0734	0.8126	0.6765	1.6869	0.2844	0.0093
PE	0.3363	-0.1048	0.3029	0.7858	-0.0583	-0.0875	0.7586	-0.5156	-0.7452	-2.6523	-0.5647	-0.0085
	0.5027	0.0594	0.1765	0.3854	0.0405	0.0351	0.0707	0.7824	0.6513	1.6243	0.2739	0.0089
Y_{nom}	-0.0102	0.0026	-0.0068	0.0050	0.0017	0.0065	0.0113	0.6462	0.1828	0.1035	0.0290	-0.0010
	0.0376	0.0044	0.0132	0.0288	0.0030	0.0026	0.0053	0.0585	0.0487	0.1215	0.0205	0.0007
Y_{spr}	-0.0144	-0.0006	0.0184	0.0037	-0.0023	-0.0044	-0.0084	0.0215	0.7842	0.0426	-0.0016	0.0006
	0.0306	0.0036	0.0107	0.0234	0.0025	0.0021	0.0043	0.0476	0.0396	0.0987	0.0166	0.0005
Cr_{spr}	0.0092	0.0013	-0.0138	-0.0194	0.0003	-0.0005	-0.0012	-0.0198	-0.0011	0.7925	0.0045	0.0010
	0.0135	0.0016	0.0048	0.0104	0.0011	0.0009	0.0019	0.0211	0.0175	0.0438	0.0074	0.0002
$ntis$	-0.0961	0.0106	-0.0177	-0.0051	0.0025	-0.0029	0.0014	-0.1967	-0.1518	0.1491	0.9085	-0.0012
	0.0449	0.0053	0.0158	0.0344	0.0036	0.0031	0.0063	0.0698	0.0581	0.1450	0.0244	0.0008
Var	-6.1149	0.4726	-2.5670	-5.4110	0.2098	-0.5217	0.0301	-3.8466	-4.4843	58.2268	0.4973	0.5195
	2.9531	0.3490	1.0369	2.2638	0.2377	0.2064	0.4152	4.5963	3.8265	9.5420	1.6089	0.0526

Panel B: OLS estimates of covariance matrix of residuals

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	1.0901	-0.0537	0.2521	-0.0658	0.0288	-0.0180	-0.0505	-0.1265	0.0017	0.2609	-0.1156	-0.0318
X_s		10.2926	0.0626	0.2393	-0.9746	-0.8620	0.9949	0.0040	-0.0538	-0.5542	0.1460	-0.4292
X_b			4.0179	-0.3821	-0.0598	-0.0713	0.0643	-0.5539	0.0919	0.1569	0.0488	0.0788
Def_{pr}				1.7339	-0.2520	-0.2034	0.2432	0.0537	0.1500	-0.1481	0.1673	-0.1693
DP					10.5747	0.8528	-0.9757	0.0030	0.0413	0.5291	-0.1241	0.4269
BM						10.6419	-0.8651	0.0106	0.0511	0.4657	-0.1112	0.3443
PE							10.2468	-0.0014	-0.0484	-0.5375	0.1409	-0.4372
Y_{nom}								0.7662	-0.8454	-0.1325	-0.0593	-0.0325
Y_{spr}									0.6228	0.0665	0.0686	-0.0013
Cr_{spr}										0.2760	-0.2346	0.2580
$ntis$											0.9145	0.1313
Var												60.1963

Table 3: **Posterior probability of including a variable**

This table reports the posterior probability of including a variable in the weighted Bayesian Model Averaging model. The different equations are given in the different rows. Note that the right-hand-side variables are lagged by one period. Results are based on 500,000 retained draws from the posterior distribution.

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	1.0000	0.3189	0.0580	0.1081	0.1292	0.2880	0.2262	0.0808	0.1199	0.4725	0.2264	0.0779
X_s	0.3320	0.0852	0.0695	0.0534	0.8406	0.0986	0.3831	0.0667	0.1741	0.8633	0.4850	0.0900
X_b	0.0679	0.0622	0.0915	0.1969	0.0670	0.0619	0.0619	0.5457	1.0000	0.0884	0.0816	0.1498
Def_{pr}	0.0696	0.0854	0.1752	0.6810	0.1397	0.1802	0.2651	0.0761	0.0605	0.0936	0.7020	0.1954
DP	0.6725	0.1766	0.0795	0.2692	1.0000	0.1088	0.2329	0.0487	0.0587	0.6288	0.5781	0.1040
BM	0.0960	0.0702	0.0707	0.0621	0.1216	1.0000	0.1367	0.0572	0.0799	0.0776	0.4615	0.0777
PE	0.2568	0.0621	0.0590	0.0519	0.2752	0.0620	1.0000	0.0664	0.1529	0.4437	0.7225	0.1532
Y_{nom}	0.0620	0.0521	0.0832	0.0976	0.0696	0.0698	0.0822	1.0000	0.8359	0.1894	0.1371	0.2267
Y_{spr}	0.0717	0.0613	0.1143	0.0954	0.0590	0.0913	0.0776	0.0692	1.0000	0.4268	0.0982	0.2630
Cr_{spr}	0.0693	0.0521	0.1280	0.0595	0.1578	0.1505	0.3671	0.0937	0.0634	1.0000	0.0793	1.0000
$ntis$	0.1077	0.5003	0.1346	0.0702	0.0979	0.0926	0.1805	0.2396	0.2192	0.1096	1.0000	0.2251
Var	0.1696	0.0562	0.0809	0.1125	0.1049	0.9850	0.2030	0.0696	0.1049	1.0000	0.0572	1.0000

Table 4: **Relevance of explanatory variables for predicting excess stock returns at different horizons**

This table reports the posterior probability of including a variable in the weighted Bayesian Model Averaging model when the model is iterated forward by H periods. It shows which variables are important for predicting excess stock returns at different horizons. Results are based on 500,000 retained draws from the posterior distribution.

H	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
1	0.3320	0.0852	0.0695	0.0534	0.8406	0.0986	0.3831	0.0667	0.1741	0.8633	0.4850	0.0900
2	0.8287	0.5303	0.2946	0.3877	0.9967	0.4646	0.7203	0.3538	0.4897	0.9982	0.8150	0.9055
4	0.9226	0.8058	0.5809	0.6629	0.9990	0.9939	0.8984	0.6930	0.8775	0.9995	0.9814	0.9993
8	0.9261	0.8185	0.5997	0.6794	0.9995	0.9946	0.9056	0.7207	0.9045	0.9995	0.9861	0.9995
20	0.9261	0.8185	0.5997	0.6794	0.9995	0.9946	0.9056	0.7207	0.9045	0.9995	0.9861	0.9995
40	0.9261	0.8185	0.5997	0.6794	0.9995	0.9946	0.9056	0.7207	0.9045	0.9995	0.9861	0.9995

Table 5: **Posterior means and standard deviations**

This table reports the posterior mean and standard deviation of the coefficients in the weighted Bayesian Model Averaging model where the different equations are given in different rows. Panel A reports the posterior mean and standard deviations of the slope coefficients. Panel B reports the posterior mean of the elements of the covariance matrix of the error term. The elements on the diagonal are standard deviations x100, the off-diagonal elements are correlations. Note that constants are suppressed in the table. Results are based on 500,000 retained draws from the posterior distribution.

	Panel A: Posterior means and standard deviations											
	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	0.4745	-0.0037	0.0001	-0.0047	0.0004	-0.0010	0.0013	-0.0060	-0.0094	0.1474	0.0104	0.0000
	0.0538	0.0062	0.0038	0.0175	0.0019	0.0020	0.0034	0.0298	0.0340	0.1871	0.0229	0.0003
X_s	-0.0393	-0.0005	0.0007	-0.0013	0.0234	-0.0005	-0.0163	-0.0026	0.0163	-0.9449	-0.1049	0.0000
	0.0844	0.0030	0.0054	0.0132	0.0125	0.0034	0.0264	0.0225	0.0448	0.5820	0.2242	0.0005
X_b	0.0067	-0.0001	-0.0040	-0.0221	0.0000	0.0000	0.0000	0.3406	0.9993	0.0049	0.0002	0.0004
	0.0353	0.0030	0.0204	0.0565	0.0010	0.0009	0.0013	0.3655	0.2677	0.1085	0.0248	0.0014
Def_{pr}	0.0027	0.0007	0.0069	-0.0939	-0.0002	0.0006	-0.0018	-0.0066	0.0029	0.0087	0.0749	0.0003
	0.0207	0.0036	0.0184	0.0768	0.0027	0.0016	0.0043	0.0388	0.0247	0.0662	0.0589	0.0008
DP	-0.2152	-0.0037	0.0023	-0.0396	0.9812	-0.0009	0.0106	-0.0019	-0.0020	-0.6841	0.1890	-0.0003
	0.1824	0.0097	0.0128	0.0764	0.0133	0.0043	0.0242	0.0338	0.0349	0.6388	0.2600	0.0011
BM	0.0265	0.0016	-0.0034	0.0064	0.0023	0.9708	-0.0042	-0.0036	0.0143	-0.0119	0.2100	-0.0002
	0.1152	0.0097	0.0236	0.0503	0.0087	0.0149	0.0148	0.0806	0.0925	0.2474	0.2954	0.0014
PE	0.0165	0.0000	-0.0002	-0.0004	-0.0023	-0.0002	0.9659	0.0005	-0.0119	-0.1336	-0.0176	-0.0002
	0.0750	0.0024	0.0045	0.0122	0.0115	0.0029	0.0252	0.0207	0.0390	0.5529	0.2269	0.0007
Y_{nom}	0.0003	0.0000	-0.0005	-0.0003	0.0000	0.0000	0.0001	0.6644	0.1034	-0.0121	0.0012	-0.0002
	0.0032	0.0004	0.0035	0.0039	0.0001	0.0005	0.0007	0.0404	0.0591	0.0462	0.0042	0.0005
Y_{spr}	0.0003	0.0000	0.0008	-0.0007	0.0000	-0.0001	0.0000	0.0017	0.8510	0.0320	0.0004	0.0002
	0.0032	0.0004	0.0038	0.0037	0.0001	0.0005	0.0007	0.0147	0.0417	0.0542	0.0028	0.0005
Cr_{spr}	0.0004	0.0000	-0.0006	0.0002	0.0001	-0.0001	-0.0005	-0.0014	0.0004	0.8090	-0.0003	0.0009
	0.0032	0.0003	0.0018	0.0019	0.0003	0.0004	0.0009	0.0063	0.0033	0.0347	0.0021	0.0002
$ntis$	-0.0050	0.0057	-0.0025	0.0019	0.0002	-0.0001	0.0007	-0.0287	-0.0214	-0.0104	0.9230	-0.0002
	0.0197	0.0066	0.0081	0.0109	0.0014	0.0006	0.0022	0.0622	0.0487	0.0458	0.0223	0.0005
Var	-0.7063	0.0010	-0.0646	-0.2388	0.0160	-0.3910	-0.0844	0.1473	-0.3725	50.1991	-0.0229	0.5321
	1.8988	0.0735	0.3188	0.8930	0.0743	0.1640	0.2257	1.0625	1.4352	8.0194	0.3682	0.0456
	Panel B: Posterior mean of covariance matrix											
	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	1.1483	-0.0588	0.2525	-0.0743	0.0381	-0.0127	-0.0556	-0.1318	0.0114	0.2672	-0.0945	-0.0092
X_s		10.9314	0.0695	0.2406	-0.9752	-0.8669	0.9951	-0.0265	-0.0256	-0.5449	0.1320	-0.4340
X_b			4.2376	-0.3799	-0.0670	-0.0775	0.0715	-0.5603	0.1003	0.1725	0.0435	0.0883
Def_{pr}				1.8222	-0.2538	-0.2039	0.2444	0.0458	0.1588	-0.1516	0.1540	-0.1776
DP					11.2089	0.8575	-0.9763	0.0316	0.0157	0.5215	-0.1101	0.4338
BM						11.2405	-0.8697	0.0403	0.0229	0.4642	-0.1038	0.3499
PE							10.8847	-0.0317	-0.0207	-0.5285	0.1275	-0.4415
Y_{nom}								0.8111	-0.8468	-0.1296	-0.0510	-0.0312
Y_{spr}									0.6562	0.0569	0.0612	-0.0059
Cr_{spr}										0.2882	-0.2190	0.2711
$ntis$											0.9637	0.1508
Var												63.4195

Table 6: **Jointness measure of explanatory variables in equation for excess stock returns**

This table reports a jointness measure for pairs of variables in the equation for excess stock returns. The jointness measure is based on Ley and Steel (2007) and ranges from 0 (decisive evidence in favour of disjointness of the variables) to ∞ (decisive evidence in favour of jointness of the variables). Results are based on 500,000 retained draws from the posterior distribution.

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	-	0.0881	0.0707	0.0510	0.4690	0.0914	0.2428	0.0711	0.1148	0.4321	0.3180	0.0619
X_s	0.0881	-	0.0470	0.0267	0.0959	0.0457	0.0776	0.0405	0.0671	0.0914	0.0909	0.0441
X_b	0.0707	0.0470	-	0.0355	0.0755	0.0407	0.0601	0.0220	0.0467	0.0773	0.0632	0.0441
Def_{pr}	0.0510	0.0267	0.0355	-	0.0589	0.0503	0.0510	0.0294	0.0385	0.0564	0.0464	0.0521
DP	0.4690	0.0959	0.0755	0.0589	-	0.0905	0.2986	0.0702	0.2103	2.5790	0.8327	0.0917
BM	0.0914	0.0457	0.0407	0.0503	0.0905	-	0.1560	0.0483	0.0581	0.1102	0.1015	0.0553
PE	0.2428	0.0776	0.0601	0.0510	0.2986	0.1560	-	0.0726	0.1331	0.6960	0.3991	0.0866
Y_{nom}	0.0711	0.0405	0.0220	0.0294	0.0702	0.0483	0.0726	-	0.0439	0.0700	0.0662	0.0319
Y_{spr}	0.1148	0.0671	0.0467	0.0385	0.2103	0.0581	0.1331	0.0439	-	0.2158	0.1755	0.0736
Cr_{spr}	0.4321	0.0914	0.0773	0.0564	2.5790	0.1102	0.6960	0.0700	0.2158	-	0.7769	0.0985
$ntis$	0.3180	0.0909	0.0632	0.0464	0.8327	0.1015	0.3991	0.0662	0.1755	0.7769	-	0.0881
Var	0.0619	0.0441	0.0441	0.0521	0.0917	0.0553	0.0866	0.0319	0.0736	0.0985	0.0881	-

Table 7: **Posterior means and standard deviations Best Model**

This table reports the model that receives the highest posterior probability where the different equations are denoted by different rows in the table. The table reports the posterior means and standard deviations. A "-" means that a variable is excluded in the highest posterior probability model. Results are based on 500,000 retained draws from the posterior distribution.

	Panel A: Posterior means and standard deviations											
	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	0.5036	-0.0124	-	-	-	-	-	-	-	-	-	
	0.0452	0.0055										
X_s	-	-	-	-	0.0272	-	-	-	0.0835	-0.7729	-0.5190	-
					0.0037				0.0525	0.1267	0.1891	
X_b	-	-	-	-	-	-	-	-	0.7994	0.2096	-	-
									0.1707	0.1707		
Def_{pr}	-	-	-	-0.1515	-	-	-0.0042	-	-	-	0.0834	-
				0.0470			0.0022				0.0365	
DP	-0.4315	-	-	-	0.9751	-	-	-	-	-	0.6849	-0.0038
	0.1044				0.0046						0.1972	0.0014
BM	-	-	-	-	-	0.9812	-	-	-	-	0.7078	-
						0.0065					0.2091	
PE	0.0937	-	-	-	-	-	0.9794	-	-	-	-0.4347	-
	0.0474						0.0047				0.1897	
Y_{nom}	-	-	-	-	-	-0.0001	-	0.6974	0.1259	-	-	-
						0.0002		0.0127	0.0367			
Y_{spr}	-	-	-	-	-	-	-	-	0.8464	-	-	-
									0.0293			
Cr_{spr}	-	-	-	-	0.0008	-	-	-	0.0034	0.7931	-	0.0009
					0.0003				0.0111	0.0299		0.0002
$ntis$	-	-	-	-	-	-	-	-	-	-	0.9212	-0.0012
											0.0205	0.0005
Var	-	-	-	-	-	-0.2931	-	-	-	49.7713	-	0.5350
						0.0681				6.8995		0.0441
	Panel B: Posterior mean of covariance matrix											
	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	1.1465	-0.0648	0.2373	-0.0727	0.0413	-0.0073	-0.0616	-0.1303	0.0210	0.2682	-0.0773	0.0017
X_s		10.8456	0.0590	0.2470	-0.9746	-0.8644	0.9951	-0.0192	-0.0275	-0.5427	0.1243	-0.4319
X_b			4.2459	-0.3886	-0.0512	-0.0637	0.0601	-0.5577	0.0967	0.1727	0.0342	0.0919
Def_{pr}				1.8163	-0.2620	-0.2103	0.2509	0.0490	0.1593	-0.1552	0.1625	-0.1850
DP					11.0966	0.8545	-0.9757	0.0204	0.0190	0.5209	-0.1057	0.4325
BM						11.1029	-0.8672	0.0283	0.0281	0.4639	-0.0976	0.3497
PE							10.7938	-0.0234	-0.0234	-0.5266	0.1201	-0.4396
Y_{nom}								0.8126	-0.8469	-0.1308	-0.0525	-0.0348
Y_{spr}									0.6594	0.0589	0.0679	-0.0032
Cr_{spr}										0.2876	-0.2130	0.2704
$ntis$											0.9661	0.1561
Var												63.4407

Table 8: **Robustness: posterior probability of including a variable using a different prior distribution**

This table reports the posterior probability of including a variable in the weighted Bayesian Model Averaging model. We use a beta-prior for parameter q and an inverse gamma prior for parameter g. The different equations are given in the different rows. Note that the right-hand-side variables are lagged by one period. Results are based on 500,000 retained draws from the posterior distribution.

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	1.0000	0.1826	0.0255	0.0434	0.0429	0.1124	0.0794	0.0410	0.0421	0.1552	0.0836	0.0359
X_s	0.2332	0.0385	0.0271	0.0273	0.9650	0.0410	0.1524	0.0348	0.1028	0.7348	0.2405	0.0454
X_b	0.0520	0.0236	0.0344	0.1101	0.0275	0.0230	0.0243	0.3555	0.9983	0.0438	0.0282	0.0653
Def_{pr}	0.0257	0.0339	0.0980	0.4661	0.0630	0.0822	0.1331	0.0281	0.0277	0.0522	0.5011	0.0998
DP	0.4593	0.1002	0.0349	0.1322	1.0000	0.0430	0.0654	0.0198	0.0213	0.8010	0.3096	0.0420
BM	0.0375	0.0347	0.0284	0.0268	0.0592	1.0000	0.0530	0.0211	0.0239	0.0337	0.2106	0.0305
PE	0.1204	0.0190	0.0235	0.0310	0.1156	0.0299	1.0000	0.0277	0.0818	0.3829	0.8025	0.0733
Y_{nom}	0.0253	0.0227	0.0239	0.0519	0.0270	0.0303	0.0243	1.0000	0.7649	0.0858	0.0472	0.0771
Y_{spr}	0.0232	0.0239	0.0336	0.0480	0.0265	0.0309	0.0302	0.0361	1.0000	0.2199	0.0355	0.1094
Cr_{spr}	0.0271	0.0257	0.0544	0.0305	0.1028	0.0743	0.2813	0.0427	0.0343	1.0000	0.0339	0.9992
$ntis$	0.0379	0.3455	0.0606	0.0400	0.0454	0.0414	0.1052	0.0822	0.0798	0.0705	1.0000	0.1539
Var	0.0761	0.0340	0.0320	0.0504	0.0487	0.9789	0.0934	0.0270	0.0398	1.0000	0.0266	1.0000

Table 9: **Robustness: posterior means and standard deviations using a different prior distribution**

This table reports the posterior mean and standard deviation of the coefficients in the weighted Bayesian Model Averaging model where the different equations are given in different rows. We use a beta-prior for q and a inverse gamma prior for g. Panel A reports the posterior mean and standard deviations of the slope coefficients. Panel B reports the posterior mean of the elements of the covariance matrix of the error term. The elements on the diagonal are standard deviations x100, the off-diagonal elements are correlations. Note that constants are suppressed in the table. Results are based on 500,000 retained draws from the posterior distribution.

Panel A: Posterior means and standard deviations

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	0.4865	-0.0021	0.0000	-0.0017	0.0001	-0.0003	0.0003	-0.0029	-0.0021	0.0407	0.0036	0.0000
	0.0496	0.0051	0.0025	0.0107	0.0010	0.0012	0.0018	0.0199	0.0161	0.1112	0.0139	0.0002
X_s	-0.0289	-0.0002	0.0002	-0.0005	0.0265	-0.0002	-0.0047	-0.0013	0.0099	-0.6633	-0.0336	0.0000
	0.0638	0.0018	0.0029	0.0086	0.0081	0.0014	0.0141	0.0153	0.0352	0.4973	0.1300	0.0003
X_b	0.0052	-0.0001	-0.0010	-0.0123	0.0000	0.0000	0.0000	0.2242	0.9490	0.0072	0.0005	0.0001
	0.0298	0.0017	0.0103	0.0421	0.0005	0.0005	0.0007	0.3375	0.2717	0.0669	0.0117	0.0007
Def_{pr}	0.0009	0.0002	0.0039	-0.0625	0.0001	0.0002	-0.0007	-0.0024	0.0015	0.0063	0.0513	0.0002
	0.0123	0.0019	0.0140	0.0749	0.0012	0.0010	0.0023	0.0234	0.0172	0.0475	0.0577	0.0006
DP	-0.1422	-0.0022	0.0010	-0.0193	0.9802	-0.0003	0.0023	-0.0005	-0.0007	-0.9725	0.0779	-0.0001
	0.1743	0.0078	0.0081	0.0567	0.0091	0.0022	0.0118	0.0206	0.0213	0.6287	0.1671	0.0007
BM	0.0089	0.0007	-0.0014	0.0031	0.0013	0.9763	-0.0016	-0.0015	0.0041	-0.0102	0.0789	-0.0001
	0.0656	0.0066	0.0149	0.0337	0.0065	0.0117	0.0090	0.0478	0.0498	0.1481	0.1905	0.0008
PE	0.0077	0.0000	-0.0001	-0.0001	0.0005	0.0000	0.9768	0.0006	-0.0067	0.1164	0.0573	-0.0001
	0.0448	0.0012	0.0025	0.0083	0.0072	0.0011	0.0142	0.0126	0.0292	0.4709	0.1351	0.0004
Y_{nom}	0.0001	0.0000	-0.0001	-0.0002	0.0000	0.0000	0.0000	0.6794	0.0932	-0.0023	0.0003	-0.0001
	0.0018	0.0002	0.0013	0.0025	0.0001	0.0001	0.0001	0.0370	0.0618	0.0249	0.0019	0.0003
Y_{spr}	0.0001	0.0000	0.0002	-0.0003	0.0000	0.0000	0.0000	0.0011	0.8657	0.0123	0.0001	0.0001
	0.0017	0.0002	0.0016	0.0023	0.0001	0.0001	0.0001	0.0110	0.0422	0.0321	0.0013	0.0003
Cr_{spr}	0.0001	0.0000	-0.0002	0.0001	0.0001	0.0000	-0.0004	-0.0007	0.0002	0.8169	-0.0001	0.0008
	0.0019	0.0002	0.0012	0.0014	0.0002	0.0003	0.0007	0.0043	0.0025	0.0353	0.0013	0.0002
$ntis$	-0.0015	0.0040	-0.0010	0.0013	0.0000	0.0000	0.0003	-0.0086	-0.0065	-0.0083	0.9235	-0.0002
	0.0108	0.0062	0.0052	0.0088	0.0008	0.0004	0.0014	0.0350	0.0269	0.0391	0.0204	0.0005
Var	-0.3037	-0.0003	-0.0231	-0.1077	0.0058	-0.3310	-0.0319	0.0619	-0.1239	49.8479	-0.0133	0.5443
	1.2676	0.0571	0.1879	0.6065	0.0492	0.1270	0.1538	0.6543	0.8219	7.4173	0.2447	0.0461

Panel B: Posterior mean of covariance matrix

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
R_{tbill}	1.1532	-0.0644	0.2508	-0.0719	0.0448	-0.0077	-0.0612	-0.1366	0.0216	0.2658	-0.0941	-0.0056
X_s		10.9552	0.0694	0.2344	-0.9751	-0.8670	0.9951	-0.0264	-0.0264	-0.5422	0.1286	-0.4319
X_b			4.2419	-0.3801	-0.0678	-0.0779	0.0715	-0.5615	0.1032	0.1741	0.0378	0.0894
Def_{pr}				1.8280	-0.2463	-0.1965	0.2380	0.0447	0.1616	-0.1474	0.1493	-0.1768
DP					11.2450	0.8577	-0.9763	0.0322	0.0163	0.5184	-0.1067	0.4327
BM						11.2647	-0.8699	0.0414	0.0224	0.4631	-0.1014	0.3486
PE							10.9089	-0.0317	-0.0214	-0.5258	0.1243	-0.4395
Y_{nom}								0.8143	-0.8477	-0.1317	-0.0454	-0.0331
Y_{spr}									0.6593	0.0590	0.0573	-0.0043
Cr_{spr}										0.2882	-0.2186	0.2710
$ntis$											0.9661	0.1541
Var												63.4438