Strategic asset allocation

The effect of uncertainty on portfolio choice

Strategic asset allocation

The effect of uncertainty on portfolio choice

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. mr. G.P.M.F. Mols
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op woensdag 25 mei 2011 om 12.00 uur

door

Bart Franciscus Diris

UNIVERSITAIRE
PERS MAASTRICHT

To my parents, my brother and Suzan

Acknowledgements

The past years as a PhD student have been some of the best years of my life. I learned a lot - from econometrics and finance to academic writing and presenting - met many interesting people along the way and lived in interesting cities such as Stockholm, Boston and - of course - *Mestreech*. The research process was a long, challenging, but ultimately rewarding process. It was a journey that I could not have made without the support of the following people.

Firstly, I would like to thank my PhD supervisors, Peter Schotman and Franz Palm. Peter and Franz, thank you for all the freedom you gave me in formulating my own ideas and solving my own problems, for the trust you showed in me when I was working independently and for the support you gave me throughout the process. Franz, you taught me how to take the helicopter view, how to clearly motivate all my choices and how to always keep the reader in mind. Peter, you challenged me to think out-of-the-box, to confront my research with very different (your) approaches and to keep things sophisticatedly simple.

A next word of gratitude goes to the members of my assessment committee Jean-Pierre Urbain, Antoon Pelsser and Luc Bauwens. Jean-Pierre, I would like to thank you as well for some important early career advice. I would also like to express my sincere thanks to my hosts at respectively SIFR and Harvard, Per Strömberg and John Campbell.

Next, I would like to thank my friends and colleagues at the Finance and KE departments of Maastricht University. Bas, thank you for many years of friendship and support. I am happy that you are my paranymph. Thies, Daniel, Simone, Daniela and Kathrin, thank you for all the good times we had during and after office hours. Hopefully, when you read this, we have finally been to Cologne and Berlin. Francien,

Cécile, Carina, Els, Karin en Haydée, thank you for the support with administrative duties, certainly in the last phase of my PhD after I left Maastricht.

Mijn grootste dank gaat uit naar mijn familie. Pap, mam en Ron, bedankt voor alle liefde en steun gedurende mijn leven. Ron, ik ben er trots op dat we beiden een PhD doen en dat je mijn paranymph bent. Suzan, het is onmogelijk je slechts in een paar woorden te bedanken. Heel veel dank voor al je steun, liefde en onvoorwaardelijk geloof in mijn kunnen gedurende de laatste jaren. Dank je!

Breda, March 2011

Contents

CONTENTS

List of Figures

List of Tables

1

Introduction

Individuals and institutions (e.g. pension funds) invest financial wealth in assets to meet a long-term goal. Individuals might want to save money for retirement or for the education of their children. Pension funds invest on behalf of their participants to provide them with retirement income. The funds want to guarantee a safe pension to them at the lowest cost (in terms of contributions). Given a set of long-term preferences, the investor determines the allocation of financial wealth to different asset classes, such as stocks, bonds, short-term deposits, real-estate, commodities and hedge funds. Strategic asset allocation is the choice how to invest in these broad asset classes to meet ones long-term goal. The allocations will vary over time due to changes in investment opportunities, the investment horizon, and the long-term macro-economic risk factors like inflation and interest rates.

Already 40 years ago, Merton (1969,1971) showed that long-term investors should hold hedge portfolios that anticipate future changes in the investment opportunities. At the end of the 1990s, early 2000s there was a huge surge in popularity of the strategic asset allocation literature due to the finding that stock returns might be predictable (by e.g. the dividend-to-price ratio and interest rates) and due to the large increase in computer power.

Campbell and Viceira (2002) provide an overview of the advancements of the literature in the early 2000s. Empirically, the fact that the dividend-to-price ratio might predict stock returns leads to mean-reversion in stock returns. It means that lower than expected stock returns are followed by higher future expected stock returns. This negative autocorrelation in stock returns makes stocks safer (in terms of variance) and

therefore more attractive in the long-run. The mean-reversion is also the most important driver of the hedge portfolio of long-term investors. Such investors want to invest in a security whose return is high when future stock returns are expected to be low. Stocks turn out to be such a security. Another important consideration for long-term investors is inflation and interest rate risk. The riskfree asset of a long-term investor is a long-term inflation-indexed bond. Short-term T-bills are not the riskfree asset, since they must be rolled over repeatedly. Furthermore, the long-term real returns of nominal bonds are also not save, since they are subject to inflation risk which makes them unattractive in the long-run. Long-term investors need to take these risks into account in their hedge portfolio.

The results above are obtained using stylized models in which the true parameters and the true model are assumed to be known. Clearly, any form of model misspecification can have a large impact on the composition of the calculated portfolios. In the 2000s a new branch of the literature emerged that analyzes the quality of the models and the sensitivity of the results to changes in the setting. Barberis (2000) dropped the common assumption that investors know the true set of parameters by incorporating parameter uncertainty in the decision process of the investor. He finds that stocks are still more attractive in the long-run, although the difference between long-term and short-term stock allocations is reduced by the incorporation of parameter uncertainty. Xia (2001) and Brandt, Goyal, Santa-Clara, and Stroud (2005) add learning about predictability and find that ignoring learning may lead to a portfolio with a lower utility for the investor. Guidolin and Timmermann (2007) and Pettenuzzo and Timmermann (2010) consider the effects of model instability by considering regime-switching models and conclude that ignoring structural breaks can also lead to substantial utility costs.

There are many reasons to be skeptical about the predictions of the strategic asset allocation literature. Firstly, optimal portfolios turn out be extreme, unrealistic and very sensitive to changes in predictor variables. As an example, in Campbell, Chan, and Viceira (2003) stock weights vary between -1000% and 1000% and yearly changes of 500% are not uncommon. Secondly, Goyal and Welch (2008) document the poor out-of-sample predictability of stock returns by showing that the historical average of returns predicts stock returns equally well as all considered predictors. This casts doubt on the mean-reversion in stock returns and the time-variation of optimal portfolios. Thirdly, strategic asset allocation is even more complicated than myopic asset allocation due

to the presence of the hedge component. While myopic portfolios are only affected by estimation error in the myopic component, strategic portfolios are effected by errors in both the myopic and hedge component.

A rich literature documents the poor performance of myopic portfolios. The reason why short-term portfolios perform so badly is error maximization. The inputs of portfolio optimizers (means, variances etcetera) are estimated with error and optimizers overweight (underweight) securities with large (small) returns. These are the ones most likely to have large estimation errors as argued in Michaud (1989). DeMiguel, Garlappi, and Uppal (2009) analyze 14 different models to calculate the inputs of mean-variance optimization and show that none of them consistently outperforms a simple $1/N$ rule (equal weight in all assets). Since strategic portfolios are even more susceptible to errors, they might even perform worse.

1.1 Strategic asset allocation: The effect of uncertainty on portfolio choice

The literature cited suggests that it is unclear whether the potential gains from strategic asset allocation can be realized in practice. On one hand the bad (out-of-sample) performance of (simpler) myopic portfolios does not bode well, but on the other hand recent advancements such as the incorporation of parameter uncertainty in decision making might improve performance considerably. The review also suggests that constructing myopic portfolios consisting of many individual stocks leads to highly unstable portfolios.

In this thesis, our main objective is to analyze whether the premises of the strategic asset allocation literature hold in realistic settings that include potential parameter uncertainty, model uncertainty and model instability. As a side objective we also investigate whether we can construct robust myopic portfolios (i.e. portfolios that are stable over time) consisting of almost 2,000 individual stocks that are able to generate large expected returns with minimal risk. This thesis addresses (and answers) the following research questions:

- How can we set-up a panel data model to explain individual stock returns using multiple firm characteristics and can we use this model to construct robust portfolios?

1. INTRODUCTION

- Can the potential gains from the strategic asset allocation literature be realized in an out-of-sample test and how can the performance be improved?

- How can we develop a methodology to incorporate model uncertainty regarding long-term predictions and what is the impact of model uncertainty on long-term investors?

- How can we set-up and estimate a rich time-varying parameter model and what kind of time-variation should long-term investors take into account?

Firstly, we analyze how to set-up a model to explain the cross-section of individual stock returns using multiple firm characteristics. The standard sorting methodology does not work in such a setting and therefore we develop a panel data model that is able to handle many different characteristics jointly. Secondly, we investigate the performance of the strategic portfolios out-of-sample. Even though the out-of-sample performance is very relevant for long-term investors, such an out-of-sample test has not been done. Thirdly, we document the impact of model uncertainty on the predictive distribution of stock returns and the decisions of long-term investors. Although different models could lead to completely different predictions of future stock returns, model uncertainty is almost always ignored in the strategic asset allocation literature.[1] Fourthly, we use a time-varying parameter model to analyze the impact of model instability on the predictive distribution and asset allocations of long-term investors and to get insights in the importance of time-variation of parameters. Other papers (e.g. Pettenuzzo and Timmermann (2010), Guidolin and Timmermann (2007)) use regime-switching models instead and can therefore not distinguish between the importance of the different parameters.[2]

The concepts parameter uncertainty, model uncertainty and model instability play an important role in this thesis. In order to explain these concepts, we consider an example. Suppose we use the following model for stock returns y_t

$$y_t = \mu + \epsilon_t,$$

---

[1] An exception is Avramov (2002), but his setting is not appropriate for long-term predictions as we argue in chapter 4.

[2] It is well-known in the literature that the (error) volatility of stock returns changes considerably and persistently over time. By letting all parameters change jointly in a regime-switching model, it is unclear whether the documented model instability just reflects these well-known changes in volatility or changes in other parameters such as intercepts or slopes.

with

$$\epsilon_t \sim N(0, \sigma^2),$$

where μ is the expected return, ϵ_t is the error term and σ^2 is the variance of the error term. Usually, historical data is used to estimate parameters μ and σ^2.

It is instructive to make a distinction between a model and the data-generating process (DGP). The DGP is the unknown mechanism that actually generated the data. A model is a set of data-generating processes (different parameters lead to different DGPs) and is correctly specified when the true DGP belongs to the model under study. Note however that a correctly specified model can still be useless in practice, think for example about models with a very large number of parameters that have to be estimated.

Let us consider the equation above. The first cause of uncertainty is called "error uncertainty". It reflects the fact that the error term ϵ_t is unknown a priori, since it captures unexpected events (e.g. unexpected major news about interest rates). The second determinant of uncertainty is coined "parameter uncertainty". It is present due to the fact that μ and σ^2 are unknown and need to be estimated using historical data. A slightly different sample might already lead to very different parameter estimates and therefore return predictions. A third cause for uncertainty is called "model uncertainty". Model uncertainty is present, since we do not know the "best" model.[1] Instead we select a model (often using data) and our choice can be wrong. Another model that says that stock returns are a function of the interest rate might be equally likely. A final determinant of uncertainty is model instability. It captures the fact that model parameters might change over time in a partly unpredictable way. For example, it could be the case that μ decreased from 8% several decades ago to 4% more recently.

One might argue that model instability is a proof of model misspecification. After all, it could be an indication for missing variables. We provide two answers. Firstly, it is an implausible assumption anyway that there exists a correct model, i.e. all models are wrong. Some models however are actually useful. A time-varying parameter model is a very flexible and therefore useful model. Secondly, there is no reason why there cannot be any time-varying parameters (latent variables) in a particular parametrization of

---

[1]The term "best" is of course quite general. It depends on the criteria one uses. In chapter 4, we use posterior model probabilities.

the DGP. Features such as regulatory regimes might cause the parameters of the DGP to change over time.

1.2   Outline

Before assessing long-term strategies, chapter 2 first shows how to construct portfolios consisting of individual stock returns for short investment horizons. Here, we ignore parameter uncertainty, since it is well-known (e.g. Barberis (2000)) that parameter uncertainty plays only a minor role at short horizons. The chapter constructs a panel data model that is able to explain the cross-section of expected stock returns incorporating multiple explanatory variables, multiple (short) forecast horizons, industry effects, several alternative factor structures and unobserved heterogeneity at the firm level. The chapter shows that the predictive power increases if multiple firm characteristics are combined. The most important predictor variables are size, dividend-to-price and turnover, followed by earnings revisions and momentum. We obtain portfolios by sorting on expected returns and find that these are well-diversified, only have moderate risk exposures and have high risk-adjusted returns. Longer forecast horizons hardly have an effect on abnormal returns, but strongly reduce portfolio turnover and hence transaction costs.

In chapter 3 we analyze whether the expected potential gains from strategic asset allocation can be realized in an out-of-sample test. We find that risk-averse investors should time the market if they use a shrinkage prior to downplay predictability. It reduces the losses in extreme events and is therefore very valuable for risk averse investors. The inclusion of the hedge component of long-term strategies hardly affects performance due to estimation error, since repeated myopic strategies approximate the true unknown optimal dynamic portfolio equally well as estimated dynamic strategies. We also find that the incorporation of parameter uncertainty leads to a small performance improvement and that portfolio weight restrictions hurt the good models but help the bad models.

Chapter 4 extends the basic strategic asset allocation framework by incorporating model uncertainty in investment decisions. We develop a method to take model uncertainty into account with respect to a series of Vector Autoregressions by using Bayesian Model Averaging techniques. We find that the dividend-to-price ratio and the credit

spread are the most important predictors at short horizons, but that almost all considered predictors play a role at longer horizons by either predicting stock returns directly or by predicting the predictors of stock returns. Stocks mean-revert but are at least as risky at long horizons as at short horizons due to the model and parameter uncertainty effect. Stock returns are even riskier in the long-run in crisis periods due to model uncertainty, since the different models strongly disagree about future stock returns in such periods. The incorporation of model uncertainty has a large impact on the asset allocations of long-term investors, since its inclusion lowers the mean, increases the variance, leads to more negative skewness and increases the kurtosis of future stock returns. Investors with horizons of 20 years or more should invest as much in the stock market as short-term investors. Again, these effects are especially strong in crisis periods.

In chapter 5, we analyze the effect of model instability on long-term investor decisions by developing a specification that is able to handle time-varying intercepts and slopes, time-varying error volatility and correlation, the leverage effect and fat-tailed error distributions. We find that the persistence of time-varying parameters plays a large role. The time-variation in the mean equation (intercepts and slopes) is not persistent enough to have an effect on long-term investors, while the time-variation in the error covariance matrix (volatility and correlation) is very persistent and very relevant for long-term investors. The fat tails of the error distributions disappear once investors take time-varying volatility into account. In this time-varying setting, stocks can either be riskier/safer and stock allocations can either be higher/lower in the long-run than in the short-run. We also find that random walk or pooled specifications for the time-varying parameters lead to a large overestimation of stock market risk and a large underinvestment in stocks.

Finally, chapter 6 provides the conclusion and some directions for future research.

2

Firm Characteristics, Industry, Horizon and Time Effects, in the Cross-Section of Expected Stock Returns[1]

We construct a panel data model to explain the cross-section of individual stock returns, using monthly data for 1,880 large US firms for 1985–2005. Model specification is geared towards multiple explanatory variables, poolability across industries, multiple forecast horizons, alternative factor structures and the effects of unobserved heterogeneity among firms. We find that combining multiple firm characteristics increases the predictive power. High expected returns are mostly related to size, dividend-to-price and turnover, and somewhat to earnings revisions and momentum. Diversified portfolios sorted on expected returns have moderate risk exposures and generate significant risk-adjusted returns over all horizons. Longer forecasting horizons drastically reduce portfolio turnover and hence lower costs.

2.1   Introduction

A huge body of empirical research has found that various firm characteristics help to explain the cross section of stock returns. Prominent explanatory variables are size,

---

[1]This chapter is based on Bauer, Diris, Pavlov, and Schotman (2011).

2. FIRM CHARACTERISTICS, INDUSTRY, HORIZON AND TIME EFFECTS, IN THE CROSS-SECTION OF EXPECTED STOCK RETURNS

valuation ratios, momentum, liquidity and industry classification.[1] Most empirical studies have considered one or two predictors in isolation. The typical statistical procedure for documenting return predictability starts with the construction of portfolios. Stocks are first sorted according to a particular firm characteristic, and then allocated to portfolios. If the average returns of the portfolios are significantly different, the characteristic has predictive power. With multiple characteristics the stocks are sorted along different dimensions. The best known two-dimensional sort is the Fama and French (1995) procedure, which sorts portfolios with respect to five size and five book-to-market categories.[2] With only one or two characteristics this methodology is simple and statistically powerful.

Much less is known about the combined effect of multiple characteristics. When the number of explanatory variables grows, the portfolio formation methodology is bound to become problematic, since the number of portfolios grows exponentially with the number of characteristics. With ten different firm characteristics and just two categories per characteristic, we would already need 2^{10} different portfolios. Adding the industry dimension multiplies the number of portfolios even further.

An alternative approach is to use the full cross-section of individual stocks in a multivariate context. Although firm characteristics are correlated and sometimes interact, these instruments clearly contribute to the overall cross-sectional prediction of stock returns. Haugen and Baker (1996), Brennan, Chordia, and Subrahmanyam (1998) and more recently Fama and French (2008) and Ang, Liu, and Schwarz (2010) are among the few multivariate studies in which a large set of predictive variables is analyzed. Instead of sorting stocks in portfolios according to a particular firm characteristic, they work with cross-sectional regressions in a panel of individual stock returns.

In this paper we extend this literature in several directions. Our first extension deals with the interaction between industry effects and firm characteristics. Fama and

---

[1]The literature is so large that it will be impossible to cite more than a few books and empirical studies. Some book references are Bodie et al. (2007, ch. 12, 13), Cochrane (2005, ch. 20) and Campbell, Lo, and MacKinlay (1997). Important empirical studies include De Bondt and Thaler (1985), Fama and French (1992, 1996, 1997), Daniel and Titman (1997), Davis, Fama, and French (2000), Jegadeesh and Titman (1993, 2001).

[2]The returns of these portfolios are used in many empirical studies. Some examples are Fama and French (1996), Hodrick and Zhang (2001) and Campbell and Vuolteenaho (2004)

French (1997) show that only a part of the cross-sectional return differences can be attributed to risk factors. Interaction between firm characteristics and industries occurs in various forms. For example, Moskowitz and Grinblatt (1999) and Lewellen (2004) study the interaction between industries and the momentum effect. Both report evidence that industry portfolios exhibit strong momentum. In our analysis, we employ a methodology in which we allow for the inclusion of a set of firm characteristics and industry-specific intercepts and time effects, as well as industry-specific slope coefficients. This setup enables us to shed new light on the relevance of industries in the cross-section of expected stock returns.

As a second extension, we consider the return predictability over longer horizons. In the cross-sectional regressions of Haugen and Baker (1996), Avramov and Chordia (2006) and Brennan, Chordia, and Subrahmanyam (1998) the dependent variable is without exception the monthly (excess) return. Instead, we consider the cumulative returns over one, three and six months as dependent variables. Longer holding periods are common in many studies using the portfolio sorting methodology, but do not seem to have been considered in panel regression models. If we can predict firm returns over longer periods, given a firm's current characteristics, portfolios sorted on expected returns will be much more stable in terms of turnover. Eventually, this will improve the net returns of managed portfolios in practice.

Thirdly, we introduce firm-level individual effects as a diagnostic check. The inclusion of individual effects implies even more heterogeneity than adding industry effects. In a specification with individual effects, each stock has its own unconditional expected return, irrespective of its average characteristics. Adding the individual effects is interesting for two reasons. First, the amount of unobserved cross-sectional heterogeneity in average stock returns is a measure of the fit of the cross-section of expected returns. Important unobserved heterogeneity at the level of individual firms is an indication for missing predictive variables. Second, if the individual effects are correlated with one or more firm characteristics, this will affect estimates of the slope parameters in the panel. In our panel structure, we formally test whether firm characteristics have significantly different effects in models with and without individual effects.

Finally, in order to model the cross-sectional dependence between stock returns, we estimate several factor models to assess the robustness of our results. We consider basic specifications that either correct for one common factor in total or one common

factor per industry, but we also look at more sophisticated specifications that correct for multiple observed and unobserved factors.

Our predictive variables (size, value, momentum and turnover) have featured in many previous studies. These variables have been subjected to predictability tests for different countries and sample periods. As such, it is not surprising that many show up statistically significant in our panel regressions. More interesting is the result that almost all characteristics interact with industry effects, meaning that their effect is heterogeneous across industries.

When we sort stocks on the predicted returns, we find the largest dispersion in average returns for models with heterogeneous slope coefficients. Portfolios of stocks in the bottom 30% have returns that are on average about 1.8% per month below the return of portfolios of stocks in the top 30%. Moreover, we find that the characteristics in these models have persistent predictive power. Extending the horizon to six months hardly has a negative effect on the predictive power, but stabilizes the composition of the portfolios considerably. At the six months horizon, more than 95% of the stocks in both the bottom and top 30% remain in that portfolio from one month to the other. Our results are robust to the different factor specifications.

The remainder of the paper is organized as follows. Section 2.2 discusses the specification, estimation and testing of the panel model. Section 2.3 describes the data. Section 2.4 reports estimation results for the different panel specifications. Section 2.5 explores the portfolio implications of the models. We sort portfolios on expected returns and test if standard asset pricing models can explain the average returns. In section 2.6 we discuss the impact of individual effects. Next, section 2.7 looks at alternative factor models as a robustness check. Finally, Section 2.8 concludes.

2.2   Panel Model

We consider panel regressions for returns over holding periods from one to six months. For these regressions the dependent variable $R_{i,t+J}^{(J)}$ is the cumulative return over a period of J months, following month t,

$$1 + R_{i,t+J}^{(J)} = \prod_{j=1}^{J}(1 + R_{i,t+j}), \qquad (2.1)$$

with R_{it} the single period return for firm i. Our interest is in explaining the cross sectional variation of the returns $R_{i,t+J}^{(J)}$ of individual stocks using K firm characteristics in a vector $x_{it}^{(J)}$ observed in month t. The characteristics $x_{it}^{(J)}$ are the same for each J in the empirical application.

2.2.1 Specification

The basic model is the panel

$$R_{i,t+J}^{(J)} = \beta_J' x_{it}^{(J)} + v_{i,t+J}^{(J)}, \ i = 1, ..., N, \ t = 1, ...T, \tag{2.2}$$

where β_J is a K-vector of parameters and $v_{i,t+J}^{(J)}$ is an error term, the properties of which we will discuss in the next subsection. We consider four types of firm characteristics: *size*, measured as the logarithm of market value; various *valuation ratios* like earnings-to-price and book-to-price; *momentum*, measured as various functions of past returns; and *turnover*, measured as the logarithm of trading volume in previous months. In addition we use dummies for the industry classification of each firm and also consider interactions between the industry dummies and the other characteristics.

Panels of individual stock returns are inherently unbalanced. In each period t complete data for returns and characteristics are observed for N_t firms. Return data are observed for T months. A total of N different firms are observed. The total number of data points is $n = \sum_t N_t$. In our application the cross-sectional dimension N_t ranges between 238 and 1185 companies, while $T = 248$ months. We make the usual assumption that observations are missing at random and not triggered by events related to returns themselves. Missing data are indicated by the dummy variable I_{it}, which is equal to one if firm i has complete data for period t, and is equal to zero otherwise. Complete data for the entire J-month horizon are observed if

$$I_{i,t+J}^{(J)} = \prod_{j=1}^{J} I_{i,t+j} = 1.$$

For ease of notation we will generally suppress the missing value dummy $I_{i,t+J}^{(J)}$ and the J index for the horizon to write $y_{it} = I_{i,t+J}^{(J)} R_{i,t+J}^{(J)}$ and thus $y_{it} = 0$ for missing data. Similarly we use β instead of β_J and define $v_{it} = I_{i,t+J}^{(J)} v_{it,J}$, $x_{it} = I_{i,t+J}^{(J)} x_{it}^{(J)}$ to include

the missing data.[1] We write

$$y_{it} = \beta' x_{it} + v_{it}. \tag{2.3}$$

Even though the firm characteristics remain the same for all horizons, different values of J give rise to different dependent variables y_{it}, different parameters β and different errors v_{it}. When returns are measured over a horizon longer than the sampling interval ($J = 3, 6$), the panel regression uses overlapping data. We will take the resulting autocorrelation in the errors into account in computing standard errors and test statistics.[2]

Model (2.3) has been specified without reference to an asset pricing model. Any predictable component $\beta' x_{it}$ can, however, be consistent with an asset pricing model. As an example, suppose that excess returns are generated by the factor model

$$y_{it} \;\; = \delta'_{it} f_t + e_{it}, \tag{2.4}$$

$$f_t \;\;\;\; = \pi + u_t, \tag{2.5}$$

with f_t a vector of unobserved common factors, δ_{it} the factor risk loadings, e_{it} idiosyncratic shocks with variance σ^2 unrelated to f_t, π the factor risk premia and u_t the factor risk. All that is needed to obtain the link with asset pricing is a relation between factor loadings and firm characteristics. For example, assume the following relation between δ_{it} and x_{it}

$$\delta_{it} = A x_{it}. \tag{2.6}$$

Substituting (2.5) and (2.6) in (2.4) gives

$$y_{it} = x'_{it} A' \pi + x'_{it} A' u_t + e_{it} \tag{2.7}$$

which has the same form as (2.3) with $\beta = A'\pi$ and $v_{it} = x'_{it} A' u_t + e_{it}$.

---

[1]Here, y_{it} and x_{it} share the same time index. Note however that there is no simultaneity issue, since the firm characteristics are known at least one period in advance (refer to 2.2).

[2]Here, we use direct forecasts, i.e. we regress $R_{i,t+J}^{(J)}$ directly on $x_{it}^{(J)}$. Alternatively, we could set-up an auxiliary model for the conditioning variables and iterate the one-period forecasts forward as we do in the next three chapters. However, although Marcellino, Stock, and Watson (2006) find that direct forecasts perform poorly at forecast horizons of several years (also refer to chapter 4), they find that direct forecasts work well at short forecast horizons such as in our setting. Therefore, we only consider direct forecasts in this chapter.

Consequently, β can be interpreted as a function of the factor risk premia. Without further restrictions on the factors f_t or factor loadings δ_{it}, it is not possible to separately identify risk premia $\delta'_{it}\pi = x'_{it}A'\pi$ or cross-sectional mispricing.

Our methodology is to construct portfolios based on (2.3) and subsequently test how well standard asset pricing models can explain the returns. In the first stage the panel (2.3) is estimated and stocks are sorted into portfolios with maximally different predicted returns according to the fitted values $\hat{\beta}'x_{it}$. This procedure is closest to the Fama-French type portfolio sorting methodology, where stocks are sorted on one or more characteristics like size, value, momentum and liquidity. In a second stage the focus will be on exposures to risk factors and evaluation of abnormal returns.

2.2.2 Cross-sectional dependence

There is strong cross-sectional dependence between individual stock returns. A first method to correct for this dependence is by using a simple common time fixed effect,

$$v_{it} = \lambda_t + e_{it}, \tag{2.8}$$

where the time effects are fully unrestricted parameters. One way to interpret these effects is that λ_t is an unobserved common factor against which all stocks have a beta equal to one. The time effects are a crude way to adjust for systematic risk. Since not all beta's are equal to one, and returns are generated by multiple factors, considerable cross-sectional covariance among the returns will remain. Still, time effects take out a large common noise component from the returns, and thus reduce the cross-sectional correlation of the errors. This improves the efficiency of the estimator.[1]

The cross-sectional effects of firm characteristics are our main focus. Time series effects are eliminated by the fixed time effect λ_t, which implies that all data are taken in deviation of the cross-sectional average. Being an unrestricted parameter the time effect λ_t accounts for a possible aggregate effect on all stocks, for example the effect of a historically low *EP* ratio on all returns. As a common factor the time effect cancels in ranking stocks on their predicted returns, since return differences $y_{it} - y_{jt}$ do not depend on λ_t.

---

[1]The predictor variables are persistent. By taking the factor structure in the error terms into account, we remove common components from the predictor variables and reduce therefore the persistency of the predictor variables.

2. FIRM CHARACTERISTICS, INDUSTRY, HORIZON AND TIME EFFECTS, IN THE CROSS-SECTION OF EXPECTED STOCK RETURNS

The second method we use is to perform a risk adjustment prior to constructing portfolios. In that case the panel specification becomes

$$v_{it} = \delta_i' f_t + e_{it}, \tag{2.9}$$

where δ_i are constant factor loadings and f_t is a set of M common risk factors. The factors are measured over the same period as the returns y_{it}. We consider two specifications for the common factors. In the first specification the vector f_t is observed and consists of the three Fama-French factors plus a momentum factor.[1] In the second specification the vector of observed common factors is augmented by unobserved common factors. We use the fitted value $\hat{\beta}' x_{it}$ to rank stocks into portfolios. An alternative method would have been to predict and time the factors. We do not pursue this alternative, because (i) we want to focus on the predictive power of firm characteristics and industry-effects and (ii) it would make a transparent comparison between both methods difficult.

There are two important differences relative to specification (2.8). First, the common factor components $\delta_i' f_t$ can explain a large part of the cross sectional correlation in the error terms v_{it} in (2.3). Second, asset pricing theories imply that the δ_i should explain most of the cross-sectional dispersion in expected returns. If the factors are returns, the term $\beta' x_{it}$ in (2.9) can be interpreted directly as abnormal returns, and directly match the object of interest. A test of the null hypothesis $\beta = 0$ is then a test of asset pricing restrictions. After correcting for the common factors, the remaining cross-sectional correlation in e_{it} will be very limited and therefore leads to more efficient estimates. This alternative approach is pursued in Brennan, Chordia, and Subrahmanyam (1998), Avramov and Chordia (2006) and Ang, Liu, and Schwarz (2010).

The approach, however, also brings significant costs. Interpretation of β relies on the specification of the common factors and the factor loadings. Factor loadings vary over time with macro-economic variables and cross-sectionally with some of the

---

[1]Recent results in Kleibergen (2010) cast some doubt on the added value of momentum. He finds that there is no factor structure left in the residuals once the three Fama-French factors are taken into account. However, this depends on the particular data-set at hand (particularly on the way stocks are sorted in portfolios). In the empirical section 2.5 we find that momentum has additional explanatory power (besides the three Fama-French factors) as a factor for some of our sorted portfolios.

same characteristics x_{it} as used for the expected returns.[1] Choice of factors can range from only the market portfolio, the three Fama-French factors and other factors like momentum, liquidity and volatility. The approach also requires estimates of conditional factor loadings of individual firms, which is a major challenge by itself.[2] Tests of the asset pricing implications are thus always subject to the joint hypothesis problem.

2.2.3 Industry effects

Define industry dummies $D_{i\ell}$ that take the value one if firm i belongs to industry ℓ ($\ell = 1, \ldots, L$). Industry effects are introduced in three ways. First, we add them to the characteristics and augment the panel model to

$$y_{it} = \sum_{\ell=1}^{L} D_{i\ell}\tau_\ell + x_{it}'\beta + v_{it}, \tag{2.10}$$

where τ_ℓ is a fixed industry specific effect. Industry-specific intercepts are less restrictive than a single pooled intercept, and yet allow for considerable cross-sectional heterogeneity.

Second, we introduce a vector of L industry-specific time effects $\lambda_{\ell t}$ instead of the single time effect λ_t. This generalizes (2.3) to

$$y_{it} = \sum_{\ell=1}^{L} D_{i\ell}\lambda_{\ell t} + \beta' x_{it}\beta + e_{it}. \tag{2.11}$$

We do not augment the second factor specification in subsection (2.2.2) with industry-specific time effects in order to avoid overfitting due to an abundance of factors.

As for the single time effect, we will assume that all industry-specific $\lambda_{\ell t}$ are unrestricted parameters. They can be interpreted as industry risk factors. A direct consequence is that cross-sectional predictions will be made only within the same industry. For firms i and j that belong to the same industry ℓ the relative return $y_{it} - y_{jt}$ does not involve the industry time effects. For firms in separate industries the industry

---

[1] The results in Kleibergen (2010) also cast some doubt on the asymptotics one should use when macro-economic factors are incorporated.

[2] In order to reduce the total number of parameters, we do not allow for time-varying factor loadings. Ang and Kristensen (2009), Christoffersen, Jacobs, and Vainberg (2008), Ghysels and Jacquier (2007), and Jacquier, Titman, and Yaçın (2010) are part of a growing research effort aimed at improving the measurement of individual beta's that may be varying over time. Dealing with the measurement error in factor loading estimates is an issue that is far from being settled.

time-effect will not disappear and would require a separate time series model for $\lambda_{\ell t}$. To avoid this in the portfolio sorts, we first sort firms into industries, and then use firm characteristics to find the best stocks within each industry. Trading strategies with industry-specific time effects thus imply picking the best stocks within each industry in every period.

Industry-specific time effects can change the estimates of some of the slope parameters in β. A typical example is the Moskowitz and Grinblatt (1999) hypothesis that momentum is actually an industry effect. They find that momentum does not help predict the relative returns of individual firms, but rather the relative performance of entire industries. If the hypothesis of Moskowitz and Grinblatt (1999) is correct, and we estimate the panel with industry-specific time effects, we should expect that the momentum parameters will become smaller and less significant. Otherwise we would be able to predict the relative returns within the same industry using individual firm momentum.

The third way of accounting for industry effects is by allowing separate slope parameters β_ℓ for each industry. We extend both factor specifications by these separate slope parameters. If we allow for industry-specific intercepts, slopes and time-effects, we get the following model

$$y_{it} = \sum_{\ell=1}^{L} D_{i\ell} \left(\lambda_{\ell t} + \beta_\ell' x_{it} \right) + e_{it}. \tag{2.12}$$

In this model industries are completely separated and firm characteristics are only valuable for within industry prediction. Without any pooling on either $\lambda_{\ell t}$ or β_ℓ we have L separate panel data models.

2.2.4 Estimation of β

The following two subsections explain the estimation of β and its standard errors in respectively the standard panel model and in the augmented factor model.

2.2.4.1 Standard panel model

Let $\tilde{y}_{it}$ and $\tilde{x}_{it}$ be the transformed data after partialling out the time and firm dummies, and collect all elements at time t in the N-vector $\tilde{y}_t$ and the $(N \times K)$ matrix $\tilde{X}_t$. For an unbalanced panel the transformation is somewhat more complicated than the usual

correction for the cross-sectional and time series averages. Exact formulas are derived in Wansbeek and Kapteyn (1989). Estimation of the alternative specification with industry time dummies $\lambda_{\ell t}$ involves a minor adjustment in the data transformation, where cross-sectional averages must be replaced by industry-specific averages. After the transformation the pooled OLS estimator of β is

$$\hat{\beta} = \left(\sum_{t=1}^{T} \tilde{X}_t' \tilde{X}_t \right)^{-1} \left(\sum_{t=1}^{T} \tilde{X}_t' \tilde{y}_t \right). \tag{2.13}$$

Our panel contains lagged returns as momentum variables. It is well-known that lagged dependent variables cause biases in dynamic panel data models with individual effects μ_i. In case of cross-section dependence, Phillips and Sul (2007) show that the bias is random and hence could potentially be large. The bias disappears when T is large, as we assume.[1]

We use a robust estimator of the covariance matrix of $\hat{\beta}$ proposed by Driscoll and Kraay (1998) that is valid for arbitrary cross-sectional correlation and heteroscedasticity to calculate standard errors and test statistics. The basic set-up of Driscoll and Kraay (1998) is quite general, but does not allow for individual effects. Goncalves (2010) shows that the Driscoll and Kraay (1998) covariance estimator is also valid in settings with individual effects allowing for any level of remaining cross-sectional dependence (including strong dependence, i.e. an error factor structure is for example allowed). Vogelsang (2008) extends this result and shows that the Driscoll and Kraay (1998) covariance matrix for $\hat{\beta}$ is valid even when time fixed effects and/or more general time trends are included.[2]

As in Driscoll and Kraay (1998) define the time series of K-vectors h_t as

$$h_t = \frac{1}{\sqrt{N_t}} \tilde{X}_t' e_t, \tag{2.14}$$

where e_t is the N-vector of errors, i.e. including the zeros for the $N - N_t$ missing data at time t. The cross-sectional sum is scaled by N_t to account for the unbalancedness

---

[1]In this chapter, we use a frequentist perspective. It allows us to be robust to unknown forms of cross-sectional correlation. We could have approached the problem as well in a Bayesian way. This would however complicate inference in this particular case, because it would have required us to take a stance on the form (or different forms) of cross-sectional correlation in the error terms. Therefore, we do not pursue this alternative here.

[2]Note that Vogelsang (2008) derives results for the more general fixed-b asymptotics. Standard asymptotics can be obtained by letting b go to 0 in his setting.

2. FIRM CHARACTERISTICS, INDUSTRY, HORIZON AND TIME EFFECTS, IN THE CROSS-SECTION OF EXPECTED STOCK RETURNS

of the panel. We estimate the covariance matrix of $\hat{h}_t$ using the Newey-West weights on the first m autocorrelations,

$$\hat{S} = \frac{\bar{N}}{T} \sum_{s=-m}^{m} \left(1 - \frac{|s|}{m+1} \right) \sum_{t=1}^{T} \hat{h}_{t-s} \hat{h}_t', \qquad (2.15)$$

where $\hat{h}_t$ uses the estimated residuals $\hat{e}_t$ and $\bar{N}$ is the average number of firms in the sample. The autocorrelation lag (m) in the Newey-West estimator depends on the forecast horizon of the model and is as big as the number of months over which the cumulative returns y_{it} are computed. The final expression for the standard errors is

$$Var\left(\hat{\beta} \right) = \frac{1}{T} V^{-1} \hat{S} V^{-1}, \qquad (2.16)$$

where $V = \frac{1}{T} \sum_t \tilde{X}_t \tilde{X}_t$.

The parameter β is affected by heterogeneity assumptions on the intercepts. Both with individual effects μ_i as well as industry time dummies $\lambda_{\ell t}$ a Hausman test is used to check if β is the same in different specifications. We use a version of the Hausman test similar to Pesaran, Smith, and Im (1996) and Pesaran and Yamagata (2008) to check if estimates of β differ significantly between these models.[1]

The fixed effect estimator of a model with individual effects is denoted $\hat{\beta}_I$. The estimator for the restricted specification without individual effects is denoted $\hat{\beta}_P$. Under the null hypothesis both estimators are consistent, with $\hat{\beta}_P$ likely to be more efficient, as it omits the unnecessary individual effects. Under the alternative, $\hat{\beta}_P$ will be inconsistent. Therefore the difference $\hat{\beta}_I - \hat{\beta}_P$ can tell us if individual effects have an effect on the slope coefficients β.

From the expression of the standard errors in (2.16) we know that we can write the difference between the two estimators as

$$\sqrt{T}(\hat{\beta}_I - \hat{\beta}_P) = \frac{1}{\sqrt{T}} \sum_t (V_{II}^{-1} h_{It} - V_{PP}^{-1} h_{Pt}) = \frac{1}{\sqrt{T}} \sum_t g_t, \qquad (2.17)$$

where V_{II} and V_{PP} are the relevant matrices corresponding to the general V in (2.16), h_{It} and h_{Pt} the relevant time series related to h_t in equation (2.14), and g_t is defined as

$$g_t = V_{II}^{-1} h_{It} - V_{PP}^{-1} h_{Pt}. \qquad (2.18)$$

---

[1]It is well-known (e.g. Pesaran and Yamagata (2008)) that the Hausman test has low power in a setting such at the ours. However, in the empirical analysis, we reject the null hypothesis that both sets of coefficients are equal. Therefore, the power issue does not have an impact on our results.

Having constructed g_t, the covariance matrix of $\sqrt{T}(\hat{\beta}_I - \hat{\beta}_P)$ follows as

$$S_{IP} = \frac{\bar{N}}{T} \sum_{s=-m}^{m} \left(1 - \frac{|s|}{m+1}\right) \sum_{t=1}^{T} \hat{g}_{t-s}\hat{g}_t', \tag{2.19}$$

We use this covariance matrix to compute the Hausman test statistic

$$W_H = T(\hat{\beta}_I - \hat{\beta}_P)' S_{IP}^{-1}(\hat{\beta}_I - \hat{\beta}_P). \tag{2.20}$$

2.2.4.2 Panel with Common Factors

We use the three Fama-French factors plus a momentum factor as observed factors. In order to model the unobserved factors, we make the following assumptions

1. We augment the set of observed factors with the cross-sectional averages of y_{it} or the cross-sectional averages of y_{it} and x_{it} as suggested in Pesaran (2006);

2. We augment the common factors with the principal components of y_{it} as suggested in Kapetanios and Pesaran (2007).

In all specifications the factor loadings δ_i are assumed constant over time. In all cases we estimate β by pooled OLS and use the Driscoll and Kraay (1998) estimator of the previous subsection for the covariance matrix. In this way our standard errors are still robust to any remaining cross-sectional correlation.

Firstly, we consider a specification with observed factors. Define the $(T \times K)$ data matrix X_i as

$$X_i = \begin{pmatrix} I_{i1}x_{i1}' \\ \vdots \\ I_{iT}x_{iT}' \end{pmatrix}, \tag{2.21}$$

where I_{it} is a dummy equal to one if firm i has complete observations at time t and zero otherwise. Similarly define the T-vectors

$$y_i = \begin{pmatrix} I_{i1}y_{i1} \\ \vdots \\ I_{iT}y_{iT} \end{pmatrix}, \tag{2.22}$$

the $(T \times M)$ matrices of common factors

$$F_i = \begin{pmatrix} I_{i1}f_1' \\ \vdots \\ I_{iT}f_T' \end{pmatrix}, \tag{2.23}$$

and the transformation matrices M_{F_i}

$$M_{F_i} = I - F_i(F_i'F_i)^{-1}F_i'. \tag{2.24}$$

For specifications involving common factors, we only consider firms with at least 30 observations in order to make sure that all transformation matrices are properly defined. With this notation the pooled OLS estimator of β follows as

$$\hat{\beta}_F = \left(\sum_i X_i'M_{F_i}X_i\right)^{-1}\left(\sum_i X_i'M_{F_i}y_i\right). \tag{2.25}$$

Next, consider specifications with unobserved factors. With unobserved factors we partition the vector f_t as

$$f_t = \left(\begin{array}{c} f_{obs,t} \\ f_{un,t} \end{array}\right).$$

We consider two different estimation techniques to estimate β in such a model. Pesaran (2006) proposes the Common Correlated Effects Pooled (CCEP) estimator. The idea of this estimator is to "filter the individual-specific regressors by means of cross-section averages such that asymptotically as the cross-section dimension (N) tends to infinity, the differential effects of unobserved factors are eliminated". Define the $(T \times (K+1))$ matrix Z_{wi} as

$$Z_{wi} = \left(\begin{array}{c} I_{i1}Y_{w1}, I_{i1}X_{w1} \\ \vdots \\ I_{iT}Y_{wT}, I_{iT}X_{wT} \end{array}\right),$$

where Y_{wt} and X_{wt} are respectively the cross-sectional averages of the dependent variable and the independent variables at time t. Alternatively, matrix Z_{wi} only contains the cross-sectional average Y_{wt}.

Next define the matrix H_{wi} as

$$H_{wi} = (F_i \ \ Z_{wi})$$

with F_i already defined in (2.23), and the $(T \times T)$ transformation matrix M_{wi}

$$M_{wi} = I - H_{wi}(H_{wi}'H_{wi})^{-1}H_{wi}'.$$

This results in the following pooled OLS estimator which is shown to be consistent in Pesaran (2006)

$$\hat{\beta}_P = \left(\sum_i X_i'M_{wi}X_i\right)^{-1}\left(\sum_i X_i'M_{wi}y_i\right) \tag{2.26}$$

According to assumption (3) in Pesaran (2006) this estimator is only valid when the risk-loadings on the unobserved factors are uncorrelated with the explanatory variables. This can potentially lead to problems since it is likely that firm characteristics influence risk loadings.

The second class of estimators is the Principal Components augmentation approach (PC), explained in Kapetanios and Pesaran (2007). This estimator uses the principal components of y_{it} to approximate the unobserved factors. We use the observed common factors and the principal components to account for the common factor structure of the error terms.

Bai and Ng (2002) explain an estimation method to obtain the principal components in a balanced panel. Define the $N \times T$ matrix Y as

$$Y = \begin{pmatrix} I_{11}y_{11} & \cdots & & \cdots \\ \vdots & \ddots & & \vdots \\ \cdots & & \cdots & I_{NT}y_{NT} \end{pmatrix}. \tag{2.27}$$

Bai (2003) shows that the first p principal components of this matrix are the eigenvectors corresponding to the p largest eigenvalues of matrix $Y'Y$ multiplied by $\sqrt{T}$. As in Kapetanios and Pesaran (2007) we firstly standardize the elements in y_{it}. Greenaway-McGrevy, Han, and Sul (2010) show that these principal components can be used instead of the true (unknown) factors as long as asymptotically $\frac{T}{N} \to 0$ and $\frac{N}{T^3} \to 0$.

Since our panel is unbalanced we have to modify the estimation procedure slightly. We use an iterative algorithm proposed by Stock and Watson (2002) that works in unbalanced panels. This algorithm uses the estimation technique of Bai (2003) but replaces the missing values in Y by estimates and updates these estimates until convergence is reached. We determine the number of principal components p by using the information criterion of Bai and Ng (2002). Define P as the $T \times p$ matrix of principal components and let p'_t be the tth row of P. Subsequently define the $(T \times p)$ matrix Z_{Pi} as

$$Z_{Pi} = \begin{pmatrix} I_{i1}p'_1 \\ \vdots \\ I_{iT}p'_T \end{pmatrix}$$

and H_{pi} as

$$H_{pi} = (F_i \ \ Z_{Pi})$$

This allows us to define the $T \times T$ transformation matrix M_{pi} as

$$M_{Pi} = I - H_{pi}(H'_{pi}H_{pi})^{-1}H'_{pi}.$$

The consistent pooled OLS estimator becomes

$$\hat{\beta}_{PC} = \left(\sum_{i=1}^{N} X'_i M_{Pi} X_i\right)^{-1} \left(\sum_{i=1}^{N} X'_i M_{Pi} y_i\right). \tag{2.28}$$

Pesaran (2006) shows that his estimator is $\sqrt{N}$ consistent (in the general case), whereas Greenaway-McGrevy, Han, and Sul (2010) show that the Principal Components augmentation approach is $\sqrt{NT}$ consistent under the regularity conditions given above. The latter has (even) the same asymptotic distribution as the infeasible estimator that uses the true (unknown) factors. Therefore, we expect that the PC approach is more efficient, both asymptotically and most likely in our sample as well.

We use a similar Hausman test as above to compare estimates in models with different factor structures.

2.3 Data

Our data set is the Morgan Stanley Capital International (MSCI) US data universe. It covers the investable universe for most institutional investors. As such it contains relatively few small cap stocks. We include all US firms explicitly followed by MSCI. Some of them are the constituents of the well-known published MSCI US index. Others are followed by MSCI because of their size or relevance. We include companies in the data set only when investors were able to obtain the information provided by MSCI in real time. The MSCI index covers about 70% of the US stock market capitalization. The sample period ranges from February 1985 until September 2005.

We include eleven regressors that have been widely used over the last fifteen years, have proved to contribute to the prediction of stock returns, and are likely to capture different aspects of a company. The explanatory variables are classified into five groups: size, valuation ratios, momentum, turnover and industries.

Size: Size (MV) is defined as the logarithm of the market capitalization of firm i in month t. The relation between size and stock returns is known since the early 1980's. It is one of the main characteristics on which stocks are sorted into portfolios in Fama and French (1992).

Valuation ratios: We include the ratios book-to-price (BP), earnings-to-price (EP), dividend-to-price (DP), cashflow-to-price (CP) and sales-to-price (SP). Valuation ratios are the second standard characteristic on which stocks are sorted. A few of the many studies analyzing the link between valuation ratios and stock return are Rosenberg, Reid, and Lanstein (1985), Fama and French (1992), Lakonishok, Shleifer, and Vishny (1994), and Daniel and Titman (1997). Cochrane (2005) discusses the use of valuation ratios like EP and DP for the prediction of stock returns. Vuolteenaho (2002) finds that cashflow news influences stock returns.

Momentum: We include two types of momentum variables. Short-term price momentum ($R2$-7) is defined as the cumulative return over the last six months. As common, the variable is lagged by an additional month to avoid any spurious relation between the current month return and the future month return caused by bid-ask spread effects and thin trading. Long-term price momentum is defined as the cumulative return over the six months prior to the last six months ($R7$-12). The second type of momentum is earnings momentum (analyst earnings revisions), denoted by $CFY1$. It reflects the expectation revisions of financial analysts about the next year's earnings of the stock, and is computed as the number of positive revisions minus the number of negative revisions, divided by the total number of revisions. The original source of this data is I/B/E/S. Momentum variables are used in Jegadeesh and Titman (1993) and Rouwenhorst (1998). Chan, Jegadeesh, and Lakonishok (1996) discuss both earnings momentum and price momentum. Frankel and Lee (1998) focus on earnings momentum.

Turnover: We use two turnover variables. The first one (VOL) is the log of monthly turnover volume. The second variable ($52W$) is the log of average turnover volume for the last 52 weeks. Stoll (1978), among others, finds that volume is the most important determinant of the bid-ask spread, while Brennan and Subrahmanyam (1995) find that it is a basic determinant of liquidity. Koski and Michaely (2000) discuss the relation between liquidity and stock prices and returns.[1]

---

[1]Brennan, Chordia, and Subrahmanyam (1998) recommend defining separate liquidity variables for NYSE and NASDAQ stocks, since trading volume is measured differently between NYSE and NASDAQ. On the other hand, the stocks traded at NASDAQ are concentrated in a small number of industries. Since our general model in equation (2.12) includes industry-specific coefficients and time effects, we do not split the turnover variables.

2. FIRM CHARACTERISTICS, INDUSTRY, HORIZON AND TIME EFFECTS, IN THE CROSS-SECTION OF EXPECTED STOCK RETURNS

Industries: Each company belongs to a specific industry.[1] The total number of industries is 22.

The raw data set contains 2913 companies. For the econometric analysis we delete all data points that contain incomplete or missing data. This reduces the data set to 1,880 companies and 153,380 data points. Table 2.1 reports the number of companies per industry in the data set. Some industries contain only a few firms, indicating that we should be careful in interpreting their industry-specific parameters.[2]

Fama and French (2008) warn for two potential drawbacks in cross-sectional regressions: the dominance of very small stocks of which there are so many, and the influence of extreme outliers. We avoid the first problem by using the MSCI database, which does not contain many microcaps. Since some firm characteristics, especially the valuation ratios, have extreme outliers, we trimmed all valuation ratio outliers to the lower and upper 1% tail of the distribution. Descriptive statistics of the final data set are reported in Table 2.2.

There are two sources of multicollinearity related to the valuation ratios. First, their numerators contain accounting information and are updated only quarterly. The denominator is the stock market capitalization and is the same for all ratios. The monthly change in the valuation ratios could be mostly due to price changes, and therefore might be correlated with short-term momentum. Yet the degree of multicollinearity seems to be limited. The strongest correlation between short-term momentum and a valuation ratio is -0.25. The maximum correlation among the five valuation ratios is 0.49. As we are mostly interested in the joint effect of the predictors, we keep all five ratios as regressors.

---

[1]We use the MSCI industry classification that was used before April 1999. In April 1999 MSCI and S&P 500 introduced the Global Industry Classification System (GICS). Using the new industry classification in all periods would result in a look ahead bias, while using it only after April 1999 would lead to unreliable results due to the short time series April 1999 - September 2005.

[2]These industries are Power Producers, Data Processing and Computer Services. The low number of firms in the last two industries can be explained by the high number of firms in the industry Technology Hardware.

Table 2.1: Summary Statistics by Industry

The summary statistics are based on 1,880 US firms observed over 248 months from February 1985 until September 2005. The number of firms per industry is the sample size remaining after deletion of incomplete data points. Average returns ($\bar{R}$) and standard deviations ($s(R)$), measured in percentage points per month, are for equally weighted (EW) and value weighted (VW) industry portfolios.

	Industry	firms	Data points	EW port. $\bar{R}$	EW port. $s(R)$	VW port. $\bar{R}$	VW port. $s(R)$
1	Basic Materials	109	12357	1.24	5.71	1.14	5.86
2	Automobiles	31	3592	1.25	7.05	1.00	7.16
3	Consumer	74	7829	1.29	5.44	1.18	5.37
4	Retail	140	10933	1.55	6.81	1.51	6.55
5	Commercial	57	2828	1.23	7.16	0.70	6.66
6	Food and Consumer	107	11140	1.65	4.64	1.40	4.70
7	Specialty	13	1797	1.30	5.54	1.32	5.93
8	Services	42	3682	1.55	5.84	1.37	5.39
9	Health Care	203	12370	1.78	6.16	1.54	5.08
10	Oil and Gas	87	8040	1.64	7.05	1.43	4.91
11	Banking and Insurance	204	15064	1.77	5.57	1.47	5.79
12	Diversified Financials	135	7043	1.37	5.04	1.50	5.91
13	Capital Goods	66	6245	1.41	6.24	1.33	5.59
14	Machinery-Diversified	85	7615	1.53	6.09	1.24	5.75
15	Technology Hardware	331	21526	1.53	9.88	1.12	7.88
16	Semiconductors	16	1353	1.97	17.02	2.07	15.89
17	Computer Services	15	1120	1.90	9.22	1.44	8.91
18	Data Processing	11	1014	2.01	6.82	1.60	6.25
19	Telecom	39	3300	1.44	8.28	1.15	5.85
20	Utilities	72	9634	1.21	4.47	1.13	4.57
21	Power Producers	4	364	0.87	13.76	0.97	13.91
22	Transport	39	4534	1.39	6.38	1.21	5.67

2. FIRM CHARACTERISTICS, INDUSTRY, HORIZON AND TIME EFFECTS, IN THE CROSS-SECTION OF EXPECTED STOCK RETURNS

Table 2.2: Summary Statistics of All Firm Characteristics

The table reports descriptive statistics for the set of complete data points, and after all valuation ratios were trimmed to the lower and upper 1% tail of the distribution. A data point is considered to be complete if all variables are available for that particular data point. The data contain 1,880 US firms observed over 248 months from February 1985 until September 2005. Variables are monthly return (RET), log of the market capitalization (MV), book-to-price (BP), cashflow-to-price (CP), dividend-to-price (DP), earnings-to-price (EP), sales-to-price (SP), analyst earnings revisions ($CFY1$), short-term momentum ($R2$-7), long-term momentum ($R7$-12), log of the monthly volume (VOL) and log of the average volume over the last 52 weeks ($52W$). Returns are measured in percentage points. The variables $R2$-7 and $R7$-12 are cumulative six-month returns.

Variable	Avg.	Std. dev.	Min	1st perc.	Median	99th perc.	Max
RET	1.46	13.76	-92.10	-34.50	1.16	41.93	640.74
MV	7.85	1.53	1.30	4.64	7.86	11.61	13.31
BP	0.51	0.36	-0.13	-0.13	0.44	2.00	2.00
CP	0.12	0.12	-0.26	-0.26	0.09	0.65	0.65
DP	0.02	0.02	0.00	0.00	0.01	0.09	0.09
EP	0.02	0.12	-0.75	-0.75	0.05	0.18	0.18
SP	1.29	1.52	0.03	0.03	0.80	9.17	9.17
$CFY1$	-0.06	0.74	-1.00	-1.00	0.00	1.00	1.00
$R2$-7	7.85	34.06	-96.83	-64.36	6.03	117.33	985.96
$R7$-12	8.52	36.33	-96.43	-63.19	6.25	125.67	1597.39
VOL	16.31	1.48	4.79	12.94	16.27	20.00	22.40
$52W$	13.27	1.43	6.36	10.11	13.21	16.90	18.46

2.4   Results

This section reports the estimation results for the specifications that are considered above.

2.4.1   Univariate sorts

As a benchmark for the predictive power of firm characteristics we first construct portfolios that are sorted on a single characteristic. At the beginning of each month t, we construct a high and a low portfolio based on sorting of the stocks by each character-

Table 2.3: Summary statistics of simple sorted portfolios

The table reports the average returns of high and low portfolios based on sorting of all stocks by each characteristic. We form a long portfolio of the top 30% of the sorted stocks and a short portfolio of the bottom 30% of the sorted stocks. The table reports results for simple high minus low portfolios and for industry neutral portfolios. The latter are first constructed within each industry as described above and then aggregated with weights proportional to the number of firms in the industry. Each month we observe the returns of the high and low portfolios, constructed in the previous month. Entries report the difference of average returns and a t-statistic for testing the equality of the mean returns of the long and short portfolios. The *t*-statistics are adjusted for autocorrelation.

Variable	Simple			Industry neutral	
	Avg	*t*-stat		Avg	*t*-stat
MV	-0.37	-1.30		-0.32	-1.42
BP	0.43	1.65		0.41	2.28
CP	0.35	1.10		0.37	2.03
DP	-0.01	-0.03		0.11	0.74
EP	0.17	0.50		0.32	1.74
SP	0.48	1.58		0.41	1.96
CFY1	0.42	2.57		0.39	3.59
R2-7	0.04	0.11		0.01	0.05
R7-12	0.56	2.13		0.41	2.26
VOL	0.54	3.05		0.39	2.88
52W	0.48	2.67		0.40	2.92

istic and buying the top 30% of the sorted stocks (the high portfolio), while selling the bottom 30% of the sorted stocks (the low portfolio).[1] The next month $(t+1)$ we record the returns of the high and low portfolios.

The resulting returns for each characteristic separately are reported in Table 2.3. When significant, the return differences between high and low portfolios are in the order of 0.5% per month. Although the negative sign for size corroborates the small

---

[1]All portfolios are equally weighted. We considered a variety of weighting schemes, like characteristic based weighting according to x_{it} or value weighting. These produce very similar results.

firm effect, it is not statistically significant in a univariate context. This might be due
to the low number of small firms in the MSCI universe. Surprising is the low predictive
power of all valuation ratios.

Two strong predictive variables are long-term price momentum ($R7$-12) and analyst
earnings revisions $CFY1$. Their effect is significant, while short-term price momentum
($R2$-7) has low predictive power. The turnover variables VOL and $52W$ have positive
and significant effects. Other studies, e.g. Brennan, Chordia, and Subrahmanyam
(1998), find a negative and significant effect in a multivariate setting. We expect that
this is related to our universe of stocks which is geared towards the investable universe
of institutional investors that includes relatively more large cap and liquid stocks and
very few micro-caps.

Industry-neutral sorts produce almost the same average return differences. The
main difference with the full sort is that t-statistics are mostly larger due to increased
diversification. As a result, several of the valuation ratios (BP, CP and SP) become
significant.[1]

2.4.2 Multivariate regressions

Table 2.4 reports the pooled estimates of β for different models. The dependent vari-
ables are one-, three- or six-month returns respectively. The models also differ in the
structure of the intercepts and time effects. The first column for each horizon consid-
ered reports coefficients for a specification with industry-specific intercepts and pooled
time effects; the second column contains industry-specific time effects; the third column
contains industry-specific intercepts and the Fama and French and Momentum factors
as a proxy for time effects. In subsection 2.7 we reports results for specifications that
additionally correct for unobserved factors.

Multivariate regression results are very different from those in simple univariate
sorts. Size (MV) is now significant irrespective of forecasting horizon and interaction
with industry effects. Its coefficient is stable across specifications and has the expected
negative sign. In a similar way, the dividend-to-price ratio (DP) becomes significant.
Valuation ratios can be expected to be important for the forecasting of three and six

---

[1] We also investigated the power of monthly firm characteristics to forecast cumulative returns over
three and six months. In general the t-statistics are similar to those reported in Table 2.3.

Table 2.4: Pooled Parameter Estimates

The table reports estimation results for the pooled coefficient model

$$y_{it} = \beta' x_{it} + v_{it}$$

under different assumptions about industry dummies on intercepts and time effects. Each column contains model coefficients and the respective t-statistics in parentheses. For convenient scaling all entries for $R2\text{-}7$ and $R7\text{-}12$ are multiplied by six. Returns are either measured over a one month, three month or six month period. Models are indicated in the header row by the coefficients that are pooled or industry-specific. The specification with (τ_ℓ, λ_t) is a model with industry intercepts and pooled time effects, the model $\lambda_{\ell t}$ contains industry-specific time effects and the model $(\tau_l, FF4)$ contains industry intercepts and the Fama French and Momentum factors as proxy for the time-effects. Absolute t-statistics are reported in parentheses. The standard errors have been computed using the robust estimator of the covariance matrix in (2.16). The Wald statistic is a test statistic against the null hypothesis that the industry-specific intercepts in model (τ_ℓ, λ_t) are all equal. The 5% critical value of the chi-square distribution with 21 degrees of freedom is 32.67. The Hausman statistics in the last row compare estimates of β in models (τ_ℓ, λ_t) and $(\lambda_{\ell t})$ and estimates of β in models (τ_ℓ, λ_t) and $(\tau_\ell, FF4)$. Its 5% critical value is 19.68.

	One Month			Three Months			Six Months		
	τ_ℓ, λ_t	$\lambda_{\ell t}$	$\tau_l, FF4$	τ_ℓ, λ_t	$\lambda_{\ell t}$	$\tau_l, FF4$	τ_ℓ, λ_t	$\lambda_{\ell t}$	$\tau_l, FF4$
MV	-0.66	-0.60	-0.52	-1.91	-1.72	-1.56	-3.77	-3.38	-3.62
	(4.91)	(4.50)	(7.62)	(5.46)	(5.10)	(7.49)	(5.68)	(5.40)	(8.75)
BP	0.55	0.33	0.48	1.36	0.80	1.68	2.45	1.37	2.80
	(1.74)	(1.50)	(2.47)	(2.17)	(1.62)	(3.37)	(3.20)	(1.88)	(2.70)
CP	0.75	0.66	0.73	2.36	2.25	1.83	5.00	5.24	4.09
	(1.23)	(1.21)	(1.51)	(1.31)	(1.40)	(1.47)	(1.13)	(1.34)	(1.45)
DP	8.85	9.07	10.48	25.65	23.39	31.42	48.79	42.35	61.65
	(2.22)	(2.74)	(2.98)	(2.48)	(2.69)	(2.98)	(2.25)	(2.30)	(2.91)
EP	0.00	0.39	-0.36	-0.48	0.04	-2.58	1.43	1.57	-2.21
	(0.00)	(0.34)	(0.40)	(0.14)	(0.01)	(1.37)	(0.22)	(0.29)	(0.65)
SP	0.07	0.08	0.00	0.18	0.22	-0.01	0.39	0.43	0.15
	(0.94)	(1.30)	(0.04)	(1.05)	(1.29)	(0.09)	(1.38)	(1.45)	(0.57)
CFY1	0.19	0.24	0.08	0.10	0.25	0.10	0.44	0.79	-0.10
	(1.92)	(3.96)	(1.31)	(0.53)	(1.89)	(0.71)	(1.76)	(4.24)	(0.48)
R2-7	0.03	0.02	-0.03	0.18	0.16	-0.04	0.38	0.35	-0.03
	(0.66)	(0.64)	(1.28)	(2.12)	(2.93)	(0.75)	(3.13)	(4.31)	(0.29)
R7-12	0.03	0.03	0.06	-0.01	0.00	0.11	-0.19	-0.16	0.14
	(1.17)	(1.90)	(3.04)	(0.09)	(0.02)	(1.71)	(1.16)	(1.71)	(1.20)
VOL	0.40	0.43	0.35	-0.19	-0.06	-0.31	-0.51	-0.26	-0.52
	(2.15)	(2.81)	(2.23)	(0.56)	(0.22)	(1.03)	(0.87)	(0.60)	(1.51)
52W	0.27	0.20	0.22	2.19	1.89	2.05	4.63	4.04	4.30
	(1.46)	(1.33)	(1.41)	(5.77)	(5.60)	(5.79)	(6.44)	(6.11)	(7.32)
$\tau_\ell = 0$	34.06		46.46	44.42		78.73	6911		244.65
Hausman		20.62	20.51		34.51	29.94		76.11	41.04

month returns because their numerators are announced quarterly. Indeed, the book-to-price variable (BP) increases in significance for three and six month horizons. The other valuation ratios are insignificant in all specifications.

Earnings momentum ($CFY1$) is only significant in models with industry time dummies $\lambda_{\ell t}$. Contrary to the univariate sorts, long term price momentum ($R7$-12) is not significant across horizons; only in the one-month horizon specification for the FF4 model it is statistically significant. Short term price momentum ($R2$-7) now is significant at longer horizons, consistent with the results in Rouwenhorst (1998). Interestingly, all momentum effects are estimated more precisely in models with industry time dummies. Hence, momentum can be interpreted as a significant intra-industry return predictor.

The effect of trading volume is generally significant. The short term volume measure (VOL) has predictive power at the one month horizon, while the trend in volume ($52W$) is significant for the three and six months horizons. However, again all significant estimates have unexpected positive signs.

Industry time effects $\lambda_{\ell t}$ have very little effect on the β estimates in table 2.4. The importance of industry time effects will become important later on when we sort portfolios on expected returns and consider the effects of industry neutral sorts.

The Wald test statistics in the bottom of the table shows that industry-specific intercepts are not equal. Moreover, the Hausman statistic shows that the β estimates differ between the three specifications.

Next, we consider the pooling hypotheses $\beta_\ell = \beta$, i.e. whether firm characteristics have the same effect in all industries. For all models in Table 2.4 we estimate a version with industry-specific coefficients β_ℓ. We formally test whether coefficients are indeed industry-specific or whether these can be pooled.

We test whether each firm characteristic has the same coefficients across industries. Table 2.5 reports the test statistics of the null hypothesis $\beta_{j\ell} = \beta_j$ for each characteristic j separately. The null hypothesis is rejected at the 5% significance level in all model specifications, for all horizons, and for all characteristics with the exception of DP and $R2$-7 at the one month horizon (in two out of three specifications). Interestingly, the heterogeneity across industries is more pronounced for longer forecast horizons as all coefficients vary across industries.

Table 2.5: Industry-Specific Slope Parameters

The table shows Wald-statistics for the null hypothesis

$$H_0 : \beta_\ell = \beta \qquad (\ell = 1, \dots, L)$$

in the model

$$y_{it} = \sum_{\ell=1}^{L} D_{i\ell} \left(\tau_\ell + \beta'_\ell x_{it} \right) + v_{it}$$

under different assumptions about the error term. Columns (τ_ℓ, λ_t) relate to pooled time effects and industry intercepts with $\lambda_{\ell t} = \lambda_t$; the columns $(\lambda_{\ell t})$ refer to industry-specific time effects in which $\lambda_{\ell t}$ are unrestricted; finally the model $(\tau_\ell, FF4)$ consists of industry intercepts and the Fama-French factors and Momentum factor as proxy for the time-effects. The 5% critical value of the chi-square distribution with 21 degrees of freedom is 32.67. The forecasting horizons are one, three and six months.

	One month			Three months			Six months		
	τ_ℓ, λ_t	$\lambda_{\ell t}$	$\tau_\ell, FF4$	τ_ℓ, λ_t	$\lambda_{\ell t}$	$\tau_\ell, FF4$	τ_ℓ, λ_t	$\lambda_{\ell t}$	$\tau_\ell, FF4$
MV	59.29	72.32	63.66	78.86	75.93	85.48	111.07	93.82	105.22
BP	50.86	46.91	56.89	102.41	83.76	92.62	158.61	82.93	175.71
CP	41.57	52.57	63.66	59.27	62.11	89.76	116.38	144.77	196.16
DP	32.48	36.56	27.00	48.39	60.12	65.24	60.87	97.82	101.31
EP	43.21	48.09	51.39	75.97	91.60	93.82	139.94	280.80	273.47
SP	47.02	50.51	68.26	93.16	71.51	134.98	131.63	82.25	149.96
CFY1	61.52	72.30	40.76	52.22	50.97	45.67	79.83	118.90	61.42
R2-7	30.38	20.91	33.74	74.38	61.01	70.36	84.91	126.43	148.55
R7-12	56.69	60.85	66.69	155.70	82.84	137.89	140.48	106.61	233.28
VOL	54.09	49.87	58.94	64.91	70.71	104.45	79.94	79.17	105.93
52W	47.12	46.96	55.20	69.50	70.37	154.53	97.32	89.44	199.24

From these results we conclude that industry-specific coefficients are very important in cross-sectional prediction strategies.

2.5 Portfolio Management Implications

We follow the standard empirical methodology to define a trading strategy with large predicted returns. Subsequently, we check whether the returns of the strategy can be explained by an asset pricing model. To investigate the asset pricing implications of the multivariate models, we consider the time series returns for a number of long-short portfolios. For each model we construct the fitted values

$$\hat{y}_{i,t+J} = \sum_\ell \left(D_{i\ell} \hat{\tau}_\ell + \hat{\beta}'_l x_{it} \right) \tag{2.29}$$

Each period t, the expected returns for the next one, three and six months $\hat{y}_{i,t+J}$ are sorted in a decreasing order. We construct equally weighted portfolios in which we

allocate the top (bottom) 30% of the sorted stocks to a long (short) portfolio.

Portfolios based on models that predict cumulative returns for J months follow the overlapping portfolio methodology advocated by Fama (1998). Each month t, we predict the returns for the following J months and construct a long and a short portfolio as described above. The portfolios are kept for the following J months and are liquidated at the end of month $t + J$. In month $t + 1$ we repeat this procedure and construct new long and short portfolios. These portfolios are liquidated at the end of month $t + J + 1$. Therefore, after the start-up period, the aggregate portfolio consists of J overlapping long-short portfolios.

In models with industry-specific time effects $\lambda_{\ell t}$, we sort stocks separately within each industry and construct industry-specific long and short portfolios. We add all industry-specific long and short portfolios, weighted by the number of stocks in the industry, to obtain the aggregate portfolio. Consequently, the aggregate portfolio is industry neutral. The selection of stocks based on sorted expected returns is performed within industries only.

Let $R_{L,t}$ be the return on the long portfolio in period t, $R_{S,t}$ the return on the short portfolio and let $y_{LS,t} = R_{L,t} - R_{S,t}$ be the return differential. The performance of the portfolios is evaluated using the standard performance attribution regression of managed portfolios,

$$y_{LS,t} = \alpha + \delta'_{t-1} f_t + \epsilon_t, \tag{2.30}$$

where f_t is the set of common factors (market return, *SMB*, *HML*, *UMD*) and δ_{t-1} a vector of conditional or unconditional factor loadings for the portfolio. We report abnormal returns (α) with both unconditional and conditional factor loadings. Following Avramov and Chordia (2006) and Jagannathan and Wang (1996) we use the default spread as conditioning variable. This implies the specification

$$\delta_t = d_0 + d_1 z_t \tag{2.31}$$

with z equal to the default spread.

The exposure to the factors will be different depending on the structure of the errors in the panel specification. If the panel has a single time effect λ_t, the cross-sectional predictions have not been risk-adjusted in the panel model. If the multi-factor asset pricing model with factors f_t would be correct, the expected returns could thus be

Table 2.6: Portfolio Performance

The table shows the performance of the long-short portfolios at different prediction horizons, alternative specifications of the intercepts, and for different risk adjustments. Long-short portfolios are formed by selecting at each t the 30% stocks for which $\sum_{\ell=1}^{L} D_{i\ell}\left(\tau_l + \beta_l' x_{it}\right)$ is largest; short portfolios contain the 30% stocks with the lowest predicted returns $\sum_{\ell=1}^{L} D_{i\ell}\left(\tau_l + \beta_l' x_{it}\right)$. All returns are equally weighted.

Panel A The table reports the intercept α in the performance regressions

$$y_{LS,t} = \alpha + \delta'_{t-1} f_t + u_t$$

where f_t is a set of common risk factors and δ_{t-1} are risk factor loadings that are either constant or conditional. The first row of results ($\delta_t = 0$) refers to the raw average returns. The second row (FF4) includes the Fama-French market ($RMRf$), size (SMB), value (HML) and momentum (UMD) factors with constant factor loadings. In the third row (Cond FF4) the conditional risk loadings are specified as

$$\delta_t = d_0 + d_1 z_t,$$

with z_t equal to the default spread.

Panel B reports the coefficients $\delta_t = d_0$ and t-statistics for the FF4 model.

Models are indicated in the header row by the coefficients that are pooled or industry-specific. Intercepts are always industry specific (τ_ℓ). The columns labelled λ_t have a single time effects and pooled slope coefficients β; the specification with (β_ℓ, λ_t) has industry-specific slopes; the model ($\beta_\ell, \lambda_{\ell t}$) contains industry-specific time effects and slopes; finally the specification (β_ℓ, F) contains industry-specific slopes and the Fama-French and Momentum factors as proxy for the time-effects

	One Month				Three Months				Six Months			
	λ_t	β_ℓ, λ_t	$\beta_\ell, \lambda_{\ell t}$	β_ℓ, F	λ_t	β_ℓ, λ_t	$\beta_\ell, \lambda_{\ell t}$	β_ℓ, F	λ_t	β_ℓ, λ_t	$\beta_\ell, \lambda_{\ell t}$	β_ℓ, F
A: Intercepts (α)												
0	1.56	2.02	1.63	1.84	1.43	1.92	1.48	1.70	1.39	1.80	1.40	1.57
	(6.28)	(10.23)	(10.16)	(7.62)	(5.56)	(9.39)	(9.24)	(7.62)	(6.13)	(9.71)	(9.50)	(7.98)
FF4	1.45	1.93	1.60	2.09	1.30	1.78	1.37	1.92	1.17	1.58	1.28	1.68
	(5.71)	(7.84)	(8.63)	(7.66)	(5.58)	(7.47)	(7.78)	(7.87)	(5.89)	(7.21)	(7.42)	(7.38)
CFF4	1.40	1.87	1.57	2.03	1.24	1.71	1.33	1.85	1.09	1.49	1.23	1.61
	(5.75)	(7.51)	(7.50)	(8.05)	(5.76)	(8.01)	(7.73)	(8.55)	(5.93)	(7.30)	(7.12)	(8.36)
B: Factor loadings FF4 (d_0)												
b_M	0.32	0.24	0.19	0.22	0.30	0.24	0.17	0.19	0.27	0.21	0.16	0.15
	(4.64)	(4.25)	(5.11)	(3.44)	(4.25)	(5.16)	(4.75)	(3.85)	(3.96)	(3.81)	(4.02)	(2.83)
b_s	0.59	0.34	0.34	0.23	0.69	0.42	0.40	0.25	0.58	0.34	0.35	0.33
	(7.77)	(4.75)	(6.39)	(3.12)	(8.56)	(6.37)	(6.80)	(3.72)	(6.57)	(4.85)	(5.98)	(5.64)
b_h	-0.01	0.16	0.03	-0.21	0.06	0.21	0.07	-0.09	0.19	0.31	0.09	-0.03
	(0.12)	(1.57)	(0.45)	(2.17)	(0.54)	(1.68)	(0.94)	(0.80)	(1.21)	(1.85)	(0.95)	(0.33)
b_u	-0.15	-0.16	-0.14	0.38	-0.15	-0.14	-0.07	-0.39	-0.05	-0.03	-0.03	-0.26
	(1.63)	(1.76)	(1.94)	(3.85)	(1.67)	(1.53)	(1.17)	(4.35)	(0.55)	(0.32)	(0.55)	(2.97)
R^2	0.51	0.27	0.42	0.42	0.52	0.31	0.42	0.43	0.42	0.22	0.38	0.38

2. FIRM CHARACTERISTICS, INDUSTRY, HORIZON AND TIME EFFECTS, IN THE CROSS-SECTION OF EXPECTED STOCK RETURNS

strongly related to risk exposures and we would see significant elements in δ, while α would be close to zero.

Table 2.6 reports average returns (α) of long-short portfolios at different prediction horizons, alternative model specifications and risk-adjustment procedures. All models have industry-specific intercepts (τ_ℓ). The first model in the first column of Table 2.6 is specified with a single time effect (λ_t) and pooled slope coefficients β. The second model additionally has industry specific slopes (β_ℓ). The third specification contains industry specific time effects ($\lambda_{\ell t}$) and slopes (β_ℓ) and the fourth has industry-specific slopes and the Fama-French and Momentum factors as proxies for the time effects. For all combinations of model specifications and forecast horizons, α is considerably larger than 1% per month. Risk adjustment hardly affects the average returns of the portfolios. Analysis of the exposures to the common factors reveals that the long-short portfolios, irrespective of the model specification and horizon, have significant exposures to the market return and the *SMB* factor, but not to the *HML* and *UMD* factor. Only for the fourth specification we find a significant exposure to *UMD*. Note that industry neutral portfolios also have sizable alpha's of comparable magnitude. The industry neutrality rule in the specification with industry-specific time effects provides a powerful and well-diversified portfolio.

The next step is to analyze the portfolio turnover. The lower the turnover, the lower the transaction costs of a particular trading strategy. Table 2.7 reports the transition frequencies among the long (L), neutral (N) and short (S) portfolios that are constructed using the cross-sectional expected returns from the model specifications in Table 2.6. Again, we provide information for the three forecasting horizons. The left panel reports transition frequencies of stocks going from one portfolio to another in the case of monthly forecasting. Table 2.7 shows that 83% of the stocks that are in the long portfolio based on the first specification in month t, remain there in month $t + 1$. The same holds for stocks in the short portfolio: 83% of the stocks that are in the short portfolio in month t remain there in month $t + 1$. The three other specifications have slightly higher turnover levels, which is most likely related to the increased relevance of industries in the portfolio selection procedure. New (N) and exiting (EX) stocks are equally distributed among the long, neutral and short portfolios.

The middle and right panels of Table 2.7 report the transition frequencies in the case of three and six month forecasting horizons. A striking difference with the monthly

Table 2.7: Portfolio Turnover

The table reports transition frequencies among the long, neutral and short portfolios that are constructed using the cross-sectional expected returns from different models. The long portfolios contain the 30% stocks with the highest expected returns, the short portfolios the 30% with the lowest expected returns, and the neutral portfolios the remaining 40%. All stocks are equally weighted. Transition frequencies for portfolios P and Q are the average percentages of stocks that go from P to Q through time. The portfolio categories "New" and "EX" refer to stocks that were not in the data set in period t and appeared in period $t+1$, and that left the data set in period $t+1$, respectively. The table consists of four panels that show transition frequencies for portfolios based on forecasting for one, three and six months. The panels refer to models with a pooled time-effect and pooled slope β; with a pooled time effect and industry-specific slopes β_ℓ; with industry-specific time-effects $\lambda_{\ell t}$ and slopes β_ℓ; and finally (β_ℓ, F) with industry-specific slope coefficients and the four Fama-French and Momentum factor as proxy for the time-effects.

| | | One month | | | | Three months | | | | Six months | | | |
| | | To | | | | To | | | | To | | | |
Model	From	L	N	S	EX	L	N	S	EX	L	N	S	EX
β, λ_t	Long	83	16	0	1	94	5	0	1	96	2	1	1
	Neutral	12	75	12	1	6	88	5	1	7	86	7	1
	Short	0	16	83	0	0	4	95	0	1	2	96	1
	New	40	39	21	-	39	39	22	-	38	38	24	-
β_ℓ, λ_t	Long	78	19	2	1	92	5	2	1	95	2	2	1
	Neutral	14	69	16	1	8	82	9	1	6	87	7	1
	Short	2	21	76	1	2	5	92	1	2	2	95	1
	New	37	35	27	-	37	35	28	-	36	34	30	-
$\beta_\ell, \lambda_{\ell t}$	Long	75	22	2	1	92	5	2	1	95	2	2	1
	Neutral	18	63	19	1	11	77	12	1	8	82	10	1
	Short	3	22	75	1	2	5	92	1	2	2	95	1
	New	38	35	28	-	37	34	29	-	34	37	29	-
β_ℓ, F	Long	80	18	2	1	92	5	2	0	95	2	2	0
	Neutral	13	70	16	0	8	82	10	0	6	86	8	0
	Short	2	21	77	0	2	6	92	0	2	3	95	0
	New	36	35	29	-	35	34	31	-	37	33	30	-

results is the sharply reduced turnover. The persistence of the long portfolio, for example, increases from 83% to 94% when the forecasting horizon grows from one to three months. Longer forecasting horizons do not deteriorate returns, while simultaneously drastically decreasing portfolio turnover. This result is found for all four specifications.

The turnover results clearly have important implications for portfolio managers employing models that are of the same family of those presented in this paper. Rebalancing strategies based on monthly expected return estimates result in a turnover of 20% to update the long portfolios and 20% to update the short portfolios. This implies a yearly turnover of 500%, which is far beyond the turnover thresholds used in professional portfolio management. A strategy based on a six-month horizon would imply a yearly turnover of just 100% and hence lower trading costs. At the same time, expected returns are hardly influenced.

2.6 Individual effects

Adding individual effects, the basic panel model (2.3) becomes

$$y_{it} = \mu_i + \beta' x_{it} + v_{it}, \tag{2.32}$$

Relative returns of stocks i and j now depend on the difference $\mu_i - \mu_j$, implying that expected returns on stocks i and j differ for some unobserved reason. The cross-sectional variation in μ_i does tell us a lot about the unmodeled systematic cross-sectional variation in the data, and thus about the goodness of fit. When the individual effects μ_i make a significant contribution to the cross-sectional variation of expected returns, there is much scope for improvement of the model. The larger the variance of μ_i, the more scope for improvement.

Individual effects μ_i serve as a diagnostic, but not as a practical means for sorting stocks. In searching for stocks with high expected returns, we would need to take into account the estimates of μ_i for $i = 1, \ldots, N$. These are likely to be poorly estimated, as information on them can only come from the time-series dimension of the data. Firms without a long history will have especially poorly determined individual effects. Furthermore, individual firm returns are very noisy – that is exactly what usually motivates portfolio formation – and the forecasting performance of the model will be negatively affected by the noisy estimates of μ_i.

Table 2.8: Individual Effects

The table reports estimation results for firstly a model with individual effects μ_i, pooled time effects λ_t and pooled slope coefficients β and secondly a model with individual effects, pooled slope coefficients and the four Fama-French and momemtum factors as proxy for the time effects. T-statistics are in parenthesis. The bottom line of the table reports a Hausman test for the hypothesis that all slope coefficients in the models are equal to the slope coefficients in models where the individual effects are replaced by industry dummies. The hypothesis is also tested for each firm characteristic separately with the t-test reported in columns t.

	One Month				Three Months				Six Months			
	(λ_t)	t	(F)	t	(λ_t)	t	(F)	t	(λ_t)	t	(F)	t
MV	-2.90	6.85	-1.24	5.41	-8.30	8.17	-3.63	5.22	-15.81	8.59	-7.42	4.74
	(8.70)		(7.95)		(9.65)		(8.10)		(9.19)		(7.73)	
BP	1.60	3.12	1.47	4.62	4.05	3.33	4.46	4.83	7.42	3.58	8.40	4.93
	(2.65)		(4.15)		(3.22)		(5.61)		(4.24)		(7.02)	
CP	1.52	1.27	1.49	2.22	4.19	1.10	3.02	1.32	7.82	1.13	4.59	0.34
	(1.71)		(2.29)		(1.89)		(1.75)		(1.92)		(1.45)	
DP	-18.57	4.21	-18.60	6.45	-48.99	4.92	-55.41	7.27	-102.15	5.36	-116.21	6.59
	(2.13)		(-3.28)		(2.44)		(-4.25)		(2.80)		(4.40)	
EP	-1.81	2.35	-1.73	3.15	-6.17	2.67	-6.81	3.37	-10.70	2.98	-10.63	4.01
	(1.35)		(-1.88)		(2.27)		(3.56)		(2.27)		(3.09)	
SP	0.07	0.01	0.21	3.13	0.18	0.03	0.60	3.73	0.54	0.40	1.52	4.47
	(0.48)		(2.17)		(0.65)		(2.55)		(1.24)		(3.45)	
CFY1	0.15	1.61	0.07	0.38	-0.04	1.89	-0.12	0.41	0.10	1.85	-0.11	0.10
	(1.64)		(1.12)		(0.26)		(0.76)		(0.48)		(0.44)	
R2-7	0.02	1.80	-0.03	0.39	0.13	1.90	-0.03	0.48	0.27	1.68	0.01	0.97
	(0.42)		(1.25)		(1.88)		(0.62)		(3.04)		(0.06)	
R7-12	0.05	2.36	0.06	2.00	0.04	1.85	0.12	0.97	-0.09	1.45	0.16	0.70
	(2.15)		(3.47)		(0.59)		(1.96)		(0.72)		(1.34)	
VOL	0.33	1.37	0.41	1.89	-0.57	2.54	-0.24	0.75	-1.24	2.55	-0.36	0.92
	(1.65)		(2.48)		(1.83)		(0.74)		(2.85)		(0.88)	
52W	-0.53	6.01	-0.26	4.19	-0.34	5.75	0.28	4.57	-0.72	4.63	0.29	4.49
	(2.37)		(-1.25)		(0.67)		(0.48)		(0.68)		(0.21)	
Ha		432.82		184.35		502.25		169.20		499.71		139.29

Because of the possible interaction between the individual effects and the explanatory variables, we will treat the μ_i's as fixed effects and examine their effect on estimates of the slope coefficients β. From the panel data literature it is known that random effects estimation is inconsistent if μ_i and the time-series averages of the characteristics x_{it} are correlated.

Correlation between the characteristics and μ_i arises for example with the momentum effect. Since momentum is a function of lagged returns of stock i, it will be positively correlated with μ_i. The larger the dispersion in μ_i, the bigger the effect on the momentum coefficients in β, and the more likely it is to wrongly conclude that momentum is significant when instead individual effects should have been included. Conrad and Kaul (1998) and Jegadeesh and Titman (2002) both estimate how much of the momentum profits can be explained by the cross-sectional variation of the un-

conditional expected returns (the cross-sectional variance of μ_i).

Table 2.8 shows that individual effects have a significant impact on the parameter estimates for some of the characteristics. In a specification with individual effects and pooled slope coefficients (first column), the size effect becomes much more pronounced in models with μ_i. The same results can be found in a specification with the Fama-French and Momentum factor as a proxy for the time effect. The positive covariance between MV and μ_i implies that big firms which have been big during the entire sample period, do not perform much worse than small firms that have been small during the entire sample period. Firms that were big for a long time, but have fallen in size however, perform very well. Apparently, the size effect picks up some long-term return reversal. Again, these results are robust to the forecasting horizon.

The dividend-to-price effect completely disappears and even obtains the opposite sign. A similar sign change occurs for the long-term volume characteristic, which becomes significantly negative, as we would expect from the literature on turnover.

Moreover, the Hausman test indicates that the two sets of parameters are indeed significantly different if individual effects μ_i are included. Not surprisingly, the t-statistics for the differences between the individual elements of β are especially large for MV, DP and $52W$. The momentum variables are generally not affected by the inclusion of individual effects. This result is consistent with Jegadeesh and Titman (2002), who also find that cross-sectional differences in expected return cannot explain profits from momentum strategies.

2.7 Alternative factor specifications

Above, we correct for cross-sectional correlation in the error terms by either using pooled time effects λ_t, industry-specific time-effects $\lambda_{\ell,t}$ or observed factors. In this section, we consider specifications that additionally correct for unobserved factors by either adding the cross-sectional average of stock returns, the cross-sectional averages of stock returns and explanatory variables or the principal components of individual stock returns to the three Fama French factors and the momentum factor. Details on the estimation procedures are given in section 5.4.1.[1]

---

[1]We also considered taking the principal components of y_{it} and x_{it} as recommended in Kapetanios and Pesaran (2007). We do not report this specification here, since the outcomes are very similar to the specification that uses cross-sectional averages of both y_{it} and x_{it}.

Table 2.9: Pooled Parameter Estimates: alternative specifications

The table reports estimation results for the pooled coefficient model

$$y_{it} = \sum_{\ell=1}^{L} D_{i\ell}\tau_l + \beta' x_{it} + \delta_i' f_t + \epsilon_{it}$$

using different specifications to correct for the (un)observed factors f_t. Each column contains model coefficients and the respective absolute t-statistics in parentheses. For convenient scaling all entries for *R2-7* and *R7-12* are multiplied by six. Returns are either measured over a one month, three month or six month period. All models contain industry-specific intercepts. The specification with $\overline{Y}$ uses the three Fama-French factors, the momentum factor and the cross-sectional average of stock returns to correct for factors. Model $\overline{YX}$ adds the cross-sectional average of x_{it}. Finally, specification *PY* uses the principal components of individual stock returns instead of the cross-sectional averages. Absolute t-statistics are reported in parentheses. The standard errors have been computed using the robust estimator of the covariance matrix in (2.16).

	One Month			Three Months			Six Months		
	$\overline{Y}$	$\overline{YX}$	*PY*	$\overline{Y}$	$\overline{YX}$	*PY*	$\overline{Y}$	$\overline{YX}$	*PY*
MV	-0.50	-6.46	-0.55	-1.40	-17.07	-1.70	-3.10	-29.31	-3.57
	(7.06)	(15.24)	(9.07)	(6.55)	(13.18)	(11.18)	(7.51)	(11.10)	(10.54)
BP	0.46	4.95	0.26	1.87	13.39	1.69	4.62	19.85	3.72
	(2.46)	(10.83)	(1.42)	(2.79)	(12.92)	(3.46)	(2.86)	(12.31)	(4.21)
CP	0.69	2.71	1.15	2.39	5.00	3.29	5.40	9.67	7.14
	(1.38)	(2.88)	(2.65)	(1.74)	(2.31)	(2.60)	(2.00)	(3.03)	(2.43)
DP	12.60	-25.06	6.07	35.36	-55.17	3.99	56.80	-63.75	-19.91
	(4.01)	(2.98)	(2.04)	(4.00)	(2.92)	(0.50)	(2.82)	(1.86)	(0.98)
EP	-0.32	-0.33	-0.73	-3.36	-3.82	-4.61	-7.96	-9.30	-8.59
	(0.50)	(0.38)	(1.02)	(2.16)	(1.86)	(2.82)	(2.22)	(4.06)	(2.68)
SP	0.04	0.43	0.01	0.23	1.26	0.06	0.67	2.25	0.67
	(0.69)	(3.27)	(0.11)	(1.75)	(3.14)	(0.44)	(2.37)	(2.82)	(2.33)
CFY1	0.14	0.07	0.13	0.05	-0.28	-0.04	0.04	-0.31	-0.06
	(2.52)	(1.29)	(2.46)	(0.41)	(2.90)	(0.36)	(0.18)	(2.66)	(-0.31)
R2-7	-0.02	-0.10	-0.03	-0.05	-0.15	-0.07	-0.14	-0.18	-0.23
	(1.19)	(5.48)	(2.40)	(1.07)	(3.88)	(1.80)	(2.04)	(2.28)	(3.01)
R7-12	0.03	0.07	0.02	0.04	0.12	0.00	0.07	0.10	-0.06
	(2.11)	(4.16)	(1.86)	(0.88)	(2.92)	(0.01)	(0.76)	(1.61)	(0.70)
VOL	0.20	0.24	0.21	-0.35	-0.53	-0.45	-0.90	-0.82	-0.37
	(1.75)	(1.79)	(1.86)	(1.60)	(2.53)	(2.55)	(3.08)	(3.13)	(1.23)
52W	0.37	0.23	0.36	1.93	0.49	2.32	4.25	-0.32	4.28
	(2.85)	(0.98)	(3.08)	(7.74)	(0.91)	(10.50)	(8.41)	(0.35)	(8.55)

Table 2.9 shows the results. We find that y_{it} contains three principal components. In general, the results for the models that use the cross-sectional averages of y_{it} (first column at every horizon) or the principal components of y_{it} (third column) give similar results as above. Size has again a negative coefficient of similar magnitude. DP is again an important predictor. However, it is remarkable that its coefficient is much smaller for the third specification. Another difference is that $52W$ is suddenly also significant for both specifications at the shortest horizon. The similarity of the results suggests that the estimated firm-specific factor loadings have some, but limited, correlation with the characteristics. Overall standard errors of the augmented factor models are smaller than in the benchmark models as a result of the increased fit due to the common factors.

The estimates for the model that uses the cross-sectional averages of both the dependent as well as the explanatory variables (second column) are completely different from all the other models considered. In general, estimated coefficients are much larger than before. The amount of KN additional nuisance parameters seems to produce exceptionally noisy estimates of β. Therefore, we conclude that this specification is misspecified.

Finally, table 2.10 shows the portfolio performance for the alternative factor specifications. The table indicates that the results for the first and third specification are again very similar to previously reported results. Abnormal returns are of similar magnitude. Interestingly, the first column shows that the addition of the cross-sectional average of stock returns leads to a portfolio with large abnormal returns, but without a strong factor structure at the one month horizon. It only loads on the momentum factor and its R^2 in the performance regression is only 20%. At longer horizons, the factor structure is however more pronounced.

The second specification is not able to generate significant abnormal returns and loads heavily on the market, the size and the momentum factor. This confirms our suspicion that this model is misspecified.

We conclude that our results are robust to changes in factor specifications. Estimated coefficients, abnormal returns and factor loadings are very similar to results reported in previous sections.

Table 2.10: Portfolio Performance: alternative specifications

The table shows the performance of the long-short portfolios at different prediction horizons for different methods to correct for (un)observed factors. Long-short portfolios are formed by selecting at each t the 30% stocks for which $\sum_{\ell=1}^{L} D_{i\ell} \left(\tau_l + \beta_l' x_{it} \right)$ is largest; short portfolios contain the 30% stocks with the lowest predicted returns $\sum_{\ell=1}^{L} D_{i\ell} \left(\tau_l + \beta_l' x_{it} \right)$. All returns are equally weighted.

Panel A The table reports the intercept α in the performance regressions

$$y_{LS,t} = \alpha + \delta_{t-1}' f_t + u_t$$

where f_t is a set of common risk factors and δ_{t-1} are risk factor loadings that are either constant or conditional. The first row of results ($\delta_t = 0$) refers to the raw average returns. The second row (FF4) includes the Fama-French market ($RMRf$), size (SMB), value (HML) and momentum (UMD) factors with constant factor loadings. In the third row (Cond FF4) the conditional risk loadings are specified as

$$\delta_t = d_0 + d_1 z_t,$$

with z_t equal to the default spread.

Panel B reports the coefficients $\delta_t = d_0$ and t-statistics for the FF4 model. All models contain industry-specific intercepts and industry-specific slope coefficients. The specification with $\overline{Y}$ uses the three Fama-French factors, the momentum factor and the cross-sectional average of stock returns to correct for factors. Model $\overline{YX}$ adds the cross-sectional average of x_{it}. Finally, specification PY uses the principal components of individual stock returns instead of the cross-sectional averages.

	One Month			Three Months			Six Months		
	$\overline{Y}$	$\overline{YX}$	PY	$\overline{Y}$	$\overline{YX}$	PY	$\overline{Y}$	$\overline{YX}$	PY
A: Intercepts (α)									
0	1.83	0.30	1.69	1.63	0.34	1.57	1.42	0.31	1.44
	(10.42)	(1.15)	(7.27)	(8.30)	(1.21)	(6.50)	(7.45)	(1.23)	(6.65)
FF4	1.95	0.42	1.92	1.73	0.55	1.73	1.49	0.46	1.51
	(8.57)	(1.65)	(8.35)	(8.37)	(1.88)	(7.21)	(7.85)	(1.85)	(7.19)
CFF4	1.91	0.35	1.87	1.66	0.48	1.67	1.42	0.38	1.44
	(8.64)	(1.38)	(8.61)	(9.26)	(1.74)	(7.77)	(8.76)	(1.56)	(8.27)
B: Factor loadings FF4 (d_0)									
b_M	0.08	0.35	0.24	0.12	0.26	0.26	0.13	0.23	0.19
	(1.76)	(5.87)	(5.32)	(3.04)	(3.16)	(6.45)	(3.11)	(3.07)	(5.13)
b_s	0.02	0.39	0.26	0.10	0.43	0.20	0.20	0.40	0.16
	(0.27)	(5.33)	(3.21)	(1.33)	(6.35)	(1.92)	(3.14)	(6.14)	(1.85)
b_h	0.13	-0.16	-0.05	0.35	-0.29	0.16	0.37	0.15	0.36
	(1.59)	(1.64)	(0.51)	(3.37)	(2.48)	(1.21)	(4.01)	(1.56)	(2.21)
b_u	-0.25	-0.35	-0.43	-0.36	-0.36	-0.48	-0.35	0.31	-0.39
	(3.16)	(3.74)	(5.14)	(4.37)	(3.39)	(5.08)	(4.41)	(3.08)	(3.93)
R^2	0.20	0.52	0.47	0.40	0.49	0.45	0.43	0.43	0.41

2.8 Conclusion

We use a panel data model to explain the cross-section of individual stock returns. Our models introduce industry-specific slope coefficients and time effects, and multiple forecasting horizons extending to six months.

Combining firm characteristics based on a multivariate prediction model clearly enhances the predictive power compared to univariate sorting methods. Size, dividend-to-price and turnover are the most powerful predictors in a multivariate context. Short-term momentum is mainly significant for longer forecasting horizons consistent with the holding periods in portfolio-based momentum strategies.

The main result of the empirical analysis is that industry effects are important in cross-sectional prediction strategies. These effects can be best captured by industry-specific coefficients and intercepts. Simulations of long-short portfolio strategies result in portfolios with a low turnover and substantial abnormal returns. These portfolios have significant exposures to the market and the size factor, but virtually no exposure to the value and momentum factor.

Another contribution of this paper is the introduction of forecasting horizons exceeding the one-month frequency typically used in previous studies on the cross-section of stock returns. Longer forecasting horizons drastically reduce the portfolio turnover and hence transaction costs, do not deteriorate alpha's, while having the same risk exposures as one month horizon strategies. Some portfolios are characterized by high and significant abnormal returns and very low turnover. Finally, we consider several specifications to model the cross-sectional dependence of the error terms and conclude that our results are robust. This holds for specifications that contain pooled time-effects, industry-specific time effects and factor models with/without unobserved factors.

Whereas we emphasized specification issues of the panel, ultimately, the predictive implications of the model should be subjected to an out-of-sample test. A precise out-of-sample analysis of the performance of various portfolio strategies is outside the scope of the present paper. Since we selected firm characteristics that the previous literature has identified as significant predictors in univariate or bivariate portfolio strategies, we ex ante know that the variables in our panel have predictive power. But this predictive power was established in very much the same sample period as we have used to develop

our multivariate model. Another possible extension could be to use a conditional model. We have not used time-varying slopes in order to limit the in-sample data snooping.

Finally, the inclusion of individual effects in the models shows that some explanatory variables are still missing in our analysis. Consequently, this strand of the literature should continue to explore other characteristics and specifications.

3

Long-term strategic asset allocation: an out-of-sample evaluation[1]

The objective of this paper is to find out whether the expected potential gains from strategic asset allocation can be realized in an out-of-sample test. Firstly, we find that long-term investors should time the market if they use our proposed shrinkage prior. This prior downplays the predictability of asset returns and leads to superior out-of-sample results compared to a standard uniform prior. Important is the use of a utility metric to evaluate prediction models. Shrinkage limits the losses in extreme negative events and this is what risk-averse investors value the most. Secondly including the hedge component of strategic portfolios only leads to a modest performance improvement out-of-sample. Repeated myopic strategies perform almost as well as a dynamic asset allocation strategy. Monte Carlo simulations relate this finding to estimation error, i.e. the estimated repeated myopic and dynamic portfolios approximate the true unknown optimal dynamic portfolio equally well. Next, our paper shows that incorporating parameter uncertainty leads to a small performance improvement. Finally, portfolio weight restrictions improve performance for bad models and hurt the good models.

---

[1]This chapter is based on Diris, Palm, and Schotman (2011).

3. LONG-TERM STRATEGIC ASSET ALLOCATION: AN OUT-OF-SAMPLE EVALUATION

3.1 Introduction

Individuals and institutions (e.g. pension funds) invest financial wealth in different asset classes to meet their long-term goal. Individuals save money for retirement. Pension funds invest on behalf of their participants to provide them with retirement income. Merton (1969, 1971) showed that under changing investment opportunities, the optimal portfolios of these long-term investors (their strategic asset allocations) 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. A second element of the strategic portfolios is inflation and interest rate risk. Long-term real returns from nominal bonds are subject to inflation risk, making them unattractive for long-term investors. Similarly short-term T-bills are not risk-free in the long-run, because they must be rolled over repeatedly. Long-term investors have to take these risks into account in their hedge portfolios. If investment opportunities are changing, optimal long-term portfolio allocation requires that investors dynamically adjust the portfolio weights every period.[1]

By now, there exists a rich literature (e.g. Campbell, Chan and Viceira, 2003 and Brandt, Goyal, Santa-Clara and Stroud, 2005) that shows how to calculate the hedge portfolio and investigates the utility gains from these long-term strategic asset allocations in-sample. However, there are reasons to doubt the utility gains from strategic portfolio choice in practice, since the models of asset returns might be subject to substantial estimation error. First, Goyal and Welch (2008) document the poor out-of-sample predictability of equity returns, thus casting doubt on the mean reversion of stock returns. If returns are indeed nearly unpredictable, the optimal portfolio composition should not exhibit much time variation.[2] Secondly, strategic asset allocation is even more demanding than myopic portfolio choice. The strategic portfolio consists of a speculative component that depends on the predictions of single period returns and a hedge component that is sensitive to the long-run predictions of returns and

---

[1]See Campbell and Viceira (2002) for a broad overview of strategic asset allocation.

[2]Formally, returns are unpredictable when the conditional distribution of future returns at time t is equal to the unconditional distribution for all t. It is common in the literature to focus on (un)conditional means when assessing predictability. We follow this convention here.

their covariance with current returns. The strategic portfolio is affected by estimation error in both components, whereas the myopic portfolio is only affected by errors in the speculative component. Therefore, the strategic portfolio is more susceptible to estimation error and might not perform very well in an out-of-sample test. Thirdly, unrestricted optimized portfolios for long-term investors based on estimates of the underlying dynamics show wildly fluctuating portfolio weights. The portfolio composition is even more extreme than the portfolio for short-term investors. This phenomenon is acknowledged by Campbell, Chan and Viceira (2003) among others. These extreme weights are subject to what is called "error maximization" and magnify any small misspecification in the return prediction model.

The performance measurement of strategic portfolios is still an open question in the academic literature, despite the relevance for (institutional) investors and the issues raised above. Therefore, our main objective in this paper is to find out whether the potential gains from strategic portfolios can be realized in an out-of-sample test. Because the gains from hedge demands apply to long investment horizons, performance evaluation of strategic portfolio choice requires long-term returns. Existing studies, however, use a single period return metric and thus cannot evaluate the out-of-sample utility gains from hedge demands.[1]

Our long-term investor optimizes the expected utility of wealth at a five year horizon using power utility. She is allowed to invest in a real T-bill, a stock index and a 5-year government bond. The predictive state variables are the price-earnings ratio, yield spread, and three-months T-Bill rate.[2] We measure the portfolio performance using the certainty equivalent returns based on the average realized utility over repeated five year horizons. In our analysis, we look at both the certainty equivalent return and the hedge component.

We use Bayesian time-series methods to estimate a model of investment opportunities.[3] We use a general Bayesian shrinkage prior advocated by Berger and Strawdermann (1996) and adapted to vector autoregressions by Ni and Sun (2003). Such a

---

[1]Some recent examples containing short-term out-of-sample results are Campbell and Thompson (2008), Goyal and Welch (2008) and Wachter and Warusawitharana (2009).

[2]As a robustness test, we also consider the dividend-yield as a predictor instead of the price-earnings ratio

[3]Some example from the growing Bayesian literature include Merton (1980), Cremers (2002), Wachter and Warusawitharana (2008), Jorion (1986), Black and Litterman (1992), Avramov (2002) and Pastor and Stambaugh (2000).

3. LONG-TERM STRATEGIC ASSET ALLOCATION: AN OUT-OF-SAMPLE EVALUATION

Bayesian prior provides more plausible parameter estimates than a uniform prior such that optimal portfolio strategies become less aggressive (risky) and therefore avoid implausible extreme positions. More specifically, the prior shrinks slope coefficients in the predictive regressions for excess returns on stocks and bonds to zero, and shrinks the coefficients of the state variables to a random walk. It downplays the predictability in the data and therefore corresponds to the prior information of an investor who is skeptical with respect to the predictability of returns. Its generality allows for applications in larger systems than the setting in Wachter and Warusawitharana (2009).

We analyze the performance of this shrinkage prior, in particular whether it outperforms a standard uniform prior and whether these differences are robust to changes in the set-up. Much of the portfolio choice literature (e.g. Barberis, 2000) advocates the use of Bayesian decision-theory to account for parameter uncertainty. Supposedly, it leads to more robust portfolios and is another way to avoid the extreme "wacky" weights (Cochrane, 2007). The second method we use, called plug-in method, ignores parameter uncertainty and conditions on a given set of estimated parameters (using the posterior mean). A third way to stabilize portfolio weights are short-sell constraints as argued in Jagannathan and Ma (2003). We consider specifications with and without constraints on the portfolio weights.

For the set-up that ignores parameter uncertainty (with unrestricted weights), Jurek and Viceira (2006) derive closed form solutions for the optimal strategic portfolios. For the version of the model that accounts for parameter uncertainty as well as the plug-in version that uses restricted portfolio weights we need numerical optimization. Our performance analysis requires a fast and stable numerical algorithm. We succeed in accelerating the method of Brandt, Goyal, Santa-Clara and Stroud (2005) by introducing a quadratic interpolation step that dramatically reduces the grid size of portfolios that must be evaluated. This makes our extensive out-of-sample analysis feasible.

Not surprisingly we find that a naive implementation of strategic asset allocations that uses a uniform prior can lead to disastrous performance in terms of certainty equivalence returns. Weights are wildly fluctuating and this leads to periods with badly performing portfolios. More interestingly, we find that using Bayesian shrinkage priors leads to superior out-of-sample performance for long-term investors. Both the strategic as well as repeated myopic portfolios substantially and significantly outperform an unconditional strategy that ignores predictability and hedging. Changing portfolio

allocations over time pays off for a long-term investor. Results are robust to small changes in the setup (such as different predictor variables) and the optimization as long as we use the shrinkage prior.

It turns out that it is very important to use a utility metric for assessing the performance of a prediction model. Risk averse investors evaluate big gains and big losses differently, since they want to avoid big losses at all costs. Due to this asymmetry in the utility function the best return prediction model for a risk averse investor is not necessarily the one that that has the lowest prediction error. It is the model that helps the investor avoid the big extreme (negative) events. It turns out that prediction models based on the shrinkage priors are best at avoiding these extreme events.

In terms of expected utility, the strategic portfolio performs only marginally better than the repeated myopic portfolio, even though both portfolios differ most of the time in terms of their asset mix. We conduct a Monte Carlo study to analyze the performance of the myopic and strategic portfolios rules. In simulated data, containing some predictability, the estimated myopic rule is more aggressive/riskier than the true myopic portfolio rule. By being more aggressive, the estimated myopic rule moves towards the optimal strategic rule. The estimated strategic rule is also too aggressive, thereby overshooting the true optimal rule. Compared to the truly optimal strategic portfolio, the estimated myopic rule is not aggressive enough, whereas the estimated strategic rule is too aggressive. In the end the estimated myopic and strategic rules produce almost the same average realized utility. Both rules suffer from estimation error, but the strategic rule is hurt more by estimation error than the (repeated) myopic rule. The hedge component of the strategic portfolio only marginally improves performance compared to a repeated myopic strategy that ignores this hedge component.

Accounting for parameter uncertainty improves performance slightly. Brandt, Goyal, Santa-Clara, and Stroud (2005) show that parameter uncertainty mainly has an impact on the weights of the hedge portfolio. As this hedge component does not have a big impact on performance (positively or negatively) in general, it is not surprising that parameter uncertainty does not have a large impact on performance. Portfolio weight restrictions have a larger impact on results. If portfolio weights are restricted, the best models perform worse and the bad models perform better.

The remainder of this article is organized as follows. Section 2 presents the data we use. Sections 3-5 describe respectively the general methodology, the modeling frame-

work and the solution method. Section 6 consists of the out-of-sample results. Section 7 provides some robustness tests and finally section 8 concludes. The appendix contains technical details on the estimation techniques and the numerical optimization algorithm.

3.2 Data

Our empirical analysis is based on monthly data for the US stock and bond market. We use data on three assets and two sets of three predictor variables; i.e. the nominal yield, the yield spread and either the price-earnings ratio or the dividend yield.

The monthly data set starts in February 1954 and ends in December 2006. The first three variables are log returns on different types of assets.[1] The first variable is the ex post real T-bill rate which is the difference between the log return (or lagged yield) on the 3-month T-bill, obtained from the FRED website[2], and log inflation, obtained from the Center for Research in Security Prices (CRSP). The second variable is the excess log stock return, which is defined as the difference between the value weighted log return on the NYSE, NASDAQ and AMEX market (including dividends) and the log return on the 3-month T-bill. The third variable, the excess log bond return, is defined in a similar way, but it uses the five-year bond return from CRSP.

The sets of predictor variables have been shown to predict stock and/or bond returns in-sample. However, their out-of-sample predictive power is doubtful as argued in Goyal and Welch (2008) for stock return predictability. Fama and Schwert (1977) and Campbell (1987) among others show that the log nominal yield on the 90-day T-Bill predicts both stock and bond returns. Next, the log dividend-to-price ratio is defined as the log of the ratio of the sum of dividend payments over the past year divided by the current stock price. Dividend payouts are extracted from stock data by combining the value-weighted return including dividends and the index level excluding dividends of the NYSE, NASDAQ and AMEX market. Campbell and Shiller (1998) show that this ratio predicts stock returns. The log yield spread is defined as the difference between the log yield on a 5-year bond obtained from the FRED site and the log yield on the 90-day T-Bill. This spread forecasts stock returns and bond returns according to

---

[1]We use log asset returns when estimating our econometric model. However, we transform the log asset returns into simple returns when evaluating portfolio performance.

[2]http://research.stlouisfed.org/fred2

Table 3.1: Summary Statistics

This table reports the means, standard deviations, minima, maxima and AR(1) coefficients for the ex post T-bill rate (R_{tbill}), the excess stock return (X_s), the excess bond return (X_b), the nominal yield (Y_{nom}), the dividend-to-price ratio (DP), the price-earnings ratio (PE) and the yieldspread (Y_{spr}). The monthly data set starts in February 1954 and ends in December 2006. Percentages are given as fractions.

	R_{tbill}	X_s	X_b	Y_{nom}	DP	PE	Y_{spr}
Mean	0.0010	0.0048	0.0011	0.0501	-3.5339	2.8565	0.0112
Std dev.	0.0030	0.0428	0.0148	0.0261	0.3820	0.4141	0.0091
Min	-0.0112	-0.2607	-0.0692	0.0058	-4.5637	1.8929	-0.0160
Max	0.0112	0.1483	0.0898	0.1443	-2.8452	3.7887	0.0421
AR(1)	0.3831	0.0722	0.1089	0.9837	0.9930	0.9968	0.9193

Campbell (1995) and Fama and French (1989). The log of the price-earnings ratio is obtained from the Irrational Exuberance data, available from the website of Professor Shiller.[1] It is defined as the log of the current price over the lagged sum of earnings over the past 10 years. Campbell and Shiller (1998) show that this yield is a predictor of stock returns. In section 3.6, we use the the nominal yield, the price-earnings ratio and the yield spread. As a robustness check, we replace the price-earnings ratio by the dividend-to-price ratio in section 3.7.

These asset return and predictor variables are commonly used in the strategic asset allocation literature, see e.g. Campbell, Chan, and Viceira (2003) and Jurek and Viceira (2010). Table 4.1 provides summary statistics of our monthly data.

3.3   Methodology

This section describes the methodology we use in this paper. The first subsection explains the general set-up of our out-of-sample analysis. The second subsection explains the difference between the plug-in and decision-theoretic method. For the plug-in method, estimates are substituted for the unknown parameters in the predictive distribution function. The last subsection gives some intuition about the relative performance of different strategies.

[1] http://www.econ.yale.edu/ shiller/data.htm

3. LONG-TERM STRATEGIC ASSET ALLOCATION: AN OUT-OF-SAMPLE EVALUATION

3.3.1 General set-up

Define the $n \times 1$ vector y_t as follows

$$y_t = \begin{pmatrix} r_{tbill,t} \\ x_t \\ s_t \end{pmatrix}, \tag{3.1}$$

where $r_{tbill,t}$ is the real return on the T-bill, x_t is a vector of excess returns on stocks and bonds, and s_t is a vector of predictor variables. Vector s_t either consists of the nominal yield $Y_{nom,t}$, the price-earnings ratio PE_t and the yield spread $Y_{spr,t}$ or the nominal yield, the dividend-yield DP_t and the yield spread. Hence, $n = 6$.

We consider investors who start with initial wealth normalized to 1 and maximize expected utility over terminal wealth K periods in the future by investing in the real T-bill, a stock index and a government bond. We choose power utility for preferences. We consider both restricted and unrestricted portfolio weights. Restricted weights impose short-sell constraints.

More formally, the investor has power utility with $\gamma > 1$ and chooses portfolio weights $w_t, ...w_{t+K-1}$ such that the value function at time point t is maximized

$$V_t(K, Z_t, W_t) = \max_{w_t,...,w_{t+K-1}} E\left(\frac{W_{t+K}^{1-\gamma}}{1-\gamma}\bigg| Z_t\right) \tag{3.2}$$

subject to the budget constraint

$$W_{s+1} = W_s\left(1 + w_s' R_{s+1}\right), s = t, ...t + K - 1, \tag{3.3}$$

where Z_t are conditioning variables that summarize all information available at time t, W_t is the wealth at time t, γ is a constant relative risk aversion parameter and R_{s+1} is the vector of simple returns on the assets in period $s + 1$. Portfolio weights add up to 1. Section 3.3.2 explains that the conditioning variables Z_t are equal to vector y_t under our assumptions and therefore we replace Z_t by y_t in the following.

Since initial wealth is 1, the following equality holds

$$W_{t+K} = \prod_{s=t}^{t+K-1} \left(1 + w_s' R_{s+1}\right). \tag{3.4}$$

We consider two types of strategies: a dynamic strategy and a myopic strategy. The dynamic strategy is the optimal solution to the long-horizon problem in equation (3.2)

and contains both a myopic as well as a hedging component, defined as the difference between the dynamic and the myopic strategy. The myopic strategy ignores the long horizon, sets portfolio weights as if the remaining horizon is only one period and hence ignores the hedging part. More formally, the dynamic $w_{t,D}$ and myopic strategies $w_{t,M}$ are defined as follows

$$\{w_{t,D}, ..., w_{t+K-1,D}\} = \arg\max E \left\{ \frac{\left(\prod_{s=t}^{t+K-1}(1 + w'_{s,D}R_{s+1})\right)^{1-\gamma}}{1-\gamma} \mid y_t \right\} \qquad (3.5)$$

$$\{w_{s,M}\} = \arg\max E \left\{ \frac{\left(1 + w'_{s,M}R_{s+1}\right)^{1-\gamma}}{1-\gamma} \mid y_s \right\}, s = t, ., t + K - 1. \qquad (3.6)$$

If horizon $K = 1$, the two strategies are obviously identical.

An econometric model is needed to evaluate the conditional expectation in equation (3.2). Following among others Campbell, Chan, and Viceira (2003) and Jurek and Viceira (2010), the dynamics of asset returns and state variables are assumed to follow a VAR(1)

$$y_{t+1} = B_0 + B_1 y_t + \epsilon_{t+1}, \qquad (3.7)$$

where B_0 is a vector of intercepts, B_1 is a matrix of slope coefficients and ϵ_{t+1} is a vector of errors for which we make the following common assumption

$$\epsilon_{t+1} \sim N(0, \Sigma). \qquad (3.8)$$

For future reference, it is useful to introduce the following decomposition for Σ, consistent with equation (3.1)

$$\Sigma = \begin{pmatrix} \sigma_{tbill}^2 & \sigma'_{tbill,x} & \sigma'_{tbill,s} \\ \sigma_{tbill,x} & \Sigma_x & \Sigma'_{x,s} \\ \sigma_{tbill,s} & \Sigma_{x,s} & \Sigma_s \end{pmatrix}. \qquad (3.9)$$

We take a Bayesian perspective and obtain posterior distributions for the parameters for various prior distributions. We either use a uniform prior or a shrinkage prior, details are explained below.

In the portfolio choice literature, there are two methods that prescribe how to use these estimation results. The plug-in method substitutes parameter estimates for the

true parameters. A second method acknowledges that there might be parameter uncertainty which can be taken into account by the posterior distribution of the parameters. This is the decision-theoretic method.

When making decisions, investors need to translate data into an econometric model and the econometric model into portfolio allocation rules. Different choices in this process lead to different portfolio weights. We mainly focus on whether investors should actively time the stock and bond market, whether they should incorporate the hedge portfolio and whether the shrinkage prior leads to improved results over the uniform prior. In order to tackle these issues, we consider the following choices for investors with risk aversion level γ ranging from 2 to 5 to 10:

- Uniform or shrinkage prior (2 choices)

- Dynamic or myopic strategy (2 choices)

- Plug-in or decision-theoretic method (2 choices)

- Restricted or unrestricted portfolio weights (2 choices).

We have to be careful in calculating and evaluating portfolio strategies for all combinations above. Firstly, although an investor with power utility can go short if she would be able to trade in continuous time, this is not possible in discrete time. Her expected utility is not finite in the latter case. However, many papers (e.g. Campbell, Chan, and Viceira (2003), Jurek and Viceira (2010)) that use the plug-in method still consider shortselling in a discrete time setting by using solution methods that approximate the continuous time solution. We follow this standard in the literature and report results for the plug-in method using unrestricted weights. Branger, Breuer, and Schlag (2010) analyze this common practice and conclude that a naive implementation of a continuous time strategy in discrete time is viable as long as derivatives are not part of the asset menu.

Secondly, the tails of the posterior predictive distribution of asset returns are fatter than the tails of the normal distribution if parameter uncertainty is incorporated. Since all assets, even the T-bill, are risky in our setting, this implies that the expected utility of all portfolio strategies is minus infinity unless we make a slight modification.[1] We

---

[1]The T-bill is risky due to inflation risk.

solve this problem by imposing a lowerbound of -20% on the monthly return on the real T-bill rate. This guarantees that at least some portfolios have finite expected utility under the decision-theoretic approach. However, portfolios that involve short-selling do not have finite expected utility under the decision-theoretic method. The optimal portfolio in such a setting therefore exactly coincides with the optimal portfolio in a setting with restricted portfolio weights. Therefore, we do not report these results separately.

Hence, for all three risk aversion levels we consider 12 different specifications. Furthermore, we also calculate five benchmark specifications. Firstly, the $1/N$ rule that invests one third of the wealth in each asset. This fixed rule does not depend on data. Next, we consider rules that dogmatically impose that excess stock and bond returns are unpredictable, either combined with restricted or unrestricted weights and a myopic or dynamic strategy.[1] Investors that follow these rules do not actively time the stock and bond market. In order to limit the total number of specifications, we combine the latter rules only with the plug-in method. The solution method we use depends on whether weights are (un)restricted, what kind of strategy we use (myopic or dynamic) and how we use the econometric estimation results (plug-in method or decision-theoretic method).

In the out-of-sample analysis, our first investor has an investment horizon of K months and uses all data available until period t_{start} to choose her first portfolio weights $w_{t_{start}}$. In the next period $t_{start}+1$, her investment horizon is $K-1$ and she updates her information set to choose portfolio weights $w_{t_{start}+1}$ etcetera. In period $t_{start} + K - 1$, her investment horizon is 1 period and she uses all data until that period to choose her last portfolio weights $w_{t_{start}+K-1}$. This sequence of K portfolio weights results in exactly one terminal wealth value at time $t_{start} + K$, the end of the horizon. The next investor follows a similar strategy but she starts in period $t_{start} + 1$ and ends in period $t_{start} + K + 1$ with again exactly one terminal wealth value. We repeat this analysis for many investors, all with horizon K, who start their strategies one month after each other. The last investor starts in $T - K$ and ends in T, the end of our sample. In this way, we obtain a time series of terminal wealth values and a time series of realized utility values. This sample of realized utility values is used to measure performance.

---

[1]The dynamic and myopic specifications are not equal in this setting, since the expected real T-Bill rate is assumed to vary over time.

3. LONG-TERM STRATEGIC ASSET ALLOCATION: AN OUT-OF-SAMPLE EVALUATION

It provides a measure of out-of-sample performance of investors, since we only use information that is available to investors in real time.

In setting up the out-of-sample experiment, we need to make several choices. Firstly, we choose our starting date t_{start} to be equal to February 1974 in order to have enough initial observations (20 years) to estimate a model and to have a representative out-of-sample period. This choice is identical to the choice made in Wachter and Warusawitharana (2009). Secondly, we choose the investment horizon $K = 60$ months. This is a medium to long-term horizon and gives us almost 7 non-overlapping out-of-sample investment periods. Next, every month we allow investors to use all available information up to this month to update their portfolio holdings. This means that we re-estimate our models every month to include the newest observations using an expanding data window. Finally, we use the certainty equivalence return (CER) as performance criterium. It is the riskfree return that would make investors indifferent between following a strategy or accepting this riskfree real return. The CER is a monotone transformation of average utility values $\overline{U}$ realized over the out-of-sample investment periods and is given as follows

$$CER = \left(\bar{U}(1-\gamma)\right)^{\frac{1}{1-\gamma}} - 1. \tag{3.10}$$

In the tables, we report the annualized certainty equivalence returns $(1 + CER)^{\frac{1}{5}} - 1$.

A small note on methodology. The strategic asset allocation literature uses both Bayesian and frequentist methods for inference. The former are mainly combined with the decision-theoretic method while the latter are usually combined with the plug-in method. We choose the Bayesian perspective. Two important practical reasons are that it is conceptually (more) straightforward (i) to include parameter uncertainty (the decision-theoretic method) in the decision process and (ii) to set-up a shrinkage estimator. It would however also have been possible to use frequentist techniques. A frequentist econometrician could use a bootstrap to incorporate estimation uncertainty in finite samples and could choose from a large array of shrinkage estimators (e.g. ridge regression) to estimate the models.[1]

---

[1] On a deeper more philosophical level, the methodology in our paper combines the Bayesian way for inference and making predictions with some frequentist elements (repeated out-of-sample strategies, significance tests as a robustness check in section 3.7.1) when evaluating predictions (portfolios) out-of-sample. In fact, many papers use such a combination implicitly, e.g. Avramov (2002), Wright (2008), Cremers (2002) etcetera. We view our out-of-sample set-up as the best approximation of the environment a long-term investor faces in real-time.

3.3.2 Plug-in method versus decision-theoretic method

In this section, we explain how to use the results from the econometric model. The first method is the plug-in method that treats the parameter estimates as the true values, ignoring any form of parameter uncertainty. This gives the following result for the conditional distribution of future values y_{t+1} for asset returns and state variables given their current values,

$$P\left(y_{t+1}|\hat{B}, \hat{\Sigma}, y_t\right), \tag{3.11}$$

where $\hat{B}, \hat{\Sigma}$ are estimates for B and Σ. In other words, the pdf of returns and state variables 1 period in the future is conditioned on estimated values. From the VAR(1) model defined in equations (3.7) and (3.8), returns are conditionally lognormally distributed. The current values of asset returns and state variables summarize the conditioning space (next to the parameter estimates). This approach is adopted by Campbell and Viceira (2002) and Jurek and Viceira (2010).

The second method is the decision-theoretic method. It uses the following conditional predictive probability density function for asset returns and state variables

$$P\left(y_{t+1}|\{y_s\}_{s=1}^t,\right) = \int P\left(y_{t+1}|B, \Sigma, y_t\right) P\left(B, \Sigma|\{y_s\}_{s=1}^t\right) d\Sigma dB. \tag{3.12}$$

Hence, a (posterior) distribution for parameters (B, Σ) is used to integrate over the parameters, i.e. parameter uncertainty is taken into account.

The advantage of this method is that it takes both parameter uncertainty and uncertainty due to the stochastic nature of the variables into account. The disadvantage is that it is difficult to specify a posterior distribution that accurately describes what we really know about the parameters. Another disadvantage is that the posterior predictive distribution of returns in (3.12) is not lognormal anymore. This implies that we have to rely on numerical simulation methods for portfolio construction. Analytical properties of returns $L > 1$ periods in the future are not known anymore, but we can simulate them. References for this method are Barberis (2000) and Brandt, Goyal, Santa-Clara, and Stroud (2005).

The dynamic strategy is equal to the myopic strategy plus a term that hedges against changes in the investment opportunity set. In case of the plug-in method, the investment opportunity set is completely determined by the current value of the vector y_t. However, if we use the decision-theoretic method, this is not necessarily

true. An investor learns more about the true unknown values of the parameters over time. This implies that her investment opportunity set also changes over time since the posterior parameter distribution is updated over time. In other words, hedging against a changing investment set means hedging against the changing posterior distribution due to learning as well when we consider the decision-theoretic approach. We ignore this learning aspect however, because it is unfeasible given the size of our VAR(1) system. Since the VAR(1) system is of dimension $n = 6$, introducing this aspect would mean that we need 69 conditioning variables in vector Z_t to describe the investment opportunity set.[1] This is infeasible as the numerical methods that are used in the portfolio literature currently only solve problems up to 11 conditioning variables (see e.g. Brandt, Goyal, Santa-Clara, and Stroud (2005)).

We follow Barberis (2000) and assume that investors take parameter uncertainty into account, but ignore the impact of changing beliefs on today's asset allocation. They invest as if they only learn about the parameters at the end of their investment horizon. Under this assumption, the values of y_t summarize the conditioning space at time t (next to the posterior distribution at time t). Note that our investors still learn about the true parameter values through time if new observations become available. The simplification we make is that they do not hedge against this learning. Brandt, Goyal, Santa-Clara, and Stroud (2005) show by means of simulations that incorporating parameter uncertainty while ignoring learning leads to improved performance relative to the case without parameter uncertainty. They show that the losses, that are incurred because learning is ignored, are cut in half if parameter uncertainty is incorporated.

3.3.3 Comparison of strategies

One of the aims of this paper is to investigate whether investors should take the hedge component of strategic portfolios into account in an out-of-sample test. In order to answer this question we analyze whether a dynamic strategy outperforms repeated myopic strategies. In case we would know the process that generates asset returns and state variables perfectly, this would be a trivial question to answer. A dynamic strategy would be superior to repeated myopic strategies, since the former strategy encompasses the latter (for the same investment horizon).

---

[1] All distinct parameters plus the values of the variables.

As we do not know the true data generating process (DGP), we have to select and estimate a model. This model is however by definition misspecified and estimates suffer from sampling errors. For the myopic portfolio, the errors are only related to estimation error in the single period expected returns. The hedge component however is also sensitive to the long-run predictions of returns and their covariance with current returns. Out-of-sample, it is therefore far from trivial which strategy works best.

3.4 Empirical modeling

This section describes how we model the time-varying investment opportunity set and gives estimation results for these models.

3.4.1 Econometric model and estimation

In order to facilitate the prior choice, we firstly re-parametrize the VAR(1) model by transforming the state variables into

$$y_{trans,t} = \begin{pmatrix} r_{tbill,t} \\ x_t \\ \Delta s_t \end{pmatrix} \tag{3.13}$$

and use the following transformed auxiliary model in the estimation stage

$$y_{trans,t+1} = B_0 + B_1^* y_t + \epsilon_{t+1}. \tag{3.14}$$

We are mainly interested in the posterior distributions for B_0 and B_1. Therefore, we generally first obtain the posterior distribution for coefficients B_0 and B_1^* in the auxiliary model and subsequently add 1 to the diagonal elements in B_1^* that correspond to the predictor variables to obtain the posterior distribution for B_1. We only report and use the latter.

In order to estimate the VAR(1) model in equation (3.14), provide inference and make forecasts, we use, in line with most of the literature, a conditional likelihood function that conditions on the first observation. The conditional likelihood function is

$$P(Y^*|B,\Sigma) \propto |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr\left[(Y^* - XB^{*\prime})'(Y^* - XB^{*\prime})\Sigma^{-1} \right] \right\}, \tag{3.15}$$

where T is the number of observations, Y^* is the $T \times n$ matrix of observations on $y_{trans,t}$, Y_{-1} is the $T \times n$ matrix of lagged observations on y_t, X is the $T \times (n+1)$

matrix $X = [\iota, Y_{-1}]$ and B^* is the $n \times (n+1)$ matrix $B^* = [B_0, B_1^*]$. We are both interested in the posterior distribution of the parameters and in their point estimates. For point estimates we use the posterior means.

Our first prior is a uniform prior on B^* and a Jeffrey's prior on Σ,

$$p(B^*, \Sigma) \propto |\Sigma|^{-(n+1)/2}. \tag{3.16}$$

We refer to this prior as the uniform prior. It is the most commonly used prior for VAR models. The corresponding posterior is given in equation (3.18) in the appendix. The posterior mean of B^* is equal to the OLS/ML estimator $\hat{B}^{*'} = (X'X)^{-1}X'Y^*$ and the posterior mean of Σ is equal to $S/(T-2n-2)$, where $S = (Y^* - X\hat{B}^{*'})'(Y^* - X\hat{B}^{*'})$. For the decision-theoretic approach, we need to simulate from the full posterior distribution of the parameters and the predictive distribution of the variables y_{t+1}. We explain this in the appendix.[1]

We consider a second Bayesian estimator which is used among others in Ni and Sun (2003) in the context of a similar VAR model. We refer to this prior as the shrinkage prior. This estimator shrinks the coefficients towards zero. The prior is given as

$$p(B^*, \Sigma) \propto \left(b^{*'} b^* \right)^{\frac{-(n(n+1)-2)}{2}} |\Sigma|^{-(n+1)/2}, \tag{3.17}$$

where $b^* = vec(B^*)$. The exponent is exactly equal to the exponent that Ni and Sun (2003) propose. It is the product of a shrinkage prior for B^* and the Jeffrey's prior on Σ. The prior itself is not proper, but Ni and Sun (2003) show that the posterior is proper in a VAR model when the ML estimator exists, which holds in our setting. Note that the prior has a negative exponent. This means that prior draws of large parameter values are relatively improbable. Shrinking the coefficients in the auxiliary model (3.14) towards a zero matrix implies that we are shrinking the coefficients in the original model towards zero except for the predictor variables which we shrink towards a random walk.[2]

This particular shrinkage prior has several advantages. Firstly, since the prior is improper, it is relatively uninformative. The likelihood dominates the prior quickly

---

[1]Results using the uniform prior are equivalent for the original and the auxiliary model.

[2]Note that if we would have combined the shrinkage prior with the original model, we would have shrunk the autocorrelation coefficients of the highly persistent state variables to 0 instead. This would have resulted in a misspecified model.

once there is sufficient data. If the data shows a lot of predictability, the posterior will reflect this. Secondly, the prior does not depend on any tuning constants. This avoids all kind of calibration issues that could arise. Finally, the prior leads to a posterior that is relatively easy to calculate using Gibbs sampling. The sampling algorithm is fast and stable, even for large VAR models.

Our shrinkage prior has a clear economic interpretation. It reflects the beliefs of an investor who is very skeptical about predictability of asset returns. As a result such an investor downplays all the predictability that is found in the data. However, the investor does not dogmatically ignore predictability. If there is sufficient evidence in the data that asset returns are predictable, this investor will take (some) asset return predictability into account.

The kernel of the posterior density is given in equation (3.21) of the appendix. The shrinkage prior is not conjugate, and hence does not lead to a known posterior density for the parameters. However, as Ni and Sun (2003) show, a straightforward MCMC sampler exists to draw from the posterior. The simulation algorithm is explained in appendix A.

If the lagged asset returns and predictor variables are not able to predict stock and bond returns, the second and third rows of B_1^* in model (3.14) are both equal to zero. As a benchmark, we consider specifications that dogmatically set these coefficients equal to zero and leave the coefficients in other equations equal to the posterior mean under the uniform prior.[1] We refer to this specification as the no-predictability prior.

The VAR(1) model introduced in equations (3.7) and (3.8) is restrictive in two ways. First, it is unlikely that all dynamics in the data are modeled by using only one lag. Second, it is unlikely that the covariance matrix of the error terms is homoscedastic, i.e. that risk is constant over time.[2].

However, if we add extra lags and model the time-variation in the variance of the error terms, we would end up with an enormous increase in the total number of estimated parameters. One extra lag already means $n^2 = 36$ extra parameters.

---

[1]Results are similar if we additionally assume that the real T-Bill rate is unpredictable.

[2]Actually, we can confirm these suspicions using results from later chapters. If we apply the methodology developed in the next chapter allowing for a maximum of two lags (using the prior distributions explained in that chapter), we indeed find that the second lags of some variables (particularly the second lag of PE in the PE equation) have posterior probabilities near 1. The conclusion from chapter 5 that error volatility varies persistently over time confirms the second suspicion.

Modeling the time-variation in all 21 unique elements of the covariance matrix leads to even more additional parameters. Since estimation efficiency is an important issue, we choose not to add extra lags and assume homoscedastic errors. This probably leads to a (slightly) misspecified, but also to a much more efficiently estimated model.[1]

Another way to put it is that we have a strong prior that incorporating extra lags and heteroscedasticity are not helpful in forecasting (we dogmatically exclude these aspects using our prior). It is quite common in the forecasting literature to shrink the coefficients of lags larger than 1 strongly towards 0, refer for example to the well-known Litterman (1986) prior. The impressive predictive performance of this Litterman prior suggests that the effect of the misspecification is usually more than offset by the larger estimation efficiency in settings such as the ours.

The state variables in the model are highly autocorrelated and close to a unit root. This does not pose a problem for inference, because (possible) non-stationarity does not require specific Bayesian methods (refer to Sims and Uhlig (1991)). Nevertheless, it is common in the strategic asset allocation literature to impose the assumption of stationarity (e.g. Campbell and Viceira (2002) and Stambaugh (1999)). For the decision-theoretic approach, we indeed impose stationarity. Since the mode of the likelihood function is generally within the stationary region, we do not impose this assumption when using the plug-in approach. This only slightly changes the point estimates, has a minor impact on the out-of-sample results, but saves on computation time.

3.4.2 Estimation results

Table 3.2 reports posterior moments for B and Σ for the model where the price-earnings ratio is one of the state variables using the uniform and shrinkage prior. The table shows that the state variables are highly autocorrelated under both priors. Furthermore, we see that the nominal yield and the price-earnings ratio have a negative effect on stock return predictions and that the yield spread has a positive impact on bond returns predictions. There is also a large positive correlation between shocks to the price-earnings ratio and excess stock returns, which means that unexpected positive shocks

---

[1]In other words, even if the true model is a VAR(2) model with heteroscedastic errors, we do not expect this model to give the best forecasting performance in the sample that we consider due to the large number of estimated parameters relative to the sample size.

Table 3.2: Estimation results PE model

This table reports estimates for the VAR(1) model based on the full data-set where we use PE among the state variables. Panel A gives results for the uniform prior and panel B for the shrinkage prior. In each panel, columns 2-7 show the posterior mean of the slope coefficients and their posterior standard deviations. The last column shows the implied R^2 (implied by the the posterior mean). Finally, the correlation matrix of the error terms is given. The elements on the diagonal are the standard deviations(x100) of the error terms, the off-diagonal elements are the correlations.

	r_{tbill}	x_s	x_b	s_y	s_{PE}	s_{spread}	R^2
Panel A: Uniform prior							
Parameter estimates							
r_{tbill}	0.3242	0.0027	0.0078	0.0276	0.0011	0.0412	0.1808
	0.0383	0.0026	0.0076	0.0058	0.0003	0.0136	
x_s	1.6434	0.0240	0.3249	-0.3521	-0.0180	-0.0822	0.0579
	0.5972	0.0400	0.1181	0.0897	0.0054	0.2117	
x_b	0.4215	-0.0569	0.0787	0.0410	0.0017	0.3127	0.0764
	0.2040	0.0137	0.0401	0.0308	0.0019	0.0723	
s_y	-0.0813	0.0144	-0.0652	0.9855	-0.0001	0.0206	0.9730
	0.0615	0.0041	0.0120	0.0093	0.0006	0.0217	
s_{PE}	1.3611	0.4168	0.3114	-0.1423	0.9917	0.1227	0.9954
	0.4025	0.0269	0.0790	0.0602	0.0036	0.1421	
S_{spread}	0.0056	-0.0048	-0.0650	0.0070	-0.0002	0.9503	0.8541
	0.0497	0.0033	0.0098	0.0075	0.0005	0.0176	
Error correlation matrix							
r_{tbill}	0.2702	0.1052	0.0757	-0.0805	0.1727	0.0560	
x_s		4.2049	0.1128	-0.0487	0.7746	-0.0331	
x_b			1.4400	-0.6237	0.0557	0.2208	
s_y				0.4328	-0.0494	-0.8516	
s_{PE}					2.8249	-0.0219	
S_{spread}						0.3503	
Panel B: Shrinkage prior							
Parameter estimates							
r_{tbill}	0.2730	0.0029	0.0068	0.0297	0.0012	0.0414	0.1782
	0.0371	0.0026	0.0075	0.0057	0.0003	0.0134	
x_s	0.0099	0.0261	0.1481	-0.2037	-0.0121	-0.0039	0.0381
	0.1173	0.0374	0.0840	0.0673	0.0048	0.0978	
x_b	0.1061	-0.0537	0.0775	0.0410	0.0016	0.2487	0.0705
	0.1054	0.0135	0.0377	0.0285	0.0018	0.0631	
s_y	-0.0173	0.0137	-0.0650	0.9854	-0.0001	0.0314	0.9729
	0.0456	0.0041	0.0117	0.0089	0.0005	0.0202	
s_{PE}	0.1038	0.4144	0.2031	-0.0564	0.9950	0.1360	0.9953
	0.1132	0.0257	0.0610	0.0469	0.0033	0.0792	
S_{spread}	-0.0086	-0.0046	-0.0644	0.0066	-0.0002	0.9480	0.8541
	0.0428	0.0033	0.0097	0.0074	0.0004	0.0170	
Error correlation matrix							
r_{tbill}	0.2706	0.1110	0.0792	-0.0828	0.1783	0.0566	
x_s		4.2400	0.1195	-0.0535	0.7789	-0.0316	
x_b			1.4432	-0.6247	0.0647	0.2209	
s_y				0.4331	-0.0552	-0.8511	
s_{PE}					2.8541	-0.0202	
S_{spread}						0.3503	

Figure 3.1: Overview of (x_s, s_{PE}) and (x_b, s_{spread}) coefficients over time
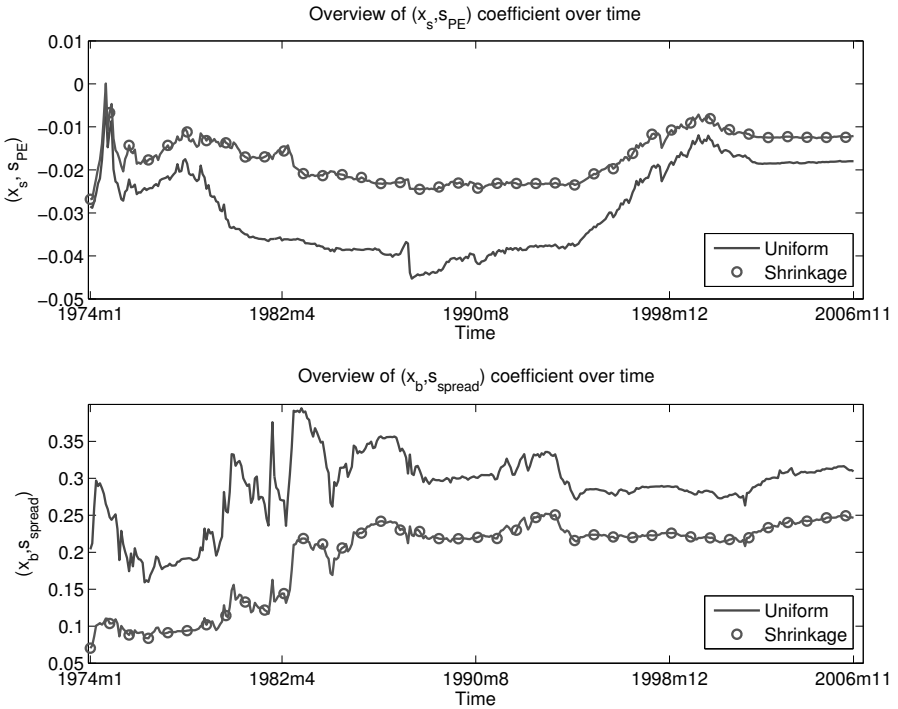
This figure plots the posterior mean of the coefficients (x_s, s_{PE}) and (x_b, s_{spread}) (y-axis) against time (x-axis) for the uniform and shrinkage prior. The model is estimated from February 1954 until the date on the x-axis.

to stock returns lead to negative future investment opportunities. This result implies that there is mean-reversion in stock returns.

Comparing the posterior means for both priors, we clearly see that the posterior mean for the return prediction coefficients are shrunk towards zero by the shrinkage estimator except for the autocorrelation coefficients which are shrunk to one. The shrinkage estimator downplays the predictability of asset returns. One way to see this is to look at the lower R^2 values under the shrinkage prior, especially for excess stock

returns. The lower R^2 values lead to less aggressive investment strategies.[1]

In the following sections, we re-estimate our models on larger and larger data-sets that include the newest observations. Since our data set starts in February 1954 and our empirical analysis in February 1974, we estimate models for which the last observation ranges from January 1974 until November 2006. Table 3.2 shows that the price-earnings ratio and the yield spread are among the most important predictors for respectively excess stock and bond returns. Therefore, we present time series plots of the estimates of the slope coefficients of (x_s, s_{PE}) and (x_b, s_{SPR}) in figure 3.1.

From the figure it is clear that the posterior means for the shrinkage prior are closer to 0 than those for the uniform prior. There seems to be a lot of uncertainty about the estimated values, since the parameters are very variable over time. However, the estimated values for the shrinkage estimator are less variable. Finally, note that the values for the two estimators slowly converge to each other once more observations are available, since the likelihood dominates when the sample size grows.

3.5   Solution method

This section explains the solution methods we use in this paper. This choice depends on whether we condition on parameter estimates (plug-in approach) or use the posterior distribution of the parameters in a decision-theoretic approach and whether we restrict portfolio weights or not. We use the semi-analytical method in Jurek and Viceira (2010) for calculating the unrestricted plug-in strategies. We have to use numerical methods for all other strategies. We propose a refinement of the method of Brandt, Goyal, Santa-Clara, and Stroud (2005) and van Binsbergen and Brandt (2007) by relying on an important observation made by Koijen, Nijman, and Werker (2010).

3.5.1   Analytical method

Given the VAR(1) model in equation (3.7), returns are lognormally distributed conditional on the parameter values. Jurek and Viceira (2010) use this fact to derive approximate-analytical solutions for the unrestricted plug-in model for the myopic and the dynamic strategy. These solutions are all based on the Campbell and Viceira (2002)

---

[1]The R^2 values we provide are implied by the posterior mean of the parameters. The mean of the posterior distribution of R^2 values does not exist when allowing for non-stationary draws.

approximation to log-portfolio returns. This approximation, and therefore Jurek and Viceira (2010)'s method, is exact in continuous time and accurate on short time intervals. Jurek and Viceira (2010) show that portfolio weights on risky assets are affine functions of the conditioning variables y_t.

3.5.2 Numerical method

There is no analytical solution available for the plug-in model combined with restricted portfolio weights. Furthermore, the predictive distribution of returns is not lognormal if parameters are integrated out and therefore there is no analytical solution available for the restricted decision-theoretic model. In these cases we have to use numerical methods.

Firstly, we consider the dynamic strategy. We solve the sequence of one-period problems by backward induction, i.e. start in period $K-1$ and iterate to period 0. We follow Brandt, Goyal, Santa-Clara, and Stroud (2005) and simulate many trajectories of asset returns and state variables and approximate the conditional expectations we need to evaluate by regressions of the value function at time $t+1$ on conditioning variables that summarize the information set at time t. Furthermore, we follow van Binsbergen and Brandt (2007) and set-up a grid of portfolio weights, evaluate the conditional expectation for the grid points and pick the maximum. Since we have to re-calculate dynamic strategies almost 400 times, computation time is an important issue. Therefore, we use a refinement by Koijen, Nijman, and Werker (2010) and parameterize the regression coefficients in regressions that approximate conditional utility by a quadratic function of portfolio weights.[1] This allows us to find the optimal weights along each path analytically by optimizing a quadratic function on a restricted set which can be done analytically. It means that we do not have to use a very fine grid since the parametrization regressions are very accurate.

This gives the following algorithm:

1. Generate N sample paths of length K of asset returns and state variables from the conditional prediction model ("plug-in") or from the predictive distribution ("decision-theory").

---

[1] Note that Koijen, Nijman, and Werker (2010) solve a life-cycle model with intermediate consumption and parametrize the first order conditions by an affine function in the portfolio weights. We parametrize the value function instead.

2. Set-up a grid of portfolio weights.

For period $K - 1$ until period 0 repeat steps 3, 4 and 5.

3. Pick one set of portfolio weights from the grid and calculate the realized utility values for all simulated paths. Hence: use the chosen portfolio weights together with the optimal portfolio weights chosen in previous steps to calculate the realized terminal wealth values for every path. Calculate the realized utility values for all paths.

4. Regress the N realized utility values on a constant and functions of the conditioning variables in order to calculate regression coefficients and conditional utility values.

Repeat step 3 and 4 for all portfolio weights on the grid.

5. Parametrize the regression coefficients in a quadratic function of the portfolio weights. This critical improvement allows us to express conditional utility as a function of constants, conditioning variables and portfolio weights. Along each path, constants and conditioning variables are known and hence along each path conditional utility is only a function of the unknown portfolio weights. For every path, choose the portfolio weights that maximize this approximate quadratic function. This can be done analytically.

The calculation of the myopic strategy is similar with $K = 1$. Appendix B gives more details on the parameterization of regression coefficients and the accuracy of the algorithm.

The decision-theoretic method combined with restricted portfolio weights gives some problems as indicated above. We guarantee that at least some portfolios have an expected utility value greater than minus infinity by imposing that the return on the real T-Bill rate is always larger than -20%.[1]

---

[1]In our numerical algorithm, we re-sample draws that would violate this boundary.

3.6 Out-of-sample performance

We investigate the out-of-sample performance of strategic asset allocations. The following subsections cover respectively benchmark results, results using the plug-in method and results using the decision-theoretic method.

3.6.1 Results for the benchmark specifications

We report results for five benchmark specifications in table 3.3. We show their certainty equivalence return (CER), their average terminal wealth and the standard deviation of terminal wealth. The first specification is the $1/N$ strategy. It invests 33% in stocks, bonds and T-bills irrespective of the data. The next four specifications are based on the no-predictability prior either combined with a dynamic or myopic strategy and

Table 3.3: Benchmark Results

This table gives benchmark results for the $1/N$, the unrestricted no predictability and the restricted no predictability strategies (NP). We either calculate a dynamic or a myopic strategy for the latter two. Specifications are based on the plug-in method. We show annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three different risk aversion levels γ.

		Unrestricted Weights			**Restricted Weights**		
		CER	$\overline{TW}$	$\sigma(TW)$	CER	$\overline{TW}$	$\sigma(TW)$
	Panel A: $\gamma = 2$						
$1/N$					0.0466	1.2885	0.2069
NP	Dyn	0.0886	1.8030	0.8159	0.0706	1.5444	0.4731
	Myop	0.0855	1.7859	0.8251	0.0706	1.5444	0.4731
	Panel B: $\gamma = 5$						
$1/N$					0.0386	1.2885	0.2069
NP	Dyn	0.0478	1.3678	0.2503	0.0450	1.3560	0.2540
	Myop	0.0443	1.3475	0.2571	0.0439	1.3586	0.2686
	Panel C: $\gamma = 10$						
$1/N$					0.0273	1.2885	0.2069
NP	Dyn	0.0265	1.2355	0.1457	0.0279	1.2350	0.1452
	Myop	0.0247	1.2151	0.1463	0.0258	1.2253	0.1519

unrestricted or restricted portfolio weights. The no-predictability prior imposes that excess stock and bond returns are not predictable.

The certainty equivalence returns (CER) in table 3.3 are all positive. This means that investors are willing to follow these simple strategies unless they are paid a positive risk-free real return. Interestingly, the dynamic strategy outperforms the repeated myopic strategy for all specifications. This implies that hedging (real) interest rate risk boosts performance for long-term investors as argued in Campbell and Viceira (2001). This is also reflected in the slightly higher average terminal wealth values and the slightly lower standard deviation of terminal wealth values. The performance difference is small however. The performance improvement is relatively more important for investors with higher risk aversion levels. Campbell and Viceira (2001) show that the hedge component for such investors is larger than for investors with lower risk aversion.

Remarkably, for a conservative investor with $\gamma = 10$, both the $1/N$ strategy and the restricted no-predictability strategies outperform the unrestricted no-predictability strategy. This suggests that imposing restrictions might improve out-of-sample performance and that non-data based methods are not necessarily inferior to data-based strategies. The former is consistent with results in Jagannathan and Ma (2003) who show that imposing weights restrictions is a form of shrinkage that boosts performance. The latter is consistent with DeMiguel, Garlappi, and Uppal (2009), who show that a $1/N$ strategy is tough to beat out-of-sample.

Figure 3.2 plots a histogram of realized utility values for the unrestricted no-predictability strategy. We set $\gamma = 5$. The figure shows that the utility value distribution is very left skewed. Most values are near zero but there are some large negative outliers (corresponding to low terminal wealth values). However, these negative outliers are the most important values for risk-averse investors. Risk averse investors want to avoid extreme negative events at all costs and will heavily weight every extreme event in their utility function. This suggests that specifications that limit the number and/or size of extreme events are the ones with the highest certainty equivalence returns (and equivalently highest average realized utility). The outliers are further emphasized when $\gamma = 10$, but are less severe when $\gamma = 2$.

Figure 3.2: Histogram of realized utility values for benchmark strategy with $\gamma = 5$
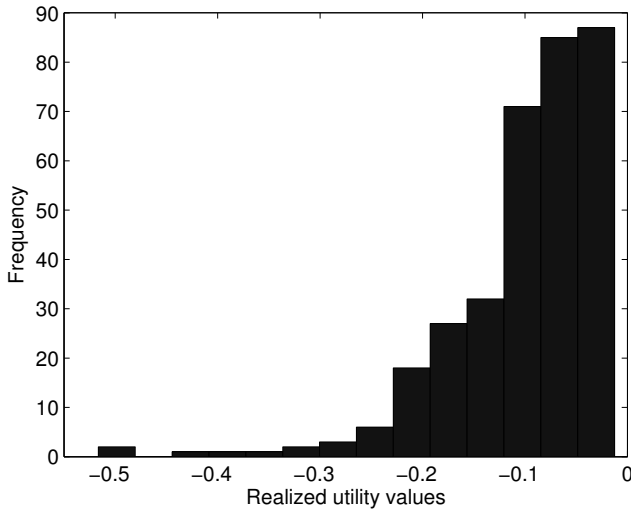
This figure gives a histogram of realized utility values for the unrestricted dynamic no-predictability strategy with $\gamma = 5$. We use the plug-in method.

3.6.2 Results for the plug-in method

Next, we show results based upon the plug-in method. We report results for dynamic and myopic strategies either using the uniform or shrinkage prior and either using restricted or unrestricted portfolio weights. Investors who use these specifications time the stock and bond market actively, since the specifications allow for predictable stock and bond returns. Results are given in table 3.4.

Firstly, we consider specifications using the uniform prior and unrestricted weights. Remarkably, the performance of an investor with low risk aversion ($\gamma = 2$) is disastrous under the standard uniform prior when weights are unrestricted. She is willing to pay a risk-free return of up to -100% to avoid adopting this strategy. The average terminal wealth and its standard deviation show why. The strategy leads to a very high average terminal wealth but with extremely high risk. Due to this risk, at least one of the terminal wealth values in our sample turns out to be zero which means that at least

one investor loses all her money during her 5-year investment period. Such an investor obtains a realized utility value of $-\infty$, since this is the outcome that such a risk-averse investor desperately wants to avoid.

The performance is better for higher γ values. These investors are less aggressive and avoid the strategies that lead to disaster for the $\gamma = 2$ investor. CERs are positive and higher than the ones for the benchmark strategies. Differences turn out to be economically important. For very risk averse investors, it pays off to time the bond and stock market.

Table 3.4: Plug-in approach - PE model

This table gives the results for the dynamic (Dyn) and myopic (Myop) strategies using the plug-in method. The results are based on a VAR(1) model with PE as one of the predictors. We report results under the uniform and shrinkage prior, either using restricted or unrestricted portfolio weights. We report annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three different risk aversion levels γ.

		Unrestricted Weights			**Restricted Weights**		
		CER	$\overline{TW}$	$\sigma(TW)$	CER	$\overline{TW}$	$\sigma(TW)$
	Panel A: $\gamma = 2$						
Uniform	Dyn	-1.0000	23.7044	68.0114	0.0821	1.5796	0.4544
	Myop	-1.0000	20.7321	56.5669	0.0814	1.5683	0.4351
Shrinkage	Dyn	0.2961	10.1712	8.9929	0.0815	1.5617	0.3935
	Myop	0.2752	8.7933	8.0463	0.0808	1.5539	0.3881
	Panel B: $\gamma = 5$						
Uniform	Dyn	0.0769	6.3876	6.8324	0.0645	1.5260	0.4046
	Myop	0.0785	4.5576	4.2663	0.0643	1.5137	0.3901
Shrinkage	Dyn	0.1231	3.3727	1.6823	0.0670	1.5222	0.3468
	Myop	0.1164	2.8407	1.2112	0.0659	1.4967	0.3320
	Panel C: $\gamma = 10$						
Uniform	Dyn	0.0430	2.9673	2.0021	0.0492	1.4826	0.3568
	Myop	0.0552	2.3598	1.2337	0.0452	1.4421	0.3235
Shrinkage	Dyn	0.0661	2.0022	0.6181	0.0500	1.4405	0.2747
	Myop	0.0622	1.7870	0.4561	0.0489	1.4000	0.2421

3. LONG-TERM STRATEGIC ASSET ALLOCATION: AN OUT-OF-SAMPLE EVALUATION

Another important finding is that repeated myopic strategies outperform the theoretically optimal dynamic strategies. Although the average terminal wealth is higher for dynamic strategies, the risk more than proportionally increases. This result implies that the hedging components of dynamic strategies are misspecified and only deteriorate performance. Dynamic strategies are more sensitive to misspecification of any form, since they do not only require us to model the evolution of asset returns correctly, but also of state variables. It is apparently sufficient to only focus on short-term changes in investment opportunities and ignore long-term changes when using the uniform prior.

Secondly, for the shrinkage prior combined with unrestricted weights, a completely different picture arises. The performance for all strategies and all risk aversion levels increases substantially and is much better than for the benchmark strategies. For all risk aversion levels, it pays off to time the bond and stock market. The shrinkage prior makes sure that investors do not take excessive risk. Although the use of the shrinkage estimator reduces average terminal wealth compared with the uniform prior, its standard deviation is more than proportionally reduced. For example, compare the dynamic strategies for an investor with $\gamma = 2$. Although average terminal wealth is reduced with a factor 2.5, its standard deviation is reduced with a factor 8. The result is that the CER for an investors with $\gamma = 2$ is not equal to -100% anymore.

It also turns out that dynamic strategies outperform myopic strategies. Apparently, we are better able to model the hedge component of strategic asset allocations when using the shrinkage prior. The risk for dynamic strategies is still higher but the extra average terminal wealth more than offsets this. In terms of economic performance, the differences between a dynamic and myopic strategy are relatively modest. Different estimation techniques lead to larger performance differences than different strategies.

In order to understand how the shrinkage model works, we plot the realized utility values for risk aversion $\gamma = 5$ using both the shrinkage and the uniform prior against time in panel A of figure 3.3. The figure shows that both series are heavily autocorrelated due to overlapping intervals and that there is a positive correlation between the series. In general, both strategies perform similarly except for a couple of extremely low realized utility values. The shrinkage prior manages to substantially reduce these losses compared to the uniform prior. Hence, the shrinkage prior improves performance by avoiding extreme losses. This is exactly why risk averse investors value this model the most.

Figure 3.3: Realized utility values and stock weights against time for different priors
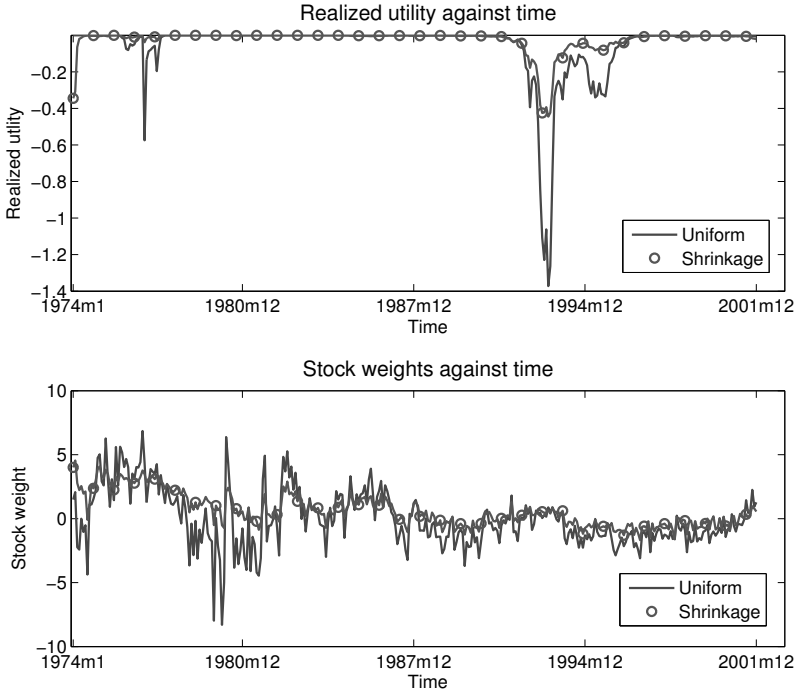
This figure plots realized utility values and stock weights against time for the uniform and shrinkage prior using the plug-in method. We consider a dynamic strategy, $\gamma = 5$ and unrestricted portfolio weights. The x-axis is indexed by the time at which the investors start investing. The second plot shows the stock weights the investors use at the beginning of their investment period.

How does the use of the shrinkage estimator reduce losses? In order to answer this question, we plot the corresponding stock weights of investors against time in panel B of figure 3.3. We plot the weights for investors with a remaining horizon of 60 months. The picture shows that the average weights for both strategies are more or less equal. The weights for the shrinkage prior are, however, much less variable and the portfolio

holdings are much less extreme. An investor who uses the shrinkage prior is still able to time the market. She can still go long in stocks or bonds if market conditions are good and short in stocks or bonds if market conditions are bad. However, the weights are not as extreme anymore and make more sense intuitively. By avoiding overly aggressive market timing, the investor using the shrinkage prior avoids the important extreme events.

The dynamic strategy outperforms the myopic strategy using the shrinkage prior. In order to illustrate this, consider panel A in figure 3.4 which plots the histogram of differences in realized utility values between a dynamic investor and a myopic investor with $\gamma = 5$. Positive values indicate outperformance by the dynamic model. The figure shows that both strategies perform similarly in general. The mass to the right of 0 indicates that most observations give a slight edge to the dynamic strategies. The figure also shows that there are more outliers on the right than on the left. However, differences are not very large.

Finally, let us consider what happens if portfolio weights are restricted to lie between 0 and 1 for all three assets. Restricting portfolio weights leads to a substantial reduction in risk and terminal wealth values for investors using either the uniform or shrinkage prior. It helps to avoid CERs of -100%. This is consistent with results in the previous section and with Jagannathan and Ma (2003). The latter show that restrictions are a form of shrinkage. Hence, in this light it is not surprising that portfolio weight restrictions can improve performance. Note that using shrinkage is substantially better than restricting portfolio weights in order to avoid extreme events.

However, imposing restrictions hurt performance for better performing specifications as it limits the possibilities of investors and leads to much lower CERs. Apparently, going short and very long in assets pays off for long-term investors, especially for those that use the shrinkage prior. Portfolio weight restrictions hurt specifications using the shrinkage prior more than specifications using the uniform prior. On average, the shrinkage prior still outperforms the uniform prior slightly when imposing portfolio restrictions. In all cases, dynamic strategies outperform myopic strategies when restricting portfolio weights. However, differences are smaller than before. The hedge component of strategic asset allocations improves performance only slightly in this setting.

Figure 3.4: Histogram of difference in realized utility: dynamic versus myopic and decision-theory vs plug-in
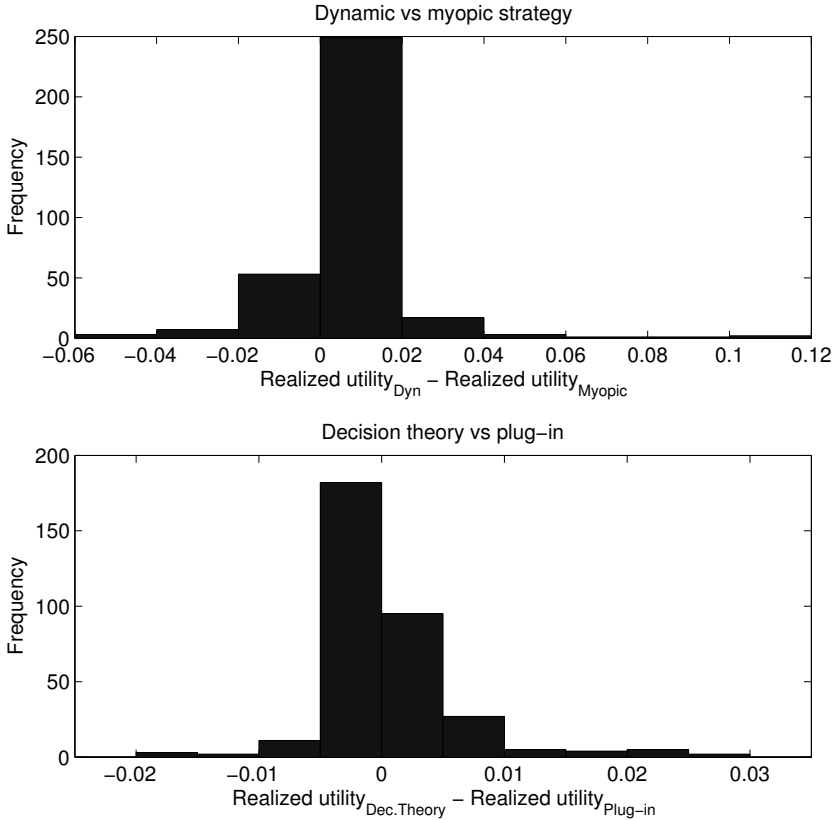
The first figure is a histogram of the difference in realized utility values between a dynamic and myopic strategy, using the unrestricted shrinkage model combined with the plug-in method for an investor with $\gamma = 5$. The second figure gives a histogram of the difference in realized utility values between the decision-theory method and the plug-in method using a restricted dynamic strategy combined with the uniform prior for an investor with $\gamma = 5$.

We conclude that it might not be optimal to time the stock and bond market unless investors use the shrinkage prior. Empirically, differences turn out to be economically important. Using shrinkage avoids extreme portfolio weights and therefore extreme

events. Such a specification is heavily favored by risk-averse investors. Dynamic strate-
gies only work satisfactorily in all cases when using the shrinkage estimator, since
shrinkage leads to a better modeling of the hedge component. However, differences
are economically relatively modest. The effect of the shrinkage prior is the largest
when portfolio weights are unrestricted. Such investors can still go short and very
long in assets without taking excessive risk. When portfolio weights are restricted to
be non-negative, the effect of the shrinkage prior is modest but still positive. Finally,
portfolio weight restrictions help the worse performing specifications, but hurt the best
performing specifications.

3.6.3 Results for the decision-theoretic approach

We consider dynamic and myopic strategies either combined with the uniform or shrink-
age prior. We only report results for restricted portfolio weights as explained in section
3.3.1. Results are given in table 3.5.

Firstly, we compare results with these in previous sections. If we consider dynamic
strategies, the CERs increase slightly when taking parameter uncertainty into account.
Results for myopic strategies are more mixed, but on average results improve when
considering parameter uncertainty. Differences however are again very small. Brandt,
Goyal, Santa-Clara, and Stroud (2005) show by means of simulation that parameter
uncertainty mainly has an impact on the hedging component of a dynamic strategy.
Since this hedging component does not have an important impact on performance
according to results in the previous subsection, it is not surprising that there is only
a small performance difference between the plug-in method and the decision-theoretic
method. The specifications that we consider in this section outperform the benchmark
strategies that ignore predictability by economically important margins, especially for
higher risk aversion levels. This implies that actively timing the bond and stock market
also pays off when taking parameter uncertainty into account.

We illustrate the performance of the decision-theoretic approach in panel B of figure
3.4. This figure plots a histogram of differences in realized utility values between
dynamic strategies using the decision-theoretic and plug-in approach. We set $\gamma = 5$
and use the uniform prior. Positive values indicate outperformance by the decision-
theoretic method. The figure shows that the plug-in approach performs better in most
cases, i.e. the median is slightly negative. However, if the decision-theoretic model

outperforms the plug-in model, the difference is relatively big, illustrated by larger positive values. On average, the decision-theoretic method slightly outperforms the plug-in method. Overall differences are negligible.

Secondly, the table shows that in terms of performance dynamic and myopic strategies are again close to each other with a slight edge for the dynamic strategy. This is again related to a finding in Brandt, Goyal, Santa-Clara, and Stroud (2005) who show

Table 3.5: Decision-theoretic approach - PE model

This table gives the results for the dynamic (Dyn) and myopic (Myop) strategies using the decision-theoretic approach. The results are based on a VAR(1) model with PE as one of the predictors. We report results under the uniform and shrinkage priors and use restricted portfolio weights. We report annualized certainty equivalence returns (CER), average terminal wealth $(\overline{TW})$ and the standard deviation of terminal wealth $(\sigma(TW))$ for three different risk aversion levels γ.

	Restricted Weights		
	Panel A: $\gamma = 2$		
		CER	$\overline{TW}$	$\sigma(TW)$
Uniform	Dyn	0.0830	1.5770	0.4312
	Myop	0.0826	1.5722	0.4227
Shrinkage	Dyn	0.0831	1.5647	0.3729
	Myop	0.0825	1.5596	0.3731
	Panel B: $\gamma = 5$		
Uniform	Dyn	0.0650	1.5313	0.4119
	Myop	0.0652	1.5127	0.3814
Shrinkage	Dyn	0.0682	1.5220	0.3483
	Myop	0.0656	1.4911	0.3255
	Panel C: $\gamma = 10$		
Uniform	Dyn	0.0516	1.4723	0.3372
	Myopic	0.0483	1.4441	0.3182
Shrinkage	Dyn	0.0509	1.4238	0.2571
	Myopic	0.0486	1.4038	0.2565

that the hedge component is relatively small when parameter uncertainty is taken into account. Portfolio weights for dynamic and myopic strategies are therefore close to each other.

Finally, the certainty equivalence returns for specifications involving the shrinkage prior are very close now to specifications using the uniform prior. Apparently, restricting portfolio weights, incorporating parameter uncertainty and using the shrinkage prior leads to portfolios that are a bit too conservative. However, on average the use of the shrinkage prior still outperforms the uniform prior slightly.

We conclude that most results from the previous section stand. Timing the stock and bond market pays off for risk-averse investors. Differences are also economically important. Furthermore, the performance of dynamic and myopic strategies are close to each other with a slight advantage for the dynamic strategies. The performance difference between specifications involving the shrinkage and uniform prior becomes smaller when incorporating parameter uncertainty. Shrinkage is economically less important in such a setting where we also restrict portfolio weights. Finally, incorporating parameter uncertainty leads to specifications with slightly higher certainty equivalence returns.

3.7 Robustness analysis

In this section, we perform some additional robustness tests. In the first subsection, we perform classical tests on the performance differences between different specifications. The second subsection investigates the performance differences between myopic and dynamic strategies. The last subsection considers a model with the dividend-to-price ratio as one of the predictor variables.

3.7.1 Classical significance tests

We investigate the classical statistical significance of the results by comparing the strategies of sections 3.6.2 and 3.6.3 with the benchmark strategies in section 3.6.1 in a repeated samples context. We test whether the difference in average realized utility between a strategy and its benchmark is statistically different from zero. As a benchmark, we take the no-predictability strategies of section 3.6.1 either unrestricted or restricted and either dynamic or myopic, depending on the context. In other words,

we test whether the extra value of market timing we find in previous sections might be spurious.

We use utility as the loss function of forecasts (after implementing strategies). In the forecasting literature, tests of equal forecasting performance are standard and we use the Diebold and Mariano (1995) test on the utility series. Diebold and Mariano (1995) generate the difference series of two forecasts and test whether this difference is equal to zero by means of a standard t-ratio. They show that this test statistic has a standard normal distribution. We estimate the covariance matrix of average realized utility non-parametrically by means of the Newey and West (1987) HAC estimator. In order to choose the lag length, we use the Newey and West (1994) lag length selection criterium.

Table 3.6 presents results. The performance of unrestricted plug-in strategies based on the uniform prior is only significantly different from its benchmark in one case. If we use the shrinkage prior instead, we see that these strategies become significant. Hence, the impressive performance for the unrestricted plug-in methods based on the shrinkage prior is not spurious. It is statistically different from its benchmark. Results

Table 3.6: Classical significance tests

This table presents classical t-statistics to test whether the performance of the portfolio strategy and its benchmark are statistically significant from each other. We give results for the plug-in approach, the decision-theoretic approach, different risk aversion levels, different types of strategies and for different weight restrictions.

		$\gamma = 2$		$\gamma = 5$		$\gamma = 10$	
		Unr	Restr	Unres	Restr	Unr	Restr
Panel A: plug-in approach							
Uniform	Dyn	n/a	0.9208	1.1372	2.7858	1.1683	3.7736
	Myopic	n/a	0.8615	1.7996	2.6184	2.5115	3.4862
Shrinkage	Dyn	4.3796	0.7716	3.6705	2.4392	2.6949	3.5770
	Myopic	4.2474	0.7136	3.9038	2.2459	3.4290	3.5485
Panel B: decision-theoretic approach							
Uniform	Dyn		0.9757		2.7631		3.7395
	Myopic		0.9476		2.6611		3.7302
Shrinkage	Dyn		0.8541		2.4885		3.4696
	Myopic		0.8039		2.2149		3.5315

Table 3.7: Monte Carlo simulation without predictability

This table gives results for 100 Monte Carlo simulations assuming no predictability. The specifications differ in the strategy (dynamic or myopic) and in the risk aversion level. The entries in panel A and B are respectively defined as $CER_{Uni} - CER_{Nopred}$ and $CER_{Shr} - CER_{Nopred}$. Data indicates the result found in the actual data-set (based on table 3.4). Mean, median, min, 1st, 5th, 95th, 99th and max respectively indicate the average difference in CERs, median difference, minimum difference, 1st percentile of differences, 5th percentile of differences, 95th percentile of difference, 99th percentile of difference and maximum difference in the Monte Carlo simulations.

	Dynamic			Myopic		
	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
Panel A: Uniform prior						
Data	-1.0886	0.0291	0.0165	-1.0855	0.0342	0.0305
Mean	-0.1965	-0.0567	-0.0264	-0.1325	-0.0387	-0.0176
Median	-0.1397	-0.0524	-0.0237	-0.1090	-0.0339	-0.0154
Min	-1.2249	-0.2571	-0.1394	-1.0669	-0.1481	-0.0745
1st	-1.1952	-0.2559	-0.1337	-0.7526	-0.1477	-0.0732
5th	-0.8486	-0.1782	-0.0840	-0.3667	-0.1171	-0.0556
95th	0.0062	0.0242	0.0162	0.0079	0.0108	0.0069
99th	0.0828	0.0438	0.0243	0.0956	0.0535	0.0298
Max	0.1101	0.0456	0.0283	0.1455	0.0649	0.0314
Panel B: Shrinkage prior						
Data	0.2075	0.0753	0.0396	0.1897	0.0721	0.0375
Mean	-0.0720	-0.0261	-0.0121	-0.0576	-0.0186	-0.0084
Median	-0.0629	-0.0227	-0.0109	-0.0589	-0.0181	-0.0061
Min	-0.3270	-0.1586	-0.0820	-0.2227	-0.0896	-0.0458
1st	-0.3070	-0.1453	-0.0737	-0.2178	-0.0855	-0.0434
5th	-0.2081	-0.1035	-0.0543	-0.1614	-0.0647	-0.0317
95th	0.0392	0.0284	0.0159	0.0413	0.0213	0.0116
99th	0.1045	0.0419	0.0248	0.1151	0.0539	0.0294
Max	0.1265	0.0422	0.0261	0.1540	0.0620	0.0297

are different for specifications that restrict portfolio weights. Results for the dynamic and myopic strategies are significant except for low risk aversion levels. Apparently, a low risk averse investor is especially hurt when weights are restricted.

In order to take issues such as autocorrelation and skewness in the realized utility series into account as well as to correct for the fact that the benchmark strategies are based on nested models, we also perform an additional Monte Carlo simulation. We generate 100 time-series of asset returns and predictor variables under the null of no predictability. The DGP is based on the parameter estimates obtained using the no-predictability prior on the full data-set. In every Monte Carlo simulation, we generate a time-series of 52 years of asset returns and state variables and perform the same out-of-sample analysis as on the real data-set. In order to make the Monte Carlo analysis feasible, we only consider specifications with unrestricted portfolio weights.[1]

Table 3.7 shows the results. Panel A reports the difference in CERs between the uniform prior and the no-predictability prior in the simulations. A positive difference implies outperformance by the uniform prior. We find that the average difference is negative which means that the no-predictability strategies perform better on average. This is not surprising, since the data is generated using the no-predictability prior on the full data-set. When comparing the differences found in the data with the different percentiles in the simulations, the results in the data are not significantly different from the results in the simulations that are generated under the null of no predictability.

Panel B shows the difference in CERs between the shrinkage prior and the no-predictability prior. The average difference is negative, but less negative than the average difference for the uniform prior. The table shows that the positive differences in the data are in all cases larger than the maxima in the 100 Monte Carlo simulations. Hence, the largest performance difference in the DGP without predictability is less than we find in the data. Therefore, the data contains predictability.

3.7.2 Difference between dynamic and myopic strategies

The results in previous sections show that there is hardly a difference in CERs between dynamic and (repeated) myopic strategies. In order to understand this result, figures 3.5 and 3.6 plot the posterior distribution of stock weights for respectively the uniform and

---

[1]Specifications involving restricted portfolio weights take approximately half a day to calculate. Repeating this 100 times is not feasible.

Figure 3.5: Posterior distribution of stock weights using the uniform prior
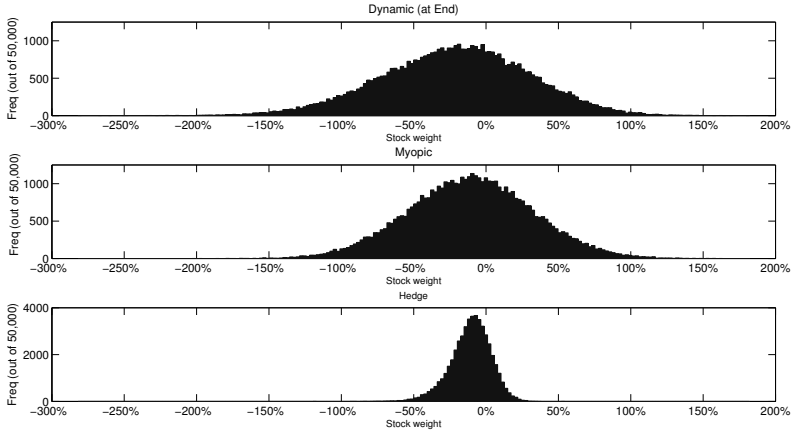
The figure shows the posterior distribution of stock weights for the dynamic strategy, the myopic strategy and the hedge component of the dynamic strategy. The horizon is 60 months, $\gamma = 5$ and state variables are equal to their values at the end of the sample. The figure is obtained by drawing 50,000 times from the posterior distribution using the uniform prior. Stock weights are calculated for every draw using the Jurek and Viceira (2010) solution.

shrinkage prior at the end of the sample for the dynamic strategy, the myopic strategy and the hedge component. We consider an investor with $\gamma = 5$ and an investment horizon of 60 months.

Figure 3.5 shows that estimation uncertainty plays a large role for the specification based on the uniform prior even when using the full sample. The standard deviation of dynamic weights (myopic weights) is 52% (43%). In other words, parameter draws that are almost as likely as the posterior mean could lead to completely different asset weights. Figure 3.6 indicates that the use of the shrinkage prior reduces the standard deviation to 41% (35%). However, it also shows that there is still a large amount of estimation uncertainty left.[1] Both figures suggest that estimation uncertainty is the

---

[1]At the end of the sample, the data dominates the prior and therefore the specifications based on different priors are relatively close to each other. If we would only use say 20 years of data, the impact

Figure 3.6: Posterior distribution of stock weights using the shrinkage prior
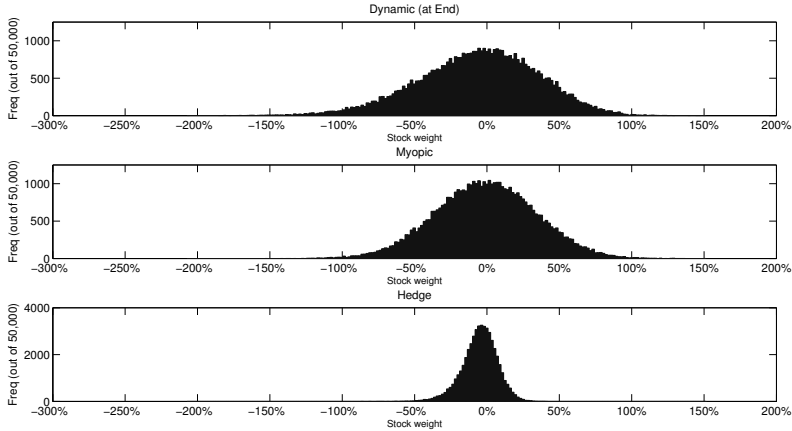
The figure shows the posterior distribution of stock weights for the dynamic strategy, the myopic strategy and the hedge component of the dynamic strategy. The horizon is 60 months, $\gamma = 5$ and state variables are equal to their values at the end of the sample. The figure is obtained by drawing 50,000 times from the posterior distribution using the shrinkage prior. Stock weights are calculated for every draw using the Jurek and Viceira (2010) solution.

highest for the dynamic strategy.

Next, we perform a Monte Carlo simulation under the null of predictability. The DGP is based on the parameter estimates obtained using the uniform prior on the full data-set. In every simulation, we generate 52 years of data and perform the same out-of-sample analysis as on the real data-set. We base the portfolio weights on the true parameters that we use to simulate the data (this is obviously infeasible in reality) or on estimated parameters based on either the uniform prior or on the shrinkage prior. To make the analysis feasible, we only consider specifications with unrestricted portfolio weights. We use 100 simulations.

The differences in CERs between a dynamic strategy and a myopic strategy are

---

of the prior and therefore the difference in estimation uncertainty would be much larger.

Table 3.8: Monte Carlo simulation with predictability

This table gives results for 100 Monte Carlo simulations assuming predictability. We compare $(CER_{dyn} - CER_{myop})$ for specifications using either the true parameters (panel A), the estimated parameters under a uniform prior (Panel B) or the estimated parameters under a shrinkage prior (Panel C). Secondly, we give the average differences (over the simulations) in root mean squared portfolio weight error between the dynamic and myopic strategies: $\overline{\Delta RMSPE_{w_s}} = RMSPE_{w_s,dyn} - RMSPE_{w_s,myopic}$ and $\overline{\Delta RMSPE_{w_b}} = RMSPE_{w_b,dyn} - RMSPE_{w_b,myopic}$. Specifications differ in their risk aversion level. Data indicates the result found in the actual data-set.

	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
Panel A: True			
Mean	0.0146	0.0179	0.0111
Median	0.0200	0.0231	0.0124
Min	-0.2904	-0.1237	-0.0551
1st	-0.2654	-0.0717	-0.0335
5th	-0.0275	-0.0072	-0.0013
95th	0.0541	0.0386	0.0225
99th	0.0696	0.0475	0.0290
Max	0.0785	0.0508	0.0311
Panel B: Uniform			
Data	0.0000	-0.0016	-0.0122
Mean	-0.1649	-0.0037	0.0023
Median	-0.0171	0.0100	0.0098
Min	-2.0126	-0.1805	-0.1062
1st	-1.9797	-0.1794	-0.1057
5th	-1.4106	-0.1057	-0.0609
95th	0.0723	0.0547	0.0375
99th	0.1548	0.0790	0.0479
Max	0.1878	0.0842	0.0498
$\overline{\Delta RMSPE_{w_s}}$	0.2113	0.2232	0.1628
$\overline{\Delta RMSPE_{w_b}}$	0.3210	0.3321	0.2442
Panel C: Shrinkage			
Data	0.0209	0.0067	0.0039
Mean	-0.0041	-0.0006	0.0002
Median	0.0084	0.0075	0.0043
Min	-0.2320	-0.1430	-0.0857
1st	-0.2265	-0.1362	-0.0773
5th	-0.0940	-0.0740	-0.0413
95th	0.0443	0.0352	0.0228
99th	0.0656	0.0499	0.0307
Max	0.0730	0.0546	0.0339
$\overline{\Delta RMSPE_{w_s}}$	0.0029	0.0040	0.0422
$\overline{\Delta RMSPE_{w_b}}$	-0.2874	-0.2454	-0.1641

given in table 3.8. A positive difference implies that the dynamic strategy outperforms the repeated myopic strategy. In panel A, we give results for strategies that are based on the true parameters. The average difference is positive, which indicates outperformance by the dynamic strategy. This is not surprising, since the dynamic strategy should be the optimal strategy when the investor knows the true DGP exactly. Remarkably, in some simulations the difference is negative even if true parameter values are used.

Panel B shows results for an investor who has to estimate the parameters using the (simulated) data and the uniform prior. The performance differences that we find in the data are in line with the differences in the Monte Carlo simulations. For investors with low risk aversion, the dynamic strategy is on average inferior to the myopic strategy. Apparently, even if the true DGP contains predictability, a dynamic strategy is not necessarily better than a myopic strategy when the parameters need to be estimated. Panel C gives similar results using the shrinkage prior. These results are in line with the results in panel B.

Why do the estimated myopic strategies perform as well as the estimated dynamic strategies? In every simulation and in every period, we can calculate the difference between the optimal dynamic portfolio weights (based on the true parameters) and the estimated portfolio weights (either myopic or dynamic). This allows us to calculate the root mean squared portfolio weight error (RMSPE) for both the stock weights as well as the bond weights for every simulation. We calculate the average RMSPE over all simulations. This is a measure of how far the estimated portfolio weights are from the true optimal dynamic portfolio weights. We compare both the estimated myopic and estimated dynamic weights with the true optimal dynamic weights. We do this for investors with a remaining investment horizon of 60 months.

Lines 10 and 11 of panel B and C show the differences in average RMSPE between the estimated dynamic and myopic weights. A positive number implies that the estimated myopic weight is closer to the optimal dynamic portfolio weight. For the uniform prior, the estimated myopic portfolio weights are indeed closer to the optimal dynamic portfolio weights. This holds for both stocks and bonds. The estimated myopic weights approximate the optimal portfolio better than the estimated dynamic weights due to the large estimation error in the long-run predictions of returns and in the covariances with current returns. This explains why the repeated myopic weights outperform dynamic strategies in the data when using the uniform prior.

A closer look at terminal wealth values shows that the portfolio weights using the uniform prior are too aggressive, i.e. the portfolios are too risky. For example considering $\gamma = 5$, the average (over the simulations) of the average terminal wealth (standard deviations of terminal wealth) is 7.81 (5.90) for the dynamic and 7.02 (5.08) for the myopic strategy using the true parameters. Using the estimated parameters under the uniform prior, we respectively get 8.76 (8.39) and 7.01 (5.76). These results show that both the estimated myopic as well as the estimated dynamic strategy are (way) too risky and aggressive. However, by being too aggressive, the estimated myopic portfolio weights move towards the optimal dynamic weights and approximates the true optimal dynamic strategy better.

Panel C shows that results are more mixed for the shrinkage prior. The estimated myopic strategy approximates the optimal dynamic strategy slightly better for stock weights, but much worse for bond weights. A closer look at terminal wealth values shows that portfolio weights based on the shrinkage prior are more conservative than weights based on the true parameters or on the uniform prior, since investors that use the shrinkage prior are more skeptical about predictability.

3.7.3 Using dividend-to-price ratio as a predictor

In section 3.6, we use the price-earnings ratio as one of the predictor variables. Another commonly used predictor variable is the dividend-to-price ratio. In this section, we give results for the plug-in and the decision-theoretic approach for a model in which the dividend-to-price ratio replaces the price-earnings ratio.

Table 3.9 shows results for the plug-in method. Firstly, we consider the plug-in method combined with unrestricted weights. Again, the performance for an investor with low risk aversion ($\gamma = 2$), unrestricted portfolio weights and the uniform prior is very bad with a CER of -100%. This time, however, the performance for higher risk aversion levels is very bad as well, i.e. CERs are often negative and are substantially lower than the ones for the benchmark models. The DP model is apparently misspecified.

Under the shrinkage prior, results substantially improve. Negative CERs become positive and benchmark models are outperformed. Investors should again time the stock and bond market when using the shrinkage prior. Despite the misspecified DP model, differences are still economically important. Next, dynamic strategies still outperform

myopic strategies when using the shrinkage prior. However, the performance differences is quite small.

The table shows that we could have restricted portfolio weights as well, instead of shrinkage, to improve out-of-sample performance for all risk aversion levels. Apparently, the misspecified DP model only gives acceptable out-of-sample results when using some form of shrinkage: either by using a shrinkage prior or by restricting portfolio weights. The results in table 3.9 indicate that double shrinkage does not work well for the DP model. Combining the shrinkage estimator with restricted portfolio weights deteri-

Table 3.9: Plug-in approach - DP model

This table gives the results for the dynamic (Dyn) and myopic (Myop) strategies using the plug-in method. We use a VAR(1) model with DP as one of the predictors for a robustness check. We report results under the uniform and shrinkage priors and either use restricted or unrestricted portfolio weights. We report annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three risk aversion levels γ.

		Unrestricted Weights			**Restricted Weights**		
		CER	$\overline{TW}$	$\sigma(TW)$	CER	$\overline{TW}$	$\sigma(TW)$
	Panel A: $\gamma = 2$						
Uniform	Dyn	-1.0000	22.7307	58.7851	0.0755	1.4929	0.3081
	Myop	-1.0000	17.7485	42.5112	0.0758	1.4941	0.3061
Shrinkage	Dyn	0.2192	4.5321	3.1417	0.0682	1.4351	0.2661
	Myop	0.2041	4.0242	2.7603	0.0683	1.4357	0.2647
	Panel B: $\gamma = 5$						
Uniform	Dyn	-0.0857	5.5379	5.4991	0.0661	1.5109	0.3324
	Myop	0.0010	3.9889	3.5361	0.0647	1.5024	0.3532
Shrinkage	Dyn	0.0946	2.1192	0.6630	0.0588	1.4420	0.2827
	Myop	0.0883	1.9085	0.5513	0.0552	1.3956	0.2449
	Panel C: $\gamma = 10$						
Uniform	Dyn	-0.0500	2.7785	1.8400	0.0499	1.4814	0.3335
	Myop	0.0139	2.2197	1.1653	0.0474	1.4365	0.3177
Shrinkage	Dyn	0.0521	1.5547	0.2946	0.0414	1.3587	0.2241
	Myop	0.0482	1.4506	0.2403	0.0376	1.3080	0.1914

orates results for all cases. Again, portfolio weight restrictions help bad-performing specifications but hurt good-performing specifications.

Table 3.10 shows analogous results for the decision-theoretic method. Incorporating parameter uncertainty improves performance for the least risk-averse investors using the uniform prior and for all investors using the shrinkage prior. Performance however deteriorates for more risk-averse investors that use the uniform prior. Again, dynamic

Table 3.10: Decision-theoretic approach - DP model

This table gives the results for the dynamic (Dyn) and myopic (Myop) strategies using the decision-theoretic approach. We use a VAR(1) model with DP as one of the predictors as a robustness check. We report results under the uniform and shrinkage priors and use restricted portfolio weights. We report annualized certainty equivalence returns (CER), average terminal wealth ($\overline{TW}$) and the standard deviation of terminal wealth ($\sigma(TW)$) for three risk aversion levels γ.

	Restricted Weights		
	Panel A: $\gamma = 2$		
		CER	$\overline{TW}$	$\sigma(TW)$
Uniform	Dyn	0.0756	1.4939	0.3084
	Myop	0.0759	1.4947	0.3041
Shrinkage	Dyn	0.0723	1.4620	0.2655
	Myop	0.0730	1.4680	0.2726
	Panel B: $\gamma = 5$		
Uniform	Dyn	0.0647	1.5085	0.3448
	Myop	0.0636	1.4923	0.3443
Shrinkage	Dyn	0.0599	1.4446	0.2778
	Myop	0.0570	1.4102	0.2530
	Panel C: $\gamma = 10$		
Uniform	Dyn	0.0476	1.4732	0.3434
	Myop	0.0471	1.4320	0.3142
Shrinkage	Dyn	0.0413	1.3515	0.2191
	Myop	0.0382	1.3181	0.2039

strategies marginally outperform myopic strategies in this setting.

We conclude that results from previous sections are confirmed. Investors should actively time the stock and bond market, including hedging components only marginally improves performance and the shrinkage prior leads to superior results. However, using double shrinkage deteriorates results in this section.

3.8 Conclusion

We investigate the out-of-sample performance of strategic asset allocations. Our aim is to evaluate if the potential gains from strategic portfolios can be realized out-of-sample. Optimal strategic portfolios are time-varying and include a hedge component. We analyze the importance of both aspects. Furthermore, we introduce a shrinkage prior that downplays the predictability of asset returns and shrinks the model for the predictor variables to a random walk. We investigate whether the shrinkage prior leads to better results for long-term investors. In our analysis, we consider several specifications. We vary the method (plug-in or decision-theoretic), the estimator (uniform prior or shrinkage prior), the strategy (myopic or dynamic) and the portfolio constraints (constrained or unconstrained) for risk aversion levels γ is 2, 5 or 10.

The first important characteristic of optimal strategic portfolios is that they are time-varying. We find that this potential gain can be realized out-of-sample. Long-term investors should let their asset allocations depend on market conditions when they use our proposed shrinkage prior. Their allocations outperform strategies that ignore asset return predictability by margins that are economically (and statistically) significant. The shrinkage prior makes sure that weights are not wildly fluctuating and not too extreme. The standard uniform prior on the other hand does not give satisfactory results. An investor with low risk aversion would have lost all her money if she would have relied on a VAR model estimated with a uniform prior.

Our analysis shows that it is very important for investors to evaluate a prediction model using an asymmetric utility metric. Risk-averse investors value models by their capability of avoiding a disaster (the extreme negative events). It turns out that the shrinkage prior does exactly this. Investors that use the shrinkage prior can still time the market and benefit from good market conditions. However, what distinguishes the

shrinkage prior from the standard uniform prior is that it is capable of limiting the losses in extreme negative events.

The second important characteristic of optimal strategic portfolios is the hedge component. We argue that this component is sensitive to estimation error in both long-run predictions of returns and in their covariance with current returns. Our analysis shows that its potential gain translates into only a modest extra performance out-of-sample. In some cases dynamic portfolios outperform repeated myopic portfolios by economically relevant margins (especially if the shrinkage prior is used), but in general differences are not very large. Monte Carlo simulations show that this result is indeed caused by estimation error. Estimated portfolios are more aggressive than their population counterparts. By being more aggressive, the estimated myopic portfolio moves towards the true (unknown) optimal dynamic portfolio. The estimated dynamic portfolio on the other hand moves away from the optimal portfolio. In the data, both rules approximate the true optimal portfolio almost equally well.

The specifications we consider in the paper also differ in the method and in the restrictions imposed. Some additional results are the following. Taking parameter uncertainty into account leads to very modest improvements over methods that only condition on parameter estimates. Brandt, Goyal, Santa-Clara, and Stroud (2005) among others show that incorporating parameter uncertainty does not significantly alter the weights for myopic portfolios. It has a much bigger impact on weights of the hedge component. Our analysis shows that this hedge component only leads to a modest improvement over myopic portfolios in general. In this light, it it not surprising that the incorporation of parameter uncertainty does not lead to a much better performance in this particular case. The effect of weight restrictions on performance is ambiguous. Badly performing specifications perform better if weights are restricted. However, the best performing specifications are hurt if weights are restricted. Hence, restrictions help bad models and hurt the good models as is commonly the case.

A risk-averse investor should combine the shrinkage prior with the plug-in method and unrestricted weights to maximize her expected utility. She should time the stock and bond market. Such an investor increases performance slightly by combining the shrinkage prior with a dynamic strategy in order to take the hedge component into account.

We can further extend our paper in several directions. Firstly, we do not take model uncertainty into account. We assume that investors only use one set of predictor variables. An alternative would be to use model selection criteria or Bayesian model averaging (see Cremers (2002) and Avramov (2002)). We do however investigate the sensitivity of performance with respect to the choice of another predictor variable. Secondly, the data generating process (DGP) of asset return and state variable dynamics is assumed not to change over time. We do not consider time-varying parameters or regime-switching models. Next, we ignore realistic aspects such as labor income, liabilities or transaction costs that matter in reality. Finally, we ignore hedging against learning due to infeasibility of the computations involved. Brandt, Goyal, Santa-Clara, and Stroud (2005) show that incorporating learning might improve certainty equivalence returns even further. A challenging task for future research will be to develop a solution method that is capable of incorporating learning in a large VAR model such as ours.

3.9   Appendix

3.9.1   Posterior distribution and MCMC algorithm

This section gives details on how we simulate from the posterior and predictive distribution for both the uniform prior, introduced in equation (3.16), and the shrinkage prior, introduced in equation (3.17). The posterior mean for the no-predictability prior is derived from the results for the uniform prior.

We first consider the uniform prior given in equation (3.16). The posterior distribution is as follows

$$P(B^*, \Sigma|Y) \propto |\Sigma|^{-(T+n+1)/2} \exp\left\{-\frac{1}{2}tr\left[(Y^* - XB^{*\prime})'(Y^* - XB^{*\prime})\Sigma^{-1}\right]\right\}. \quad (3.18)$$

It is well-known in the literature (e.g. Zellner (1971)) that the above posterior is the product of the marginal posterior distribution for Σ and the conditional posterior distribution for B^*. These distribution functions look as follows

$$P(\Sigma|Y) = iWishart\,(S, T - n - 1) \qquad (3.19)$$

$$P(\beta^{*\prime}|\Sigma, Y) = N\left(\hat{\beta}^{*\prime}, \Sigma \otimes (X'X)^{-1}\right), \qquad (3.20)$$

where $\beta^{*\prime}$ and $\hat{\beta}^{*\prime}$ are equal to vectorized $B^{*\prime}$ and $\hat{B}^{*\prime} = (X'X)^{-1}X'Y^*$ respectively, and $S = (Y^* - X\hat{B}')'(Y^* - X\hat{B}')$. We can simulate from the above posterior by first drawing Σ from the inverse Wishart distribution and then drawing $\beta^{*\prime}$ given Σ from the multivariate normal distribution.

If we impose the assumption of stationarity, it is not possible to derive an analytical expression for the marginal posterior for Σ by integrating with respect to B^* over its stationarity region. This implies that we have to rely on a Gibbs sampler with the conditional posteriors $\beta^{*\prime}|\Sigma$, given in equation (3.20), and $\Sigma|\beta^{*\prime}$. The latter distribution is an inverted Wishart distribution where S in equation (3.19) depends on B^* instead of $\hat{B}^*$ and the degrees of freedom are equal to T instead of $T - n - 1$. We use rejection sampling in order to impose stationarity, i.e. we reject draws for B^* that would result in a non-stationary model.

Secondly, consider the shrinkage prior given in equation (3.17). The posterior distribution is given in the following equation

$$P(B^*, \Sigma|Y) \propto \left(b^{*\prime}b^*\right)^{\frac{-(n(n+1)-2)}{2}} |\Sigma|^{-(T+n+1)/2} \exp\left\{-\frac{1}{2}tr\left[(Y^* - XB^{*\prime})'(Y^* - XB^{*\prime})\Sigma^{-1}\right]\right\},$$
$$(3.21)$$

The above posterior does not belong to a known distribution class. Ni and Sun (2003) develop an algorithm that allows us to simulate from the posterior distribution. In order to do so, they introduce a latent variable δ which is needed to simulate B^*. We use a Gibbs sampler, where the following conditional distributions are important

$$P(\Sigma|B^*, Y) = iWishart\left((Y^* - XB^{*'})'(Y^* - XB^{*'}), T\right) \tag{3.22}$$

$$P(\delta|B^*, Y) = iGamma\left(J/2 - 1, \frac{1}{2}\beta^{*'}\beta^*\right) \tag{3.23}$$

$$P(B^*|\delta, \Sigma, Y) = N\left(\delta(\Sigma \otimes (X'X)^{-1} + \delta I_J)^{-1}\hat{\beta}^*, (\Sigma^{-1} \otimes X'X + \frac{1}{\delta}I_J)^{-1}\right) \tag{3.24}$$

with $J = n(n+1)$ and I_J the identity matrix of dimension J. We can simply impose the assumption of stationarity by rejecting non-stationary draws as explained above. In order to increase the accuracy of point estimates, we use Rao-Blackwellization techniques if possible. This means that we average conditional means of the parameter draws in order to obtain the (un)conditional posterior means instead of averaging drawn parameter values.

No matter whether we use the uniform or shrinkage prior, we can simulate from the predictive distribution once we have a sample of simulated parameter values. This conditional distribution is given as follows

$$P(y_{t+1}|y_t, B^{(i)}, \Sigma^{(i)}) = N(B_0^{(i)} + B_1^{(i)}y_t, \Sigma^{(i)}), \tag{3.25}$$

where $B_0^{(i)}$, $B_1^{(i)}$ and $\Sigma^{(i)}$ are drawn parameter values. Note that we first have to transform $B_1^{*(i)}$ into $B_1^{(i)}$ before we are able to simulate future values of y_t. We use antithetic sampling. This means that we simulate two antithetic scenarios of future returns and state variables for each parameter draw. It is a more efficient and accurate way to simulate from the predictive distribution.

We use the ML estimates for the initialization of the Gibbs samplers. We draw 25,000 parameter estimates in total, but discard the first 5,000 draws. This results in 40,000 asset return and state variable paths. Increasing the burn-in phase or the number of simulations does not significantly impact the results. Visual inspection of the posterior draws, CUMSUM statistics proposed in Bauwens, Lubrano, and Richard (1999) and the equality of means test proposed in Geweke (2005) suggest that estimates converge.

3.9.2 Numerical method

This section elaborates on the numerical methods used in this paper. We show how the parametrization of regression coefficients works and give an indication of the accuracy of our methods. Our method is based on the observation made in Koijen, Nijman, and Werker (2010) in a different setting that the regression coefficients in step 4 of section 3.5.2 have to be a function of portfolio weights and can be parameterized. This works extremely well in our setting. For the empirical illustrations in this section, we estimate the PE model on the full data-set and assume that the estimates are the true values. Allowing for parameter uncertainty does not change conclusions in this section.

For simplicity, assume that we want to maximize power utility over terminal wealth one month in the future

$$\max_{w_t} E\left(\frac{W_{t+1}^{1-\gamma}}{1-\gamma} \middle| Z_t \right). \tag{3.26}$$

In the main paper, the conditioning variables in Z_t are equal to the asset returns and predictor variables in y_t. For illustration purposes, we set the conditioning variables equal to their historical average in this section. The standard approach for solving this problem is to set up a portfolio weight grid and simulate N asset return paths. Then, take a grid point, calculate realized utility for all paths and calculate conditional expected utility for this grid point by averaging the realized utility values. Finally, repeat this for all grid points and pick the portfolio weight that maximizes conditional expected utility.

Since different portfolio weights lead to different conditional utilities, conditional utility obviously has to be a function of portfolio weights. We illustrate this fact in figure 3.7 where we plot conditional utility versus the portfolio weights. The picture clearly shows a quadratic relation. In fact, if we regress conditional utility on a quadratic function of portfolio weights we get an R^2 near 1. Hence, the following holds

$$\max_{w_t} E\left(\frac{W_{t+1}^{1-\gamma}}{1-\gamma} \middle| Z_t \right) = \max_{w_t} f(w_t), \tag{3.27}$$

where $f(w_t)$ is a quadratic function in the portfolio weight w_t.

In other words, maximizing conditional expected utility on a constrained set is equivalent to maximizing a quadratic function on this same set. This can be done analytically. Since the R^2 in the parametrization regression is almost 1, we do not have

Figure 3.7: Conditional utility versus portfolio weights
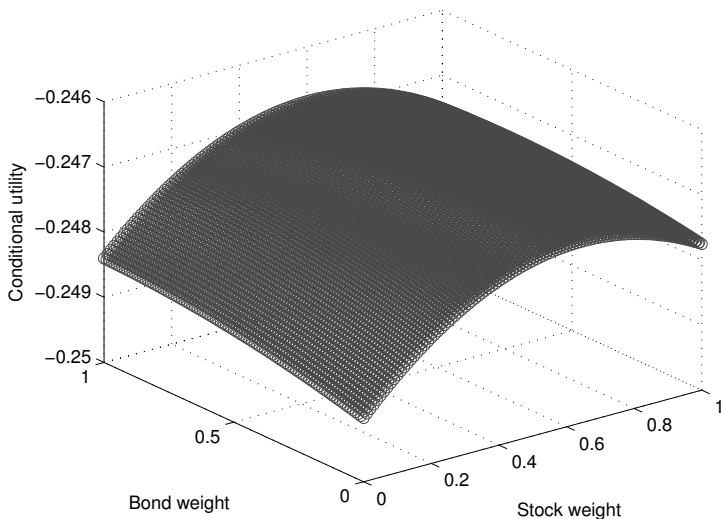
This figure plots conditional utility over terminal wealth against the portfolio weight in stocks and the portfolio weight in bonds. We impose short-selling constraints which implies that only the subregion for which weights add up to 1 is feasible.

to estimate this parametrization regression on a very fine grid: knowing a couple of points is enough. Hence, the high R^2 of 1 is the crucial here.

The figure above clearly illustrates the general idea. However, the method is much more general. We can easily generalize the above to a dynamic setting where the conditional utility depends on conditioning variables. As an illustration, assume that the conditional expectation of the value function at time t depends on one conditioning variable Z_t:

$$E\left\{V_{t+1}\left(K-1, W_{t+1}, Z_{t+1}\right) \mid Z_t\right\} = \alpha_{0w_t} + \alpha_{1w_t} Z_t, \qquad (3.28)$$

where α_{0w_t} and α_{1w_t} are coefficients depending on portfolio weights w_t. If we parameterize both coefficients in a quadratic function of portfolio weights $w_{t,s}$ for stocks and

$w_{t,b}$ for bonds depending on coefficient vectors γ_0 and γ_1, we get

$$
\begin{aligned}
E(.. \mid Z_t) = {} & (\gamma_{00} + \gamma_{10}w_{t,s} + \gamma_{20}w_{t,b} + \gamma_{30}w_{t,s}^2 + \gamma_{40}w_{t,b}^2 + \gamma_{50}w_{t,s}w_{t,b}) + \\
& (\gamma_{01} + \gamma_{11}w_{t,s} + \gamma_{21}w_{t,b} + \gamma_{31}w_{t,s}^2 + \gamma_{41}w_{t,b}^2 + \gamma_{51}w_{t,s}w_{t,b})Z_t
\end{aligned}
\tag{3.29}
$$

$$
\begin{aligned}
E(.. \mid Z_t) = {} & (\gamma_{00} + \gamma_{01}Z_t) + (\gamma_{10} + \gamma_{11}Z_t)w_{t,s} + (\gamma_{20} + \gamma_{21}Z_t)w_{t,b} + \\
& (\gamma_{30} + \gamma_{31}Z_t)w_{t,s}^2 + (\gamma_{40} + \gamma_{41}Z_t)w_{t,b}^2 + (\gamma_{50} + \gamma_{51}Z_t)w_{t,b}w_{t,s},
\end{aligned}
\tag{3.30}
$$

where the second equality follows after collecting terms. Along each path, the conditioning variables are known. Therefore, maximizing the above conditional expectations boils down to maximizing a quadratic function in portfolio weights where conditioning variables can be treated as constants.

The quality of the quadratic approximation depends crucially on the fit of the parametrization regressions. If one of the coefficients cannot be accurately parameterized in the portfolio weights, the approximation would already break down. Fortunately, this second-order approximation of the regression parameters on the portfolio weights is very accurate, i.e. the R^2's of these parametrization regressions are all larger than 0.999 for all coefficients. This holds for settings with and without parameter uncertainty, for $\gamma = 2$, $\gamma = 5$ and $\gamma = 10$, for short and for long horizons and does not depend on how we select the initial grid points. The fact that the R^2 are near one is the main motivation why we can use a global quadratic approximation.

It is important to note that the rebalancing frequency should be sufficiently high to obtain accurate results. In our setting, where investors have to choose portfolio weights every month, the method gives extremely accurate results. However, if we do not consider any intermediate rebalancing (a buy-and-hold strategy with an investment horizon of 5 years), the R^2 of the parametrization regressions decreases considerably to around 80% and the quadratic approximation breaks down. Note that there is some equivalence with the method of Campbell, Chan, and Viceira (2003) and Jurek and Viceira (2010). Firstly, the method they use is mainly accurate for high rebalancing frequencies due to - in their case - the lognormal approximation of portfolio returns. Secondly, they find that the - in their case - *log* of the (optimal) value function is quadratic in optimal portfolio weights.

In the empirical section in the paper we use 6 conditioning variables. The grid size is only 10 and the number of paths is equal to 40,000. We use a first order polynomial of the conditioning variables, refer to step 4 in section 3.5.2, and a second

order approximation in the parametrization regressions, refer to step 5. Note that this numerical method is very fast, since we only have to consider a grid size of 10 instead of more than 5,000.[1] Larger grid sizes do not influence the results because of the high R^2.

Van Binsbergen and Brandt (2007) show that their method is accurate by comparing their method with the method of Barberis (2000). Their results are similar and therefore these authors conclude that their method is accurate. We provide additional evidence that our method is accurate by comparing our numerical method with the one used in van Binsbergen and Brandt (2007). We report results in table 3.11.[2]

From the table it is clear that the two methods are equally accurate, i.e. the impact on accuracy of using our method is negligible. However, our method is around 500 times faster since we only have to consider a grid of 10 points instead of more than 5,000!

---

[1]Portfolio weights for the stock index, government bond and real T-bill rate should all be non-negative and add up to 1.

[2]We also performed additional tests including parameter uncertainty. In such a setting, both methods are also equally accurate.

Table 3.11: Comparison accuracy numerical methods

This table compares the portfolio weights obtained by
the simulation method in van Binsbergen and Brandt
(2007)(BB2007) with the portfolio weights obtained
by using the refined method of this paper (DPS2011).
We give the portfolio weights for a dynamic strategy
with K periods remaining for stocks, w_s, and bonds,
w_b. Results are based on the plug-in method. We
vary parameter K and risk aversion γ. State variables
are set to their historical average.

		BB2007		DPS2011	
K	γ	w_s	w_b	w_s	w_b
1	2	1.0000	0.0000	1.0000	0.0000
	5	0.5600	0.4400	0.5644	0.4356
	10	0.3000	0.4400	0.2974	0.4428
4	2	1.0000	0.0000	1.0000	0.0000
	5	0.6000	0.4000	0.5959	0.4041
	10	0.3100	0.2900	0.3110	0.2853
8	2	1.0000	0.0000	1.0000	0.0000
	5	0.6200	0.3800	0.6185	0.3815
	10	0.3200	0.2700	0.3198	0.2736
15	2	1.0000	0.0000	1.0000	0.0000
	5	0.6500	0.3500	0.6530	0.3470
	10	0.3400	0.3000	0.3420	0.2963
30	2	1.0000	0.0000	1.0000	0.0000
	5	0.6900	0.3100	0.6923	0.3077
	10	0.3700	0.3100	0.3651	0.3150
60	2	1.0000	0.0000	1.0000	0.0000
	5	0.7500	0.2500	0.7469	0.2531
	10	0.4100	0.2000	0.4097	0.2013

4

Model uncertainty for long-term investors[1]

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 subprime 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.

---

[1]This chapter is based on Diris (2011b).

4. MODEL UNCERTAINTY FOR LONG-TERM INVESTORS

4.1 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

---

[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.

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 - 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.[1]

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 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

---

[1]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.

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. Sec-

ondly, 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.[1] 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.

4.2 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.

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

---

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

Table 4.1: 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	

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 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 4.1 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.[1]

4.3   Methodology

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

$$y_t = \begin{pmatrix} r_{tbill,t} \\ x_t \\ s_t \end{pmatrix}, \tag{4.1}$$

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 ith column of Y.

---

[1]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.

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 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.[1]

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,

---

[1]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.

4. MODEL UNCERTAINTY FOR LONG-TERM INVESTORS

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.

4.3.1 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.[1] 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 in every equation

$$Y_i = X\beta_i + \epsilon_i, \; i = 1, ..., n, \tag{4.2}$$

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,

$$Y = XB' + E, \tag{4.3}$$

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

$$\epsilon_t \sim N(0, \Sigma). \tag{4.4}$$

---

[1]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).

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

$$Y_i = Z_i^{(j)}\beta_i^{(j)} + \epsilon_i^{(j)}, \ i = 1,......,n, \tag{4.5}$$

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.[1] 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.[2] 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 (4.5), we can express model j as follows

$$y = Z^{(j)}\beta^{(j)} + \epsilon^{(j)}, \tag{4.6}$$

where y is the $Tn \times 1$ vector vec(Y), $\beta^{(j)}$ is the $k^{(j)} \times 1$ vector $[\beta_1^{(j)'},...\beta_n^{(j)'}]'$ with $k^{(j)} = \sum_{i=1}^n k_i^{(j)}$, $\epsilon^{(j)}$ is a $Tn \times 1$ vector of vertically stacked $\epsilon_i^{(j)}$ and $Z^{(j)}$ is a $Tn \times k^{(j)}$ matrix

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

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

$$P(Y|\beta^{(j)},\Sigma^{(j)},M_j) = (2\pi)^{-\frac{Tn}{2}}|\Sigma^{(j)}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - Z^{(j)}\beta^{(j)})'(\Sigma^{(j)} \otimes I_T)^{-1}(y - Z^{(j)}\beta^{(j)})\right\}, \tag{4.7}$$

where I_T is an identity matrix of dimension T.

---

[1]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).

[2]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.

4.3.2 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

$$p(M_j) \propto q^{|M_j|}(1-q)^{n^2-|M_j|}, \tag{4.8}$$

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

$$p(M_j) = \frac{1}{n^2}. \tag{4.9}$$

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 4.7.1.

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 $\Sigma^{(j)}$

$$p(\Sigma^{(j)}|M_j) \propto |\Sigma^{(j)}|^{-\frac{n+1}{2}}. \tag{4.10}$$

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

$$p(\beta^{(j)}|\Sigma^{(j)}, M_j) \sim N(m^{(j)}, gV^{(j)}), \tag{4.11}$$

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 $\Sigma^{(j)}$

$$V^{(j)} = \left(Z^{(j)'} (\Sigma^{(j)} \otimes I_T)^{-1} Z^{(j)} \right)^{-1}. \tag{4.12}$$

In the special case where the explanatory variables $Z_i^{(j)}$ for equation i are equal to $Z^{(j*)}$ for all i, $V^{(j)} = \Sigma^{(j)} \otimes \left(Z_1^{(j*)'} Z_1^{(j*)} \right)^{-1}$. 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.[1] 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

---

[1]By looking at equation (4.16) 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.

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 4.7.1.

It is possible to impose that model M_j is stationary by adapting the prior for $\beta^{(j)}$ as in the previous chapter. However, we do not pursue this alternative, because we feel that it is too restrictive to impose such a priori in this setting. We want to consider model uncertainty over all specifications, not only the stationary ones.

4.3.3 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

$$p(\beta^{(j)}|Y, M_j, \Sigma^{(j)}) = N(\beta^{*(j)}, M^{*(j)}) \tag{4.13}$$

where

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

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

$$p(\Sigma^{(j)}|Y, M_j, \beta^{(j)}) \propto |Z^{(j)'}(\Sigma^{(j)} \otimes I_T)^{-1} Z^{(j)}|^{\frac{1}{2}} |\Sigma^{(j)}|^{-\frac{T+n+1}{2}} \exp\left\{ -\frac{1}{2} tr[(\Sigma^{(j)})^{-1}(E^{(j)'} E^{(j)} + \frac{1}{g} H^{(j)})] \right\}, \tag{4.14}$$

where

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

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

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 $\beta^{(j)}$.[1] The $n \times (n+1)$ matrix $M_f^{(j)'}$ is similarly defined with respect to the prior means in $m^{(j)}$. Expression (4.14) 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

$$
p(M_j|Y) = \frac{P(Y|M_j)p(M_j)}{\sum_k P(Y|M_k)P(M_k)}. \tag{4.15}
$$

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

$$
\begin{aligned}
p(Y|M_j, \Sigma^{(j)}) &= (2\pi)^{-\frac{Tn}{2}} |\Sigma^{(j)}|^{-\frac{T}{2}} (1+g)^{-\frac{k^{(j)}}{2}} \exp\left\{-\frac{1}{2}(y'(\Sigma^{(j)} \otimes I_T)^{-1}y)\right\} \\
&\quad \exp\left\{-\frac{1}{2}\left(m^{(j)'}\frac{1}{g}V^{(j)-1}m^{(j)} - \beta^{(j)*'}M^{*(j)-1}\beta^{(j)*}\right)\right\}.
\end{aligned} \tag{4.16}
$$

---

[1]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.

4.3.4 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.

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

$$\alpha = \min \left\{ 1, \frac{p(M_l^*)p(Y|M_l^*, \Sigma^{(j)})}{p(M_j)p(Y|M_j, \Sigma^{(j)})} \right\}. \tag{4.17}$$

   Otherwise set $M_l = M_j$.

Appendix B 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 (4.13).

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($E^{(l)'}E^{(l)} + H^{(l)}, T + n + 1$).

2. Set $\Sigma^{(l)} = \Sigma^{(l)*}$ with acceptance probability

$$\alpha = \min\left\{1, \frac{|Z^{(l)'}(\Sigma^{(l)*} \otimes I_T)^{-1}Z^{(l)}|^{\frac{1}{2}}|\Sigma^{(l)*}|^{\frac{n+1}{2}}}{|Z^{(l)'}(\Sigma^{(j)} \otimes I_T)^{-1}Z^{(l)}|^{\frac{1}{2}}|\Sigma^{(j)}|^{\frac{n+1}{2}}}\right\}. \tag{4.18}$$

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

Appendix B 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.

4.3.5 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.

4.4   Estimation Results

As a first step we estimate the unrestricted VAR(1) - the model in equation (4.3) - on the full data-set. Table 4.2 gives the OLS parameter estimates, its standard errors and the correlations and standard deviations of the residuals.

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 4.3 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.

4. MODEL UNCERTAINTY FOR LONG-TERM INVESTORS

Table 4.2: 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.

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
Panel A: OLS estimates and standard errors of slope coefficients												
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}	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

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

Table 4.3: 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

equations.[1] 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

---

[1]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.

4. MODEL UNCERTAINTY FOR LONG-TERM INVESTORS

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 for notational convenience. The following holds

$$
\begin{aligned}
E_t y_{t+1} &= A y_t \\
E_t y_{t+2} &= A^2 y_t \\
&\vdots \\
E_t y_{t+k} &= A^k y_t.
\end{aligned}
\tag{4.19}
$$

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

Table 4.4: 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

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.[1] 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 4.4 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 4.3. 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 4.5 reports the posterior moments of the coefficients of the weighted BMA model. The table shows the posterior means and standard deviations of the slope

---

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

Table 4.5: 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.

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
Panel A: Posterior means and standard deviations												
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	0.0132	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}	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	0.1003	0.1725	0.0435	0.1540	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

coefficients in panel A and the posterior mean of the covariance matrix of the residuals in panel B.[1]

Firstly, the results in the table are roughly in line with table 4.3. 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

---

[1]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.

signs are in line with the OLS 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 4.1 shows the posterior distribution of the coefficient on the credit spread in the equation for stock returns, (x_s, Cr_{spr}).
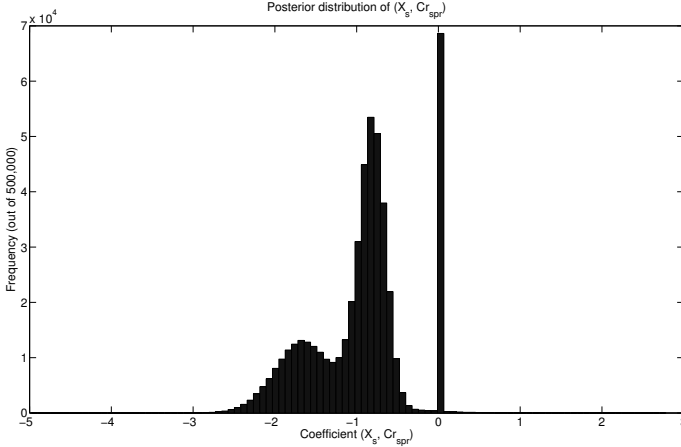
The figure shows that the posterior distribution is clearly non-normal. Firstly, there is a spike at 0. This spike corresponds to the probability that the credit spread is not included in the stock return equation. Secondly, even when the credit spread is included in the model we see that the posterior 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 4.1: Posterior distribution of the coefficent 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.

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

$$P_{ij} = \frac{P(i \bigcap j)}{P(i) + P(j) - 2P(i \bigcap j)}, \qquad (4.20)$$

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 4.6 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

126

Table 4.6: 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	-

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 4.7 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 4.3. 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.

Table 4.7: 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.

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
Panel A: Posterior means and standard deviations												
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	-		0.0441
					0.0681				6.8995			
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

4.5 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

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

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.[1] 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

---

[1]In this chapter, we ignore time-variation in parameters $\beta^{(j)}$ and $\Sigma^{(j)}$. We consider this source of uncertainty in the next chapter.

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.[1]

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

$$y_{t+1} = Ay_t + \epsilon_t. \tag{4.21}$$

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 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 4.2 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

---

[1]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.

Figure 4.2: Posterior distribution of the correlation between a shock to current excess stock returns and a shock to future expected excess stock return.
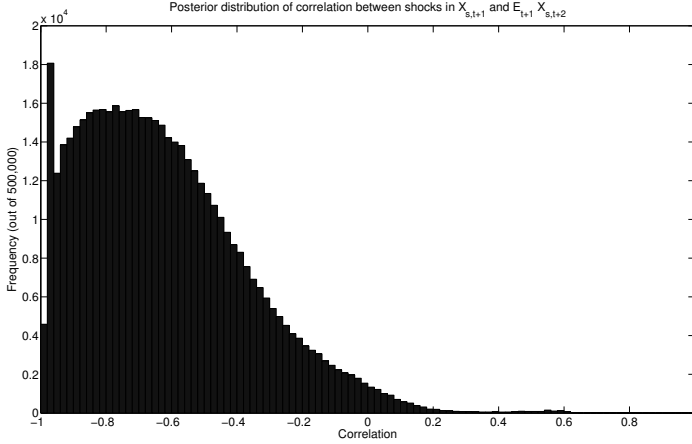
This figure shows the posterior distribution of the correlation between a shock in current excess stock returns and a shock in the expectation of future excess stock returns. The correlation is a measure for the mean reversion in stock returns. It is based on 500,000 retained draws from the posterior distribution of the weighted BMA model.

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 4.3 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.
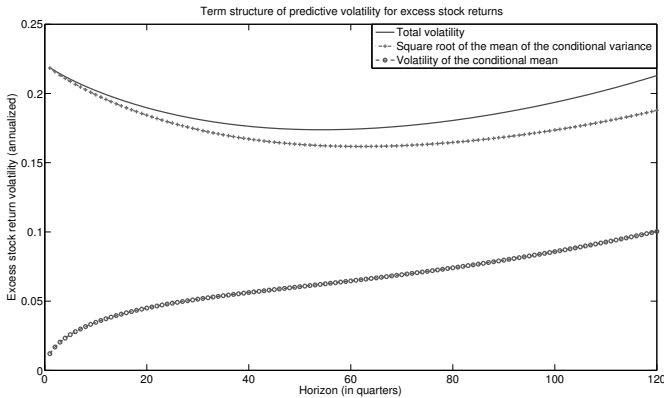
4. MODEL UNCERTAINTY FOR LONG-TERM INVESTORS

In order to obtain the posterior distribution, we estimate the specification on the full data-set.
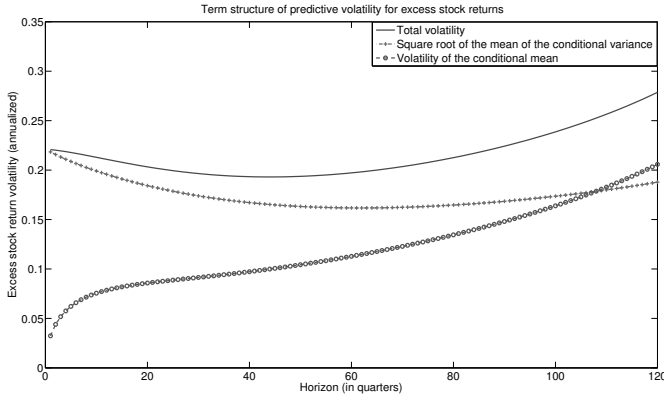
The picture shows that stocks are almost as risky in the long-run as in the short-run. The annualized total volatility is 22% at short horizons and around 20% at horizons up to 30 years. What effects play a role? Firstly, the mean-reversion effect is important. If a long-term investor sees a bad stock 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

Figure 4.3: 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 4.4: 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.

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 4.4 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 4.4 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 4.5. 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.

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,

Figure 4.5: Time-series of annualized total volatility at different horizons.
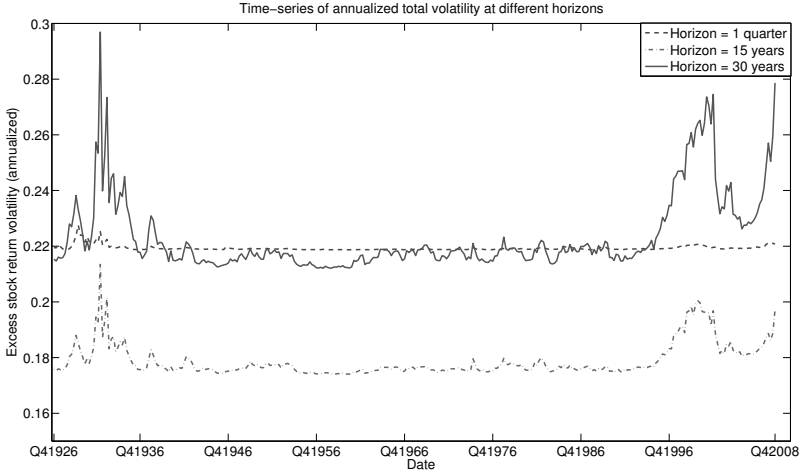
This figure shows a time-series of the square root of the annualized total variance of excess log stock returns for three different investment horizons. It is based on 50,000 retained draws from the posterior distribution of the weighted BMA model.

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
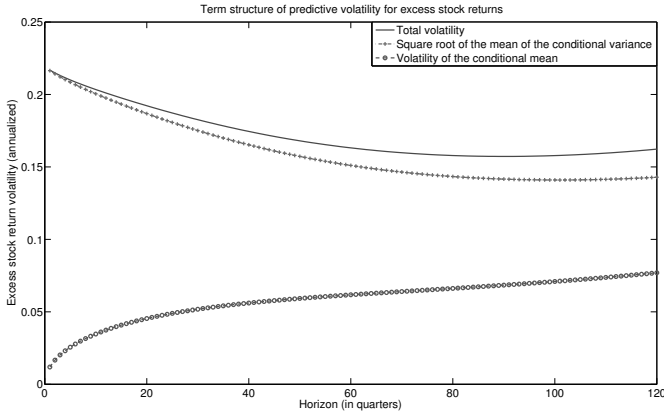
variance of the predictive distribution) changes a lot over time even though we only consider homoscedastic models!

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 4.6 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.
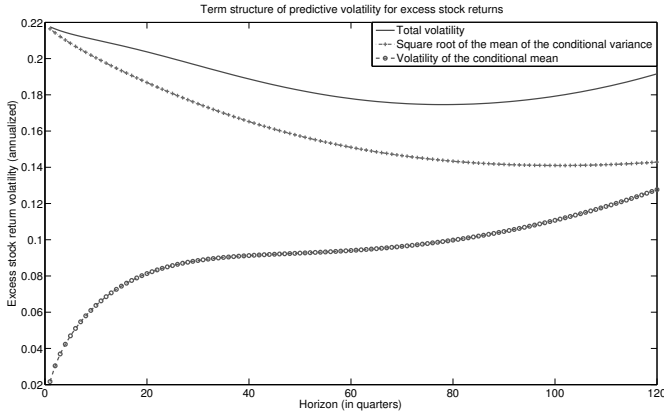
The decomposition shows that the mean of the conditional variance is the biggest component of the total variance at all horizons. The total volatility is around 22% for short investment horizons, decreases to 15.5% for moderate investment horizons and

Figure 4.6: 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 4.7: 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.

slightly increases again towards 17% for investment horizons up to 30 years. Results suggest that stocks are safer assets in the long-run.

Figure 4.7 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 4.8 and 4.9 show the term structures of risk when variables are set equal to respectively their historical average and the end-of-sample values.

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 4.9 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.
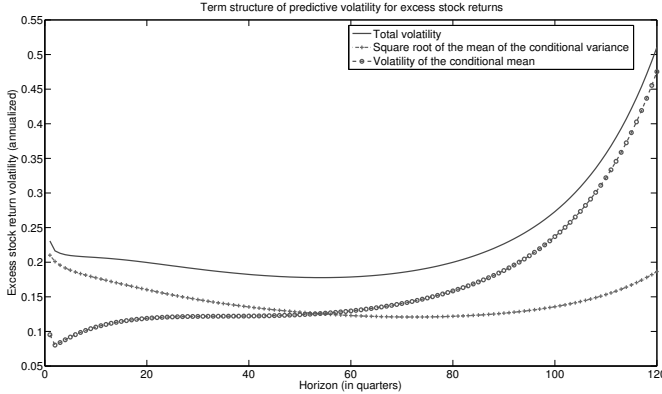
Note that the full model is one of the models that we consider in calculating the BMA specification. However, it receives an extremely low weight in the composite model, because it contains way too many parameters. Therefore, the uncertainty in the full model plays a negligible role in the composite specification. Hence, if one

Figure 4.8: 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.
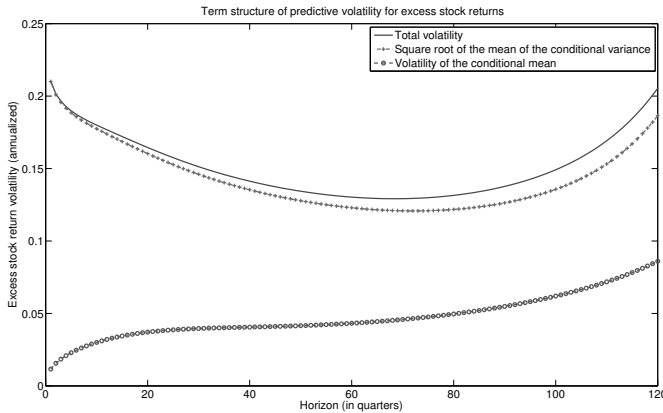
This figure shows the term structure of risk for excess log stock returns. It is based on 100,000 retained draws from the posterior distribution of the full model. The predictor variables are equal to their historical average. Note that the total variance is equal to the mean of the conditional variance plus the variance of the conditional mean and note that all values are annualized.

Figure 4.9: 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.

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.

4.6 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 $\gamma > 1$

$$\max_{w_t} E_t \left(\frac{W_{t+K}^{1-\gamma}}{1-\gamma} \right) \tag{4.22}$$

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 (4.22) 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 4.10: Higher moments stock returns. Predictor variables are equal to their historical average.
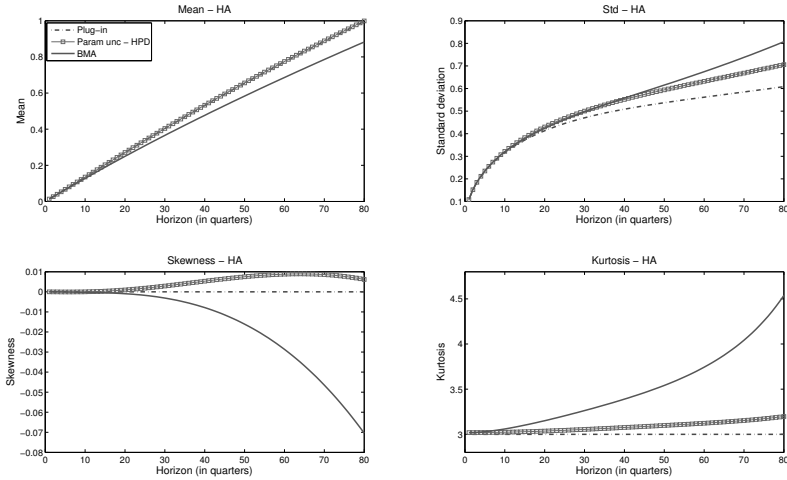
This figure plots the first four moments (mean, standard deviation, skewness and kurtosis) of the predictive distribution of excess log stock returns using either the highest posterior probability (HPD) model without parameter uncertainty (plug-in), the HPD with parameter uncertainty (param unc - HPD) 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.

Firstly, we consider the predictive distribution of future stock returns. Figure 4.10 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.[1]

---

[1]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

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.

Figure 4.11 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.[1] The means also deviate substantially from the means in figure 4.10. 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 4.10, 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
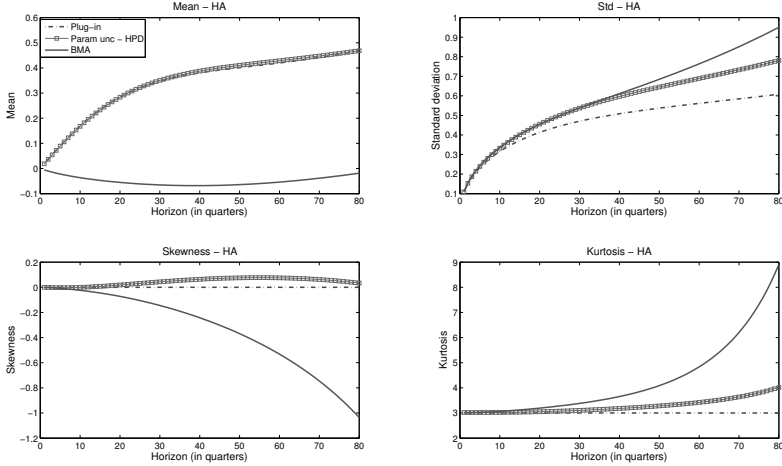
predictive distribution of excess stock returns.

[1]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 4.11: 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.

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.

How do the moments for the predictive distribution using the BMA specification change over time when variables change over time? Figure 4.12 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.

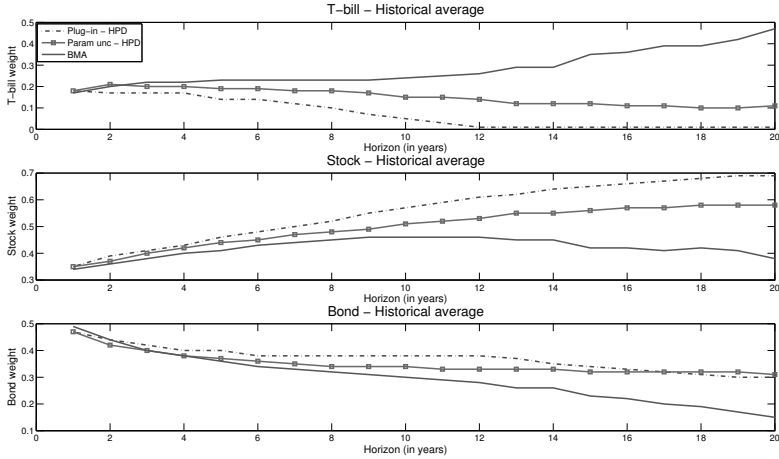**Figure 4.12: Time-series of four moments of predictive distribution.**
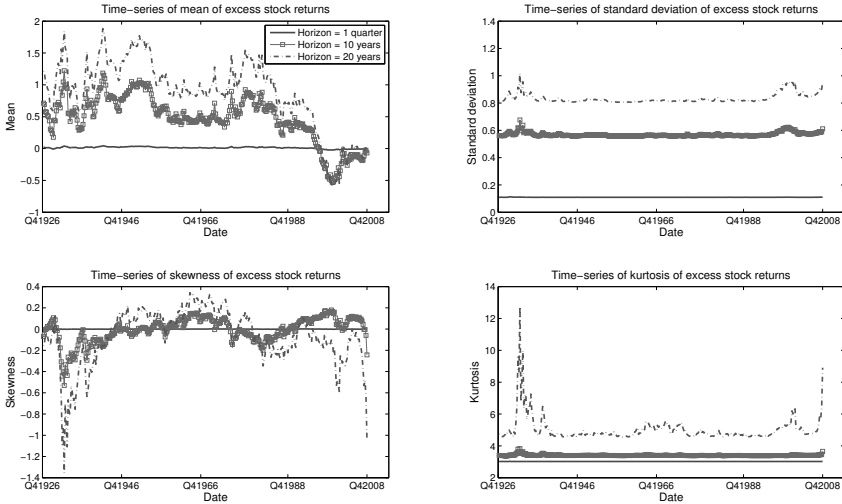
This figure shows a time-series of the four moments of the predictive distribution for three different investment horizons. It is based on 50,000 retained draws from the posterior distribution of the weighted BMA model.

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.

Figure 4.13 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

Figure 4.13: 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.

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

Figure 4.14: **Portfolio weights versus horizon for different estimation techniques. Predictor variables are equal to the values at the end of the sample.**
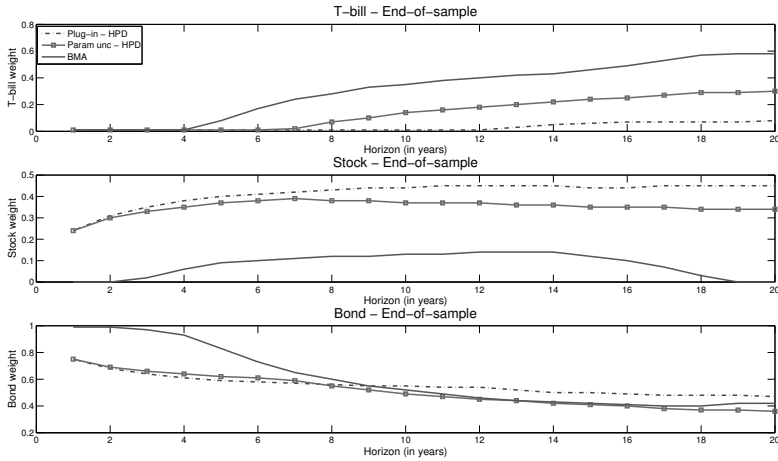

This figure plots T-bill weights (panel A), stock weights (panel B) and bond weights (panel C) against different investment horizons for different estimation techniques. The weights are the optimal weights for a buy-and-hold investor with risk aversion parameter $\gamma = 5$. Weights are either based 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.

rate compared to the investment of a short-term investor.

What happens with asset allocations when we consider the optimal buy-and-hold asset allocations at the end of the sample? Figure 4.14 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 4.13. 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 4.15 and 4.16 plot the certainty equivalence (a monotonic transformation of expected utility) against the investment horizon for

Figure 4.15: Plot of Certainty Equivalent versus horizon for different estimation techniques. Predictor variables are equal to their historical average.
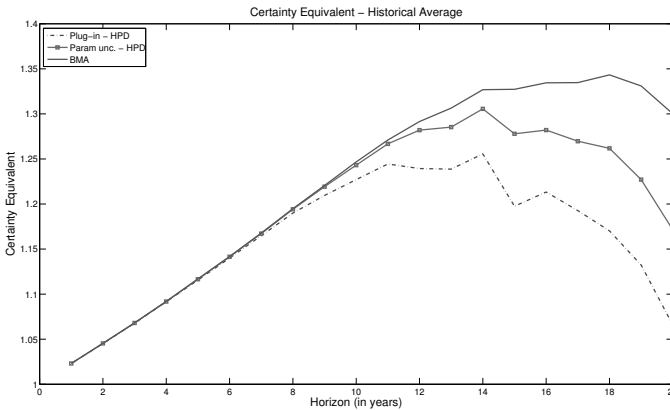
This figure plots the certainty equivalence of different strategies versus the investment horizon. The certainty equivalents are calculated using a specification based on the HPD without parameter uncertainty, a specification based on the HPD with parameter uncertainty and a specification based on the 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.
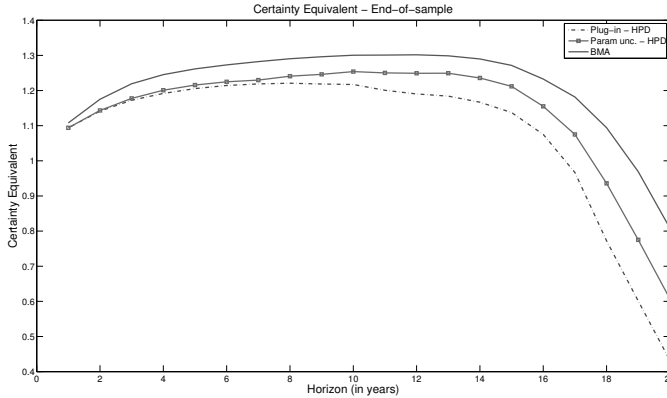
Figure 4.16: Plot of Certainty Equivalent versus horizon for different estimation techniques. Predictor variables are equal to the values at the end of the sample.

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.

the three specifications.[1] 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.

---

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

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.

4.7 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.

4.7.1 Different prior distribution

In this section, we check the prior robustness of our results by comparing the posterior distribution in section 4.4 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.

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

$$p(q) = Beta(1,b), \tag{4.23}$$

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 4.3.4 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 4.3.2 and where g is set equal to T.

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

Table 4.8: 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 4.9: 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.

	R_{tbill}	X_s	X_b	Def_{pr}	DP	BM	PE	Y_{nom}	Y_{spr}	Cr_{spr}	$ntis$	Var
Panel A: Posterior means and standard deviations												
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	0.0010	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	-0.0022	-0.0231	-0.1077	0.0058	-0.0319	-0.0319	0.0619	0.0123	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

prior for g (using the parameterization of Bauwens, Lubrano, and Richard (1999) for the inverse-gamma distribution)

$$p(g) = iG(0.01, 0.01). \tag{4.24}$$

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

$$p(g|Y, M_j, \beta^{(j)}) = iG((\beta^{(j)} - m^{(j)})'V^{(j)-1}(\beta^{(j)} - m^{(j)}) + 0.01, k^{(j)} + 0.01). \quad (4.25)$$

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 (4.16) now conditions on both $\Sigma^{(j)}$ and g, but the expression for the marginal likelihood itself stays exactly the same.

Table 4.8 shows the posterior probability that a variable is included in an equation for the alternative prior specification. If we compare this table to table 4.3, 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 4.3. Table 4.9 shows the posterior means and standard deviations for the alternative specification. If we compare the posterior moments with the posterior moments in table 4.5, 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.

4.7.2 Posterior moments over time

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 4.4 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 4.17 and 4.18 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 4.19 and 4.20 show similar figures for the posterior means of these coefficients. These figures largely confirm the results above.

Figure 4.17: Time-series of model probability (X_s, DP)
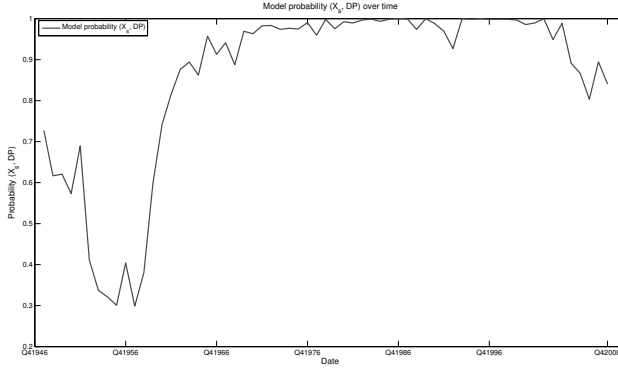
This figure shows the time-series of the model probability (X_s, DP). These are calculated by estimating the BMA specification on an expanding window. The x-axis indicates the value of the last included observation in the window. Results are based on 100,000 retained draws.

Figure 4.18: Time-series of model probability (X_s, Cr_{spr})
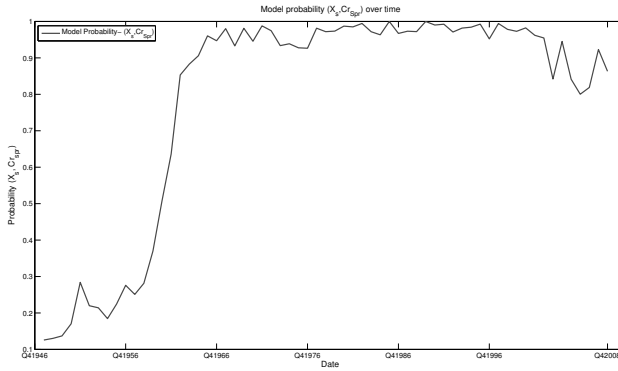
This figure shows the time-series of the model probability (X_s, Cr_{spr}). These are calculated by estimating the BMA specification on an expanding window. The x-axis indicates the value of the last included observation in the window. Results are based on 100,000 retained draws.

Figure 4.19: 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.
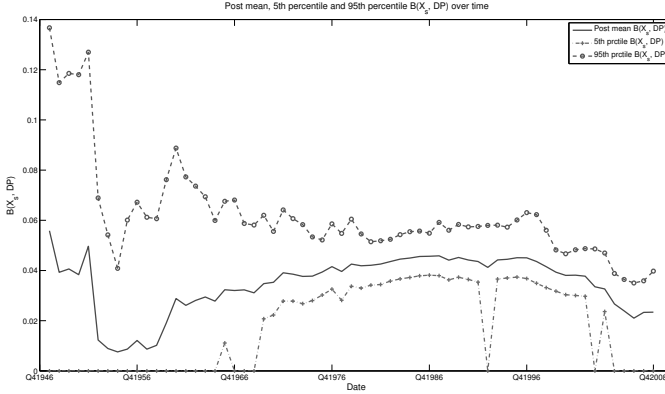
Figure 4.20: Time-series of posterior mean of coefficient (X_s, Cr_{spr})

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.
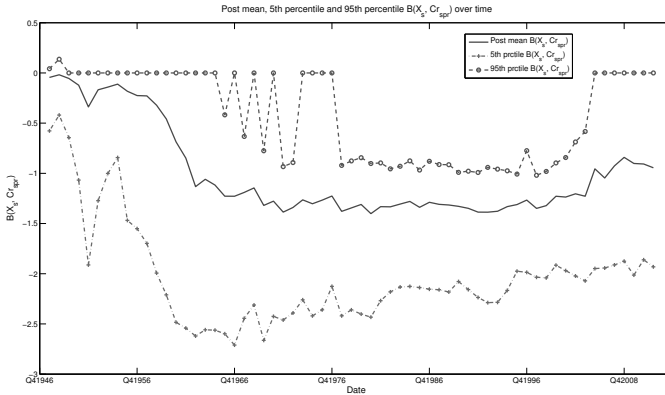
Firstly, the posterior means stabilize around 1960. Furthermore, we see that the means for both coefficients are shrunk towards 0 in the last years, because the probability that the variable is included in the BMA specification decreases in these years.

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.

4.8 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.

4.9 Appendix: 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.

4.9 Appendix: Posterior distribution and MCMC algorithm

The likelihood function for model M_j in equation (4.5) is (all three expressions are equivalent)

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

where

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

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)

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

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

$$
\begin{aligned}
p(\beta^{(j)}|M_j, Y, \Sigma^{(j)}) &\propto \exp\left\{-\frac{1}{2}(\beta^{(j)} - \hat{\beta}^{(j)})'V^{(j)-1}(\beta^{(j)} - \hat{\beta}^{(j)})\right\} \\
&\quad \exp\left\{-\frac{1}{2}(\beta^{(j)} - m^{(j)})'\frac{1}{g}V^{(j)-1}(\beta^{(j)} - m^{(j)})\right\}. \quad (4.26)
\end{aligned}
$$

Note the following:

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

where $M^{*(j)}$ is defined in equation (4.14) 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 (4.14). 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

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

and hence equals equation (4.14).

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)'} (\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)'} V^{(j)-1} \hat{\beta}^{(j)} + m^{(j)'} \frac{1}{g} V^{(j)-1} m^{(j)} - \beta^{*(j)'} M^{*(j)-1} \beta^{*(j)}\right)\right\} \\
&\quad \int \exp\left\{-\frac{1}{2}(\beta^{(j)'} - \beta^{*(j)})' M^{*(j)-1} (\beta^{(j)'} - \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)'} V^{(j)-1} \hat{\beta}^{(j)} + m^{(j)'} \frac{1}{g} V^{(j)-1} m^{(j)} - \beta^{*(j)'} M^{*(j)-1} \beta^{*(j)}\right)\right\} \\
&\quad (2\pi)^{\frac{k^{(j)}}{2}} |M^{*(j)}|^{\frac{1}{2}} \\
\\
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)'} \frac{1}{g} V^{(j)-1} m^{(j)} - \beta^{*(j)'} 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)'} \frac{1}{g} V^{(j)-1} m^{(j)} - \beta^{*(j)'} 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 (4.16).

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.

4. MODEL UNCERTAINTY FOR LONG-TERM INVESTORS

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 (4.13): $p(\beta^{(l)}|M_l^*, \Sigma^{(j)}, Y)$. 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 (4.17).

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 (4.13): $p(\beta^{(l)}|M_l = M_j, \Sigma^{(j)}, Y)$.

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 (4.14). As a proposal density, we choose an iWishart$\left(E^{(l)'}E^{(l)} + \frac{1}{g}H^{(l)}, T + n + 1\right)$.[1] The acceptance probability becomes

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

which is equivalent to the expression in equation (4.18).

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.

---

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

5

Model instability and long-term investors[1]

We analyze the effect of model instability on long-term investors using a time-varying VAR(1) model. Our specification is able to handle time-varying intercepts, time-varying slopes, time-varying volatility, time-varying correlation, the leverage effect and fat tails. We find that the persistence of time-variation is important to assess its importance for long-term investors. Time variation in intercepts and slope coefficients is not persistent enough for long-term investors to be relevant while time-variation in the error covariance matrix (especially error volatility) is persistent and therefore very important for long-term investors. Fat tails disappear once time-varying volatility is incorporated. Random walk specifications (persistence equal to 1) or regime-switching models (same persistence for all parameters) lead to a large overestimation of perceived stock market risk and an underinvestment in the stock market. Results are robust to changes in the specification.

5.1 Introduction

Long-term investors face substantial uncertainty when modeling future asset returns. Firstly, investors need to select a model to model the dynamics of asset returns. This model can be wrong and investors therefore face model uncertainty. Chapter 4 shows that this uncertainty plays a major role at long-horizons. Secondly, upon choosing a

---

[1]This chapter is based on Diris (2011a).

model long-term investors need to estimate the model parameters. They face parameter uncertainty, since the true parameters are unknown. Parameter uncertainty is also very important at long horizons as argued in Barberis (2000). Thirdly, even if the true model and true parameters are known, stock returns are still uncertain due to unexpected shocks (error term). This component is the most important component at short horizons.

Recently, there has been a lot of interest in the strategic asset allocation literature due to the finding that asset returns, specifically stock returns, might be predictable.[1] If asset returns are predictable, the optimal asset allocations of long-term investors deviate from the allocations of short-term investors. Empirically, it is found by e.g. Campbell, Chan, and Viceira (2003) that long-term investors should invest more in stocks than short-term investors, since stocks mean-revert and are therefore safer in the long-run. This conclusion is robust to the inclusion of parameter uncertainty as shown in Barberis (2000), but not to the incorporation of model uncertainty as shown in chapter 4.

The common practice in the strategic asset allocation literature is to estimate a model on a data-set of 50 years or more and to assume that model coefficients remain constant over this period. However, a priori there is no reason to believe that they indeed are. Aspects such as institutional changes, wars or changes in the stock market behavior of participants due to different risk aversion levels or different financial sophistication levels could lead to changes in the relation between asset returns and predictor variables or to changes in the properties of the error term.

While small changes in coefficients might not have a large impact on short-term investors, they can have a large impact on long-term investors if they are long-lasting and persistent. In that case, the mistakes one makes in using a constant model add up over the investment horizon and can become very large. If on the contrary the changes in coefficients are not persistent but of transitory nature, they are unlikely to be important for long-term investors. In that case the coefficients only deviate from the constant model for a fraction of the investment horizon. Hence, the persistence of time-varying parameters is extremely important at long horizons.

---

[1]Campbell and Viceira (2002) and Brandt (2010) give an excellent overview of the strategic asset allocation literature.

Another common practice in the literature is to impose Gaussian error distributions, whereas evidence clearly shows that the error distributions deviate significantly from normality. Risk-averse investors evaluate very good and very bad outcomes differently, since they want to avoid bad outcomes at all costs. They value models that are able to avoid these bad outcomes as chapter 3 shows. This suggests that the normality assumption could lead to very misguided investment advice, since it does not accurately model extreme tail events. Properly modeling the error term is therefore very important for risk-averse investors.

We consider the effect of model instability on long-term investors using a time-varying VAR(1) model in which parameters are allowed to change in every period. We develop a methodology that is able to handle time-varying intercepts, time-varying slope coefficients, time-varying error volatility, time-varying error correlation, the leverage effect and fat tails. We assess the importance of these model components for long-term investors. We focus on the persistence of the different time-varying parameters.

An alternative specification is the regime-switching model that Guidolin and Timmermann (2007) and Pettenuzzo and Timmermann (2010) implement. We do not pursue this alternative here because of three reasons. Firstly, a priori it seems more likely that the behavior of stock market participants changes smoothly over time (due to changes in risk aversion or financial sophistication) instead of abruptly, which suggests that a regime-switching model is not appropriate. Secondly, one of our objectives is to assess which kind of model instability is the most important for long-term investors. A regime-switching model does not allow us to assess the individual components, since all components change jointly. Finally, a regime-switching model pools the persistence parameter of all components. In other words, changes in say the slope of stock returns are equally persistent as changes in the volatility of a predictor variable, while there is a priori no reason to impose such a restriction.

A related paper is Johannes, Korteweg, and Polson (2011). These authors develop particle filtering techniques to assess how the views of economic decision makers evolve over time using a VAR model in which only some parameters are allowed to change over time. Our chapter differs in both scope and perspective. Firstly, we use smoothed estimates to assess ex-post whether there is model instability, while Johannes, Korteweg, and Polson (2011) use filtered estimates. Since we use smoothed parameters, we are able to assess more efficiently whether there was time variation in the past 82

years. Since Johannes, Korteweg, and Polson (2011) use filtered estimates, they would find time-variation even if there was none. Secondly, we consider time-variation in all intercepts and slope coefficients, add the leverage effect, consider non-normal distributions and also consider time-varying error correlation. Johannes, Korteweg, and Polson (2011) only consider time-variation in one slope coefficient and in the volatility of the error term. Thirdly, we look at specifications that also consider bond returns and are therefore more relevant for long-term investors and indeed find interesting effects in the correlation between stocks and bonds.

We find that it is important for long-term investors to take model instability into account. CER gains are up to 5% per year. Long-term investors should take time-varying volatility and correlation into account, but can safely ignore time-varying slopes and excess kurtosis once time-varying volatility is incorporated. The reason is that time-varying slopes are not persistent enough to be of importance for long-term investors and that fat tails are not important once stochastic volatility is incorporated. The persistence of time-varying parameters is extremely important and a random walk specification (persistence equal to 1) or a regime-switching model (same persistence for all parameters) is therefore not appropriate. Results are robust to changes in the main specification.

This chapter is organized as follows. Section 2 discusses the data-set we use and performs a preliminary analysis. Next, section 3 explains the methodology. It discusses the model, the Bayesian prior distribution and the Bayesian MCMC techniques. Section 4 shows the results for the basic specification in which the dividend-to-price ratio is incorporated as predictor of asset returns. Section 5 performs a robustness check using the yield-spread as predictor variable. Finally, section 6 provides the conclusion. The appendix contains details on the numerical techniques we use to estimate the model.

5.2   Data and preliminary analysis

We use a monthly data-set that starts in December 1926 and ends in December 2008 for the US stock and bond markets. It is based on Goyal and Welch (2008). We use three asset returns and two predictors.

The first asset return is the ex post real T-bill rate (R_{tbill}) which we obtain by subtracting log inflation from the log return on the 3-month T-bill rate. We do not

include the T-bill rate in the econometric models to keep our models parsimonious, but we use its average value, which is 0.060% per month, in portfolio construction. The second asset return is the excess log stock return (X_s). It is defined as 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 third asset return is the excess log return on a long-term government bond (maturity of approximately 20 years) and is defined in a similar way.

Our data-set contains two predictor variables. The first predictor variable is the log dividend-to-price ratio, defined as the log difference between dividends over the past four quarters and the current S&P index level. Campbell and Shiller (1998) and Cochrane (2007b) (among others) show that this ratio is an important predictor of stock returns. Secondly, we consider the yield-spread, which is the difference between the log yield on a long-term government bond and the log yield on the 90-day T-bill. It is an important predictor of both stock and bond returns, refer to e.g. Campbell (1995) and Fama and French (1989).

Table 5.1 shows the summary statistics for the data-set. Firstly, the equity risk premium of 5.4% per year is in line with other papers. Secondly, the kurtosis and skewness clearly indicate that the variables deviate significantly from normality. This holds especially for asset returns. Thirdly, the AR(1) coefficients indicate that the predictor variables are very persistent.

We consider two different specifications. The first specification is a VAR(1) model in which stock returns, bond returns and the dividend yield are regressed on a constant and the lagged dividend yield. This specification is considered in section 5.4. The second specification is a VAR(1) model in which excess stock returns, bond returns

Table 5.1: Summary Statistics of the monthly data-set

This table reports the means, standard deviations, minima, maxima, AR(1) coefficients, Skewness, Kurtosis and Sharpe ratios for excess stock returns (X_s), excess bond returns (X_b), the dividend-to-price ratio (DP) and the yield spread (Y_{spr}). The data set is monthly and starts in December 1926 and ends in December 2008. Percentages are given as fractions.

	Mean	Std	Min	Max	AR(1)	Skewness	Kurtosis	Sharpe
X_s	0.0044	0.0557	-0.3391	0.3471	0.0904	-0.4007	10.8715	0.0793
X_b	0.0015	0.0231	-0.1041	0.1342	0.0556	0.3253	7.4090	0.0640
DP	-3.3399	0.4622	-4.5074	-1.6851	0.9924	-0.2759	3.2983	
Y_{spr}	0.0153	0.0122	-0.0319	0.0438	0.9600	-0.2259	3.0648	

Table 5.2: Preliminary analysis

This table shows the OLS estimates, standard errors and covariance matrix for respectively the model with DP (panel A) and the model with Y_{spr} (panel B) as predictor. The first subpanel gives the parameter estimates, the standard errors in brackets and the skewness and kurtosis of the error terms. The second subpanel shows the estimates for the covariance matrix of the residual. The diagonal indicates the standard errors of the residuals whereas the off-diagonal elements are the correlations. Note that the different equations are given in different rows and that c indicates the constant.

	Panel A1: Est. DP model			
	c	DP	Skewness	Kurtosis
X_s	0.0270	0.0068	-0.5262	10.7775
	(0.0129)	(0.0038)		
X_b	0.0031	0.0005	0.3293	7.3907
	(0.0054)	(0.0016)		
DP	-0.0258	0.9924	0.2568	11.4675
	(0.0131)	(0.0039)		
	Panel A2: Cov. DP model			
	X_s	X_b	DP	
X_s	0.0556	0.1227	-0.9860	
X_b	0.1227	0.0563	-0.1253	
DP	-0.9860	-0.1253	0.0563	
	Panel B1: Est. Y_{spr} model			
	c	Y_{spr}	Skewness	Kurtosis
X_s	0.0023	0.1373	-0.4238	10.8776
	(0.0028)	(0.1459)		
X_b	-0.0026	0.2640	0.3516	7.6976
	(0.0012)	(0.0599)		
Y_{spr}	0.0006	0.9600	0.5155	25.1347
	(0.0002)	(0.0090)		
	Panel B2: Cov. Y_{spr} model			
	X_s	X_b	Y_{spr}	
X_s	0.0557	0.1201	-0.0185	
X_b	0.1201	0.0228	-0.2774	
YS	-0.0185	-0.2774	0.0034	

and the yield spread are regressed on a constant and the lagged yield spread. We look at this model in the robustness section 5.5. We do not consider VAR(1) models of dimension greater than 3 to keep our analysis feasible and to reduce the total number of parameters.

Table 5.2 shows the OLS estimates, standard errors, the covariance matrix of the residuals and the skewness and kurtosis of the residuals for both specifications. We find that the dividend-to-price ratio (panel A) and the yield spread (panel B) are positively related to excess stock returns. These positive coefficients combined with the negative error correlation between excess stock returns and especially the dividend-to-price ratio suggest that stocks mean-revert on average. The large standard errors indicate that there is a lot of estimation uncertainty involved. Next, the yield spread is a positive and strong predictor of excess bond returns, but hardly predicts stock returns. Its positive coefficient combined with the negative error correlation between excess bond returns and the yield spread suggests that bond returns show some mean-reversion. Finally, the skewness and kurtosis values indicate that the error terms deviate strongly from normality. The kurtosis value for the yield-spread is especially remarkable and needs to be further analyzed.

5.3   Methodology

In this paper, we use a first order time-varying Vector Autoregression - TVAR(1) - to model the investment opportunity set of long-term investors.[1] The model is able to handle time-varying intercepts and slope coefficients, a time-varying error covariance matrix with both volatility and correlation time-varying, a leverage effect in volatility and finally error terms with fat tails. The model is estimated by Bayesian MCMC techniques.[2]

---

[1]As argued in the introduction of this thesis, time-variation in the model might be a sign of time-variation in the parameters of the DGP or a sign of misspecification of the model. Even if it is a sign of misspecification, it does not invalidate the use of the time-varying model. All models are wrong, but some are actually useful. The TVAR(1) is a very flexible and therefore useful model.

[2]In theory, it would be possible to estimate the model with frequentist techniques. However, this would be extremely difficult due to the large number of parameters and the non-linearity of the problem. Bayesian methods on the other hand are well-suited to estimate a problem of this magnitude, since it divides the original estimation problem in smaller and simpler steps (the Gibbs step).

5. MODEL INSTABILITY AND LONG-TERM INVESTORS

Our methodology is an extension of the methodology of Primiceri (2005). We extend Primiceri's (2005) methodology by allowing for fat tailed error distributions, by estimating the leverage effect in volatility and by estimating AR(1) processes for all transition equations. These extensions are very relevant given the data-set we use. Firstly, it is well-known (e.g. Omori, Chib, Shephard, and Nakajima (2007), Jacquier, Polson, and Rossi (2004)) that the leverage effect in volatility and fat tailed error distributions are present in data on stock returns. Secondly, empirically we find that it is very important to estimate the persistence of time-varying processes instead of imposing random walks. We explain below that random walks lead to a large overestimation of perceived risk of long-term stock returns. Ignoring these extensions would lead to misspecified models.

5.3.1 Model

In this section, we explain the most general model we estimate. Define the $n \times 1$ vector y_t as follows ($n = 3$ in the empirical application)

$$y_t = \left(\begin{array}{c} x_t \\ z_t \end{array} \right), \tag{5.1}$$

where x_t is a $n - k \times 1$ vector consisting of the asset returns and z_t a $k \times 1$ vector of predictor variables at time t. The model we consider is as follows

$$y_t = a_t + B_t z_{t-1} + u_t, \tag{5.2}$$

for $t = 1, ..., T$, where a_t is a $n \times 1$ vector of intercepts, B_t an $n \times k$ matrix of slope coefficients and u_t an $n \times 1$ vector of error terms with covariance matrix Ω_t whose properties are indicated below.

First, we introduce some additional notation. Let $X_t = I_n \otimes [1, z_{t-1}]$ and let $b_t = vec\left([a_t, B_t]'\right)$. The model can be rewritten as

$$y_t = X_t b_t + u_t. \tag{5.3}$$

Without loss of generality, we consider a triangular reduction of covariance matrix Ω_t

$$L_t \Omega_t L_t' = \Sigma_t \Sigma_t, \tag{5.4}$$

where L_t is an $n \times n$ lower triangular matrix

$$L_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ l_{21,t} & 1 & 0 & 0 \\ \vdots & \ddots & 1 & 0 \\ l_{n1,t} & \cdots & l_{nn-1,t} & 1 \end{pmatrix} \tag{5.5}$$

and where Σ_t is an $n \times n$ diagonal matrix

$$\Sigma_t = \begin{pmatrix} \sigma_{t,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{t,n} \end{pmatrix}.[1] \tag{5.6}$$

The common assumption in the literature (e.g. Campbell, Chan, and Viceira (2003), Barberis (2000)) is to assume that the error term is normally distributed. In that case, we could write

$$u_t = L_t^{-1} D_t e_t,$$

where e_t has a standard normal distribution. Since $\Omega_t = L_t^{-1} \Sigma_t \Sigma_t L_t'^{-1}$, u_t would be distributed as $N(0, \Omega_t)$.

However, there is ample evidence that suggests that the distribution of the error term deviates from normality, see for example the preliminary results in the previous section. Therefore, we consider an alternative specification for the error term and assume that u_t has a distribution with fat tails. In order to do so we introduce scale mixture variables $\lambda_{t,1}, ... \lambda_{t,n}$ and $n \times n$ diagonal matrix

$$\Lambda_t = \begin{pmatrix} \lambda_{t,1}^{\frac{1}{2}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{t,n}^{\frac{1}{2}} \end{pmatrix} \tag{5.7}$$

and make the following assumption for u_t

$$u_t = L_t^{-1} \Sigma_t \Lambda_t e_t, \tag{5.8}$$

where the elements of e_t have independent standard normal distributions.

---

[1]Due to the triangular reduction, the results depend in theory on the ordering of the variables. Our empirical results turn out to be robust to different orderings.

Table 5.3: Moments of the normal-lognormal distribution

This table reports the first four moments of the normal-lognormal distribution $\lambda_t^{\frac{1}{2}} e_t$ with $\ln \lambda_t \sim N(-(1/2)\tau^2, \tau^2)$ and $e_t \sim N(0,1)$ for different values of τ^2.

τ^2	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
Mean	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Std	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Skewness	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Kurtosis	3.0	3.7	4.5	5.5	6.6	8.1	9.2	12.0	14.5	18.2	22.8

By multiplying $e_{t,j}$ by scale mixtures $\lambda_{t,j}^{\frac{1}{2}}$, error term $u_{t,j}$ deviates from normality $\forall\, j$. We follow Omori, Chib, Shephard, and Nakajima (2007) and use the following distribution for $\ln \lambda_{t,j}$ for $j = 1, ..., n$

$$\ln \lambda_{t,j} \sim N\left(-\frac{1}{2}\tau_j^2, \tau_j^2\right). \tag{5.9}$$

If $\tau_j^2 = 0$, $\lambda_{t,j}^{\frac{1}{2}} e_{t,j}$ would have a normal distribution. If $\tau_j^2 > 0$, $\lambda_{t,j}^{\frac{1}{2}} e_{t,j}$ has a normal log-normal distribution.[1] This is a distribution with fat tails. Its moments for different values of τ_j^2 are given in table 5.3. The table clearly shows that the multiplication of $e_{t,j}$ for $j = 1, ...n$ with the scale mixture only impacts the kurtosis. The mean, standard deviation and skewness are not influenced. Therefore, Ω_t can still be interpreted as the covariance matrix of u_t.

Next, we specify the dynamics of the time-varying parameters. Let l_t be the $n(n-1)/2 \times 1$ vector of non-zero and non-one elements of L_t (stacked by rows) and let σ_t be the $n \times 1$ vector of diagonal elements of Σ_t. The evolution of the time-varying parameters in the model - b_t, l_t and σ_t - is modeled as follows

$$
\begin{aligned}
b_{t+1} - \mu_b &= A_b\left(b_t - \mu_b\right) + \eta_t \\
l_{t+1} - \mu_l &= A_l\left(l_t - \mu_l\right) + \zeta_t \\
\ln \sigma_{t+1}^2 - \mu_s &= A_s\left(\ln \sigma_t^2 - \mu_s\right) + \xi_t,
\end{aligned} \tag{5.10}
$$

where μ_b, μ_l and μ_s are respectively the unconditional means of b_t, l_t and $\ln \sigma_t^2$, where A_b, A_l and A_s are the transition matrices and where η_t, ζ_t and ξ_t are the innovations.

---

[1] An alternative would be to specify that $\lambda_{t,j}$ has an inverse gamma iGamma(ν, ν) distribution such that $\lambda_{t,j}^{\frac{1}{2}} e_{t,j}$ would have a student-t distribution with ν degrees of freedom. However, this choice leads to an unstable numerical algorithm and therefore we do not pursue this alternative in this chapter.

These are distributed as $\eta_t \sim N(0, Q)$, $\zeta_t \sim N(0, R)$ and $\xi_t \sim N(0, S)$. The next subsection explains that we need to impose restrictions on R and S to make the analysis tractable.[1]

We follow Primiceri (2005) and assume that η_t, ζ_t and ξ_t are independent of each other and that η_t and ζ_t are independent of error term e_t. However, unlike Primiceri (2005) we do allow for a leverage effect in volatility by specifying that the correlation between $e_{t,j}$ and $\xi_{t,j}$ is (instead of 0)

$$corr(e_{t,j}, \xi_{t,j}) = \rho_j, \forall j. \tag{5.11}$$

Omori, Chib, Shephard, and Nakajima (2007) and Jacquier, Polson, and Rossi (2004) find that this correlation is negative for a shock to stock returns and an innovation in its stochastic volatility. This means that if there is a negative shock to stock returns at time t, *next* period's log volatility $\ln \sigma_{t+1}^2$ will be higher on average.

We introduce some additional notation. Firstly, define $\widetilde{b}_{t+1} = b_{t+1} - \mu_b$, $\widetilde{l}_{t+1} = l_{t+1} - \mu_l$ and $\widetilde{\ln \sigma^2}_{t+1} = \ln \sigma_{t+1}^2 - \mu_s$. Secondly, let b, l and $\ln \sigma^2$ be vectors that stack all values of b_t, l_t and $\ln \sigma_t^2$ $\forall t$. Finally, define τ^2 as the vector that vertically stacks the values for τ_j^2 $\forall j$ and define vector ρ similarly for ρ_j $\forall j$.

It is very important to note that say $e_{t,3}$ is not the error term for the third equation of the system in period t. Instead, it is the part of the third equation error term that is orthogonal to the errors of the first two equations. Likewise, τ_3^2 and ρ_3 are not the excess kurtosis and correlation coefficients for the third equation, but are the excess kurtosis and correlation coefficients for the part of the third equation error term that is orthogonal to the first two equations.

We consider several alternative models. Firstly, we impose some or all of the following restrictions

- Time-constant $b_t : Q = 0$

- Time-constant l_t and $\ln \sigma_{t+1}^2$: $R = 0, S = 0$

- Normally distributed error terms: $\tau^2 = 0$

---

[1]In order to estimate the model, we use the simulation smoother of Durbin and Koopman (2002). We use their timing convention for the innovations in the transition equations.

Secondly, we consider a pooled model in which the persistence of all time-varying parameters is equal

$$
\begin{aligned}
a_{b,(i,i)} &= \rho_P, \forall i = 1, ... nK \\
a_{l,(j,j)} &= \rho_P, \forall j = 1, ... n(n-1)/2 \\
a_{s,(k,k)} &= \rho_P, \forall k = 1, ..., n
\end{aligned}
\tag{5.12}
$$

and 0 otherwise.

Finally, we impose random walks (without drift) by setting the persistence of all time-varying parameters equal to 1

$$
\begin{aligned}
b_{t+1} &= b_t + \eta_t \\
l_{t+1} &= l_t + \zeta_t \\
\ln \sigma_{t+1}^2 &= \ln \sigma_t^2 + \xi_t.
\end{aligned}
\tag{5.13}
$$

Note that in such models, the unconditional distribution of time-varying parameters does not exist.

These alternative models are either nested in our most general model or are (in the case of the random walk specification) straightforward extensions. Since they do not lead to any further issues regarding prior choice, posterior distribution or MCMC algorithm, we do not explicitly deal with them in the next two subsections.

5.3.2 Prior

In this section, we explain the prior distributions for the most general model. It is an hierarchical model and therefore we have to define prior distributions for the initial conditions $\widetilde{b}_1$, $\widetilde{l}_1$ and $\widetilde{\ln \sigma^2}_1$ and for the hyperparameters of the model (μ_b, μ_l, μ_s, A_b, A_l, A_s, Q, R, S, ρ, τ^2). We use the first 60 months of our data-set as a training sample to estimate a time-constant VAR(1) model by OLS and use its estimates in the construction of some of the prior distributions below.

Firstly, we assume that the initial conditions are drawn from their stationary unconditional distribution

$$
\begin{aligned}
p(\widetilde{b}_1 | A_b, Q) &= N(0, \Sigma_b) \\
\end{aligned}
\tag{5.14}
$$

$$
\begin{aligned}
p(\widetilde{l}_1 | A_l, R) &= N(0, \Sigma_l) \\
\end{aligned}
\tag{5.15}
$$

$$
\begin{aligned}
p(\widetilde{\ln \sigma^2}_1 | A_s, S) &= N(0, \Sigma_s),
\end{aligned}
\tag{5.16}
$$

where $vec(\Sigma_b) = \left(I_{(nK)^2} - A_b \otimes A_b\right)^{-1} Q$ and where Σ_l and Σ_s are defined similarly.

Secondly, we consider the transition matrices. We set the off-diagonal elements of the matrices equal to 0 and assume the following for the diagonal elements

$$p(A_{b,(i,i)}) = N(m_b, v_b)I(A_{b,(i,i)}), \forall i = 1, ...nK \tag{5.17}$$

$$p(A_{l,(j,j)}) = N(m_l, v_l)I(A_{l,(j,j)}), \forall j = 1, ...n(n-1)/2 \tag{5.18}$$

$$p(A_{s,(k,k)}) = N(m_s, v_s)I(A_{s,(k,k)}), \forall k = 1, ..., n, \tag{5.19}$$

where

$$m_b = m_l = m_s = 0.9 \tag{5.20}$$

$$v_b = v_l = v_s = 0.2^2 \tag{5.21}$$

and where I(x) is equal to 1 if $-1 < x < 1$ and 0 otherwise. Hence, we impose stationary processes for the time-varying parameters. The prior correlations between the elements are equal to 0.

Thirdly, we use the training sample to specify the prior distributions for the unconditional means as follows[1]

$$p(\mu_b) = N(\widehat{b}_{OLS}, 10^6 V(\widehat{b}_{OLS})) \tag{5.22}$$

$$p(\mu_l) = N(\widehat{l}_{OLS}, 10^6 V(\widehat{l}_{OLS})) \tag{5.23}$$

$$p(\mu_s) = N(\widehat{\ln \sigma^2}_{OLS}, 10^6 V(\widehat{\ln \sigma^2}_{OLS})). \tag{5.24}$$

Mean $\widehat{b}_{OLS}$ and covariance matrix $V(\widehat{b}_{OLS})$ are respectively the standard OLS estimates and its covariance matrix. We do not have direct OLS estimates for μ_l and μ_s and its covariance matrices, but instead have an estimate of the covariance matrix of the residuals $\widehat{E}$. We draw this covariance matrix 1000 times from the inverse Wishart distribution iWishart($\widehat{E}'\widehat{E}$, T), construct l_{OLS} and $\ln \sigma^2_{OLS}$ for every draw and use its means and covariance matrix across the 1000 draws to calculate $\widehat{l}_{OLS}$, $V(\widehat{l}_{OLS})$, $\widehat{\ln \sigma^2}_{OLS}$ and $V(\widehat{\ln \sigma^2}_{OLS})$.

It is common in the literature (e.g. Primiceri (2005)) to use the variance of the estimated intercepts, slopes, volatilities and $\widehat{l}_{OLS}$'s of the time-constant model to set-up the covariance matrices of the innovations of the transition equations. We follow

---

[1]Although the parameters are unlikely to be exactly constant in the training sample, we assume that time-variation is limited over such a short time-span. Furthermore, we choose the priors as uninformative as possible such that they only have a negligible impact on the posterior distribution.

this trend and set the mean of the inverse Wishart and inverse gamma distributions equal to a constant fraction c of the covariance matrices of the estimated coefficients. For the degrees of freedom, we choose the minimum degrees of freedom such that the prior means actually exist. Constant c is specified below.

The prior distribution for Q is

$$p(Q) = iWishart(W_Q, df_Q), \tag{5.25}$$

where

$$
\begin{aligned}
W_Q &= c \times (df_Q - nK - 1) \times D_Q \tag{5.26}\\
df_Q &= nK + 2
\end{aligned}
$$

where D_Q is a diagonal matrix with the diagonal elements of $V(\hat{b}_{OLS})$ on the diagonal. It is common in the literature to impose a diagonal matrix for Q itself to reduce the total number of parameters, while there is a priori no reason to expect that the off-diagonal elements are equal to 0. In our setting, Q can be any positive definite matrix, but is shrunk towards a diagonal matrix by choosing this particular prior scale matrix. In this way, we try to find the balance between flexibility and efficiency.

There is no good guidance on how to choose the multiplication constant c. If we choose c too large, our prior implies too much time-variation. If we choose c too low, the simulation algorithm does not work smoothly. We choose c equal to 0.01. Results are not noticeably different to specifications with $c = 0.1$ or $c = 0.001$. In this way, the prior is not flat, but still diffuse and relatively uninformative. This same constant c is chosen in the priors for R and S.

In order to obtain partially analytical results for l_t, we need to impose a blockdiagonal structure for matrix R as in Primiceri (2005). Since $n = 3$, the matrix contains two blocks. The first block is formed by element $R_{1,1}$, i.e. the variance of innovations to $l_{t,(2,1)}$. The second block is formed by elements $R_{2:3,2:3}$, i.e. the covariance matrix of innovations to $l_{t,(3,1)}$ and $l_{t,(3,2)}$. This gives the following prior distributions

$$
\begin{aligned}
p(R_{(1,1)}) &= iGamma(W_{r1}, df_{r1}) \tag{5.27}\\
p(R_{(2:3,2:3)}) &= iWishart(W_{r2}, df_{r2}),
\end{aligned}
$$

where

$$
\begin{aligned}
W_{r1} &= c \times (df_{r1} - 2) \times D_{l,1} \\
W_{r2} &= c \times (df_{r2} - 3) \times D_{l,2} \\
df_{r1} &= 1 + 2 \\
df_{r2} &= 2 + 2
\end{aligned}
\tag{5.28}
$$

and where $D_{l,1}$ is $V(\hat{l}_{OLS,(1,1)})$ and $D_{l,2}$ is a diagonal matrix with the diagonal elements of $V(\hat{l}_{OLS,(2:3,2:3)})$ on its diagonal. As above, matrix $R_{(2:3,2:3)}$ can be any positive definite matrix, but we shrink it towards a diagonal matrix to get more efficient estimates.

Next, let us consider the prior covariance matrix for S and the prior for ρ jointly. In order to make the analysis tractable, we choose a diagonal matrix for S. This allows us to consider the three stochastic volatility equations separately. Let us consider the covariance matrix Σ_j^* of $e_{t,j}$ and $\xi_{t,j}$ jointly for $j = 1, .., n$

$$
\Sigma_j^* = \begin{pmatrix} 1 & \rho_j \sqrt{S_{j,j}} \\ \sqrt{\rho_j S_{j,j}} & S_{j,j} \end{pmatrix}.
\tag{5.29}
$$

It is difficult to formulate a prior for Σ_j^*, since its (1,1) element is equal to 1. Therefore, we reparameterize $S_{j,j}$ and ρ_j to be able to choose prior distributions in the way proposed in Jacquier, Polson, and Rossi (2004)

$$
\Sigma_j^* = \begin{pmatrix} 1 & \psi_j \\ \psi_j & \theta_j + \psi_j^2 \end{pmatrix}.
\tag{5.30}
$$

As in Jacquier, Polson, and Rossi (2004), we choose an inverse gamma prior for θ_j and a normal prior for $\psi_j | \theta_j$ such that we obtain a tractable algorithm. We get for $j = 1, ..., n$

$$
\begin{aligned}
p(\theta_j) &= iGamma(W_{\theta,j}, df_\theta) \\
p(\psi_j | \theta_j) &= N(0, \theta_j / p),
\end{aligned}
\tag{5.31}
\tag{5.32}
$$

where

$$
\begin{aligned}
W_{\theta,j} &= c(df_\theta - 2) V(\widehat{\ln \sigma^2}_{OLS,(j,j)}) \\
df_\theta &= 3 \\
p &= 3
\end{aligned}
\tag{5.33}
$$

Clearly, $S_{j,j} = \theta_j + \psi_j^2$ and $\rho_j = \frac{\psi_j}{\sqrt{\theta_j + \psi_j^2}}$ for $j = 1, .., n$.

Finally, we choose a prior distribution for τ_j^2 for $j = 1, ...n$. As in Omori, Chib, Shephard, and Nakajima (2007) we use a gamma prior

$$p(\tau_j^2) = Gamma(a_\tau, b_\tau) \tag{5.34}$$

with

$$
\begin{aligned}
a_\tau &= 0.25 \\
b_\tau &= 2
\end{aligned}
\tag{5.35}
$$

This prior has a mean of 0.5 and a relatively large variance of 1.

5.3.3 Posterior and MCMC

In this section, we sketch how we estimate the general model. Exact details are given in appendix 5.7. The simulation algorithm is more complicated than Primiceri (2005) due to the formulation of the initial conditions in equations (5.14), (5.15) and (5.16), the estimation of transition matrices, the estimation of the leverage effect and the presence of fat-tailed error distributions.

We use Markov Chain Monte Carlo techniques to obtain posterior distributions of the parameters of interest. In general, time-varying parameters and their unconditional means are drawn using the Kalman filter - smoother technique of Durbin and Koopman (2002). Transition matrices and covariance matrices are simulated using Metropolis-Hastings steps.

The system for b is clearly a linear Gaussian state space model (conditional on l, σ and hyperparameters) and b and μ_b can therefore be easily simulated. The system for l is in general not a linear Gaussian state space model (conditional on b, σ and hyperparameters), but can be transformed to $n-1$ linear Gaussian state space models if we impose that R is blockdiagonal. Under this assumption, sampling l and μ_l is straightforward. The system for σ^2 is also not a linear Gaussian state space model (conditional on b, l and hyperparameters), but can be transformed into n approximately linear Gaussian state space models by transforming σ_t^2 to $\ln \sigma_t^2$ and by approximating the errors of the observation equation of the state space model by a mixture of normals as suggested in Kim, Shepherd, and Chib (1998) and Omori, Chib, Shephard, and

Nakajima (2007). In this way, we can still use the simulation smoother of Durbin and Koopman (2002) to simulate $\ln \sigma^2$ and μ_s.

If the initial conditions in equations (5.14), (5.15) and (5.16) would not depend on the transition matrices and covariance matrices, we could easily simulate them using a Gibbs sampler by respectively a normal distribution and an inverse Wishart distribution. Since they do, we use these distributions as proposal densities in a Metropolis-Hastings step. Since the acceptance probabilities are all larger than 80%, this hardly deteriorates the sampling performance of the MCMC algorithm.

We do not sample the correlations ρ directly, but instead consider the transformation in equation (5.30) to make the sampling step easier. These transformed parameters are sampled using a Metropolis-Hastings algorithm using inverse gamma and normal proposal densities.

Finally, the conditional posterior distribution for τ_j^2, for $j = 1, ..., n$ is not a known distribution. Therefore, we use another Metropolis-Hastings step. We sample $\ln \tau_j^2$, because its posterior distribution is easier to approximate by a student-t proposal density. This gives an acceptance probability of 97%.

In the empirical section, we retain 10,000 iterations after a burn-in period of 5,000 iterations. We draw 10 asset return paths per iteration (hence 100,000 in total) to calculate predictive distributions and portfolio weights. Increasing the number of iterations does not significantly impact results.

5.4 Results basic specifications

We report results for the basic specification with the dividend-to-price ratio as predictor. We explain estimation results, consider the term structures of risk and portfolio weights for the time-varying model, provide an assessment of the individual components of the time-varying model and finally analyze the importance of the persistence of time-varying parameters.

5.4.1 Estimation results

Figure 5.1 plots the time-series of (smoothed) posterior means of the time-varying intercepts and slope coefficients for both the time-varying and constant model. A few results stand out. Firstly, the dividend-to-price ratio is a positive predictor of excess

Figure 5.1: Time-series of the posterior means of intercepts and slopes
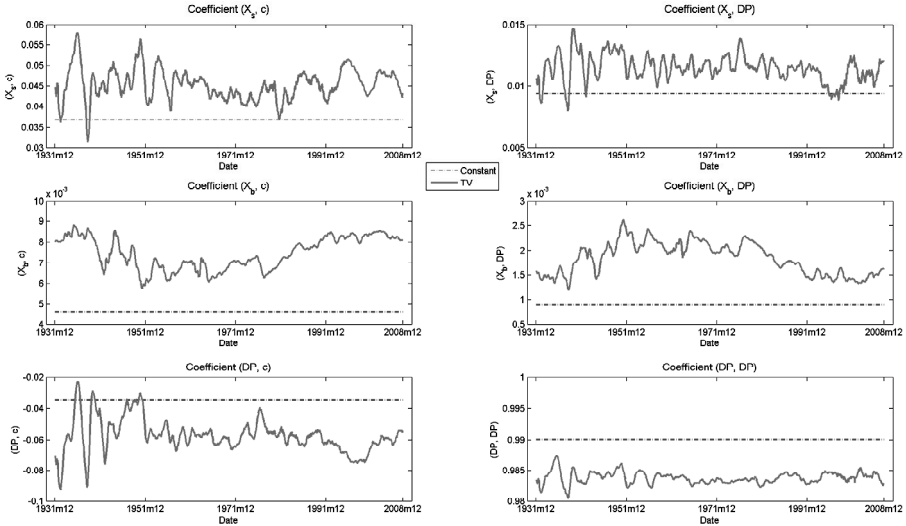
This figure shows the (smoothed) posterior means of intercepts and slopes for the time-varying and the time-constant model. Results are based on 10,000 retained draws. Note that the equations are given in the different rows.

stock and bond returns in every period. Both coefficients are on average quite a bit higher than the ones in the constant model. Secondly, both prediction coefficients show some modest time-variation. For example, the posterior mean in the stock return equation varies between 0.008 and 0.015. The time-variation is however not very persistent, since deviations from means do not last very long. Thirdly, the AR(1) coefficient for the dividend-to-price ratio is on average lower than the one in the constant model and varies over time. Such a change in persistence of a predictor can potentially have a large impact on long-term investors. The constant in the predictor equation varies quite a lot over time as well.

Table 5.4 shows the posterior means and standard deviations of the most important hyperparameters of several specifications. The time-varying model is in the last column. The other specification are considered in subsection 5.4.3. The first 12 rows show the posterior means and standard deviations of the diagonal elements of A_{slope}. The table

Table 5.4: Posterior means and standard deviations of hyperparameters

This table shows the posterior means and standard deviations (between brackets) of the most important hyperparameters of seven specifications. The specifications differ in whether b is time-varying (1) / time-constant (0), whether Ω is time-varying (1) / time-constant (0) and finally whether error terms have fat tails (1) / are gaussian (0). Matrices A_b, A_s and A_l are the transition matrices for b_t, $\ln \sigma_t^2$ and l_t. Vectors ρ and τ^2 are respectively the correlation between error term e_t and ξ_t, and a kurtosis measure for e_t.

TV b	0	0	0	1	1	1	1
TV Ω	0	1	1	0	0	1	1
Kurt	1	0	1	0	1	0	1
$A_{b,(1,1)}$				0.3110	0.3209	0.9153	0.9109
				(0.0480)	(0.0483)	(0.0318)	(0.0347)
$A_{b,(2,2)}$				0.6703	0.6623	0.8804	0.8837
				(0.0884)	(0.1051)	(0.0319)	(0.0321)
$A_{b,(3,3)}$				0.8112	0.9333	0.8605	0.8637
				(0.1354)	(0.1107)	(0.1416)	(0.1213)
$A_{b,(4,4)}$				0.7764	0.8337	0.8383	0.8779
				(0.1627)	(0.1519)	(0.1463)	(0.1310)
$A_{b,(5,5)}$				0.9112	0.8853	0.9411	0.9436
				(0.0178)	(0.0205)	(0.0163)	(0.0144)
$A_{b,(6,6)}$				0.4837	0.5874	0.8973	0.9021
				(0.0650)	(0.0784)	(0.0355)	(0.0340)
$A_{s,(1,1)}$		0.9538	0.9589			0.9495	0.9555
		(0.0184)	(0.0175)			(0.0183)	(0.0169)
$A_{s,(2,2)}$		0.9784	0.9788			0.9773	0.9808
		(0.0090)	(0.0090)			(0.0091)	(0.0083)
$A_{s,(3,3)}$		0.9865	0.9860			0.9984	0.9985
		(0.0067)	(0.0065)			(0.0016)	(0.0015)
$A_{l,(1,1)}$		0.9964	0.9959			0.9962	0.9960
		(0.0028)	(0.0033)			(0.0030)	(0.0032)
$A_{l,(2,2)}$		0.8168	0.8069			0.8352	0.8506
		(0.1328)	(0.1313)			(0.1183)	(0.1292)
$A_{l,(3,3)}$		0.7925	0.7733			0.8182	0.7599
		(0.1276)	(0.1567)			(0.1361)	(0.1525)
τ_1^2	0.8802		0.0592		0.9145		0.1269
	(0.1368)		(0.0506)		(0.1422)		(0.0702)
τ_2^2	1.3335		0.0445		1.6542		0.1626
	(0.2011)		(0.0723)		(0.2569)		(0.1147)
τ_3^2	1.8091		0.0003		0.2506		0.0750
	(0.2191)		(0.0004)		(0.3696)		(0.0618)
ρ_1		-0.3249	-0.3513			-0.3629	-0.4266
		(0.1016)	(0.1027)			(0.1200)	(0.1297)
ρ_2		-0.0856	-0.1024			-0.1108	-0.1532
		(0.0902)	(0.0981)			(0.1094)	(0.1250)
ρ_3		0.0714	0.0711			-0.5311	-0.4787
		(0.0511)	(0.0530)			(0.2148)	(0.2351)

confirms that time-variation in intercepts and slopes is not very persistent. The average posterior mean of these persistence parameters is around 0.90, which implies a half life of an innovation of only a bit more than 6 months.

We conclude that there is time-variation in intercepts and slopes, but that this time-variation is rather small and not very persistent.

Figure 5.2 reports the time-series of (smoothed) posterior means of the covariance matrix of the error terms. It reports volatilities on the diagonal, correlations above the diagonal and values for l_t below the diagonal. The figure shows some interesting results. Firstly, the error volatility in the stock return equation varies considerably over time. It ranges from 17% per month in the 1930s to 2.5% in the 1960 and 1970s. The figure also shows that time-variation is very persistent, since volatility reverts only slowly to

Figure 5.2: Time-series of the posterior means of error volatilities, correlations and l_t
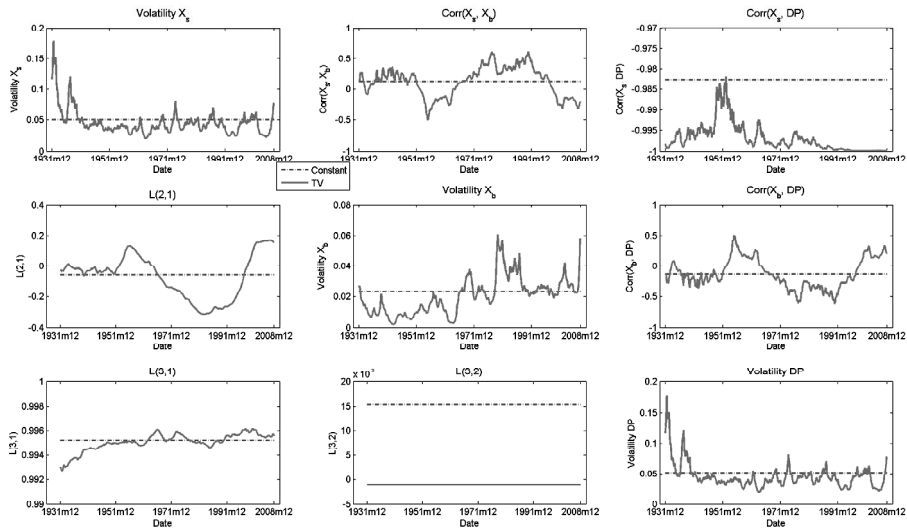
This figure shows the (smoothed) posterior means of the standard deviation of the residuals (diagonal), the correlation between the residuals (above the diagonal) and the l_t coefficients (below the diagonal) for the time-varying and time-constant specifications. Results are based on 10,000 retained draws. Note that the equations are given in the different rows.

its mean. Secondly, there is a lot of time-variation in the correlation between stocks and bonds. Correlation can be positive as in the 1970s and 1980s, but can also be negative as in more recent times. The time-variation in correlation is also very persistent. The correlation between stock and dividend-to-price ratio innovations also varies somewhat, but stays very close to -1. Thirdly, the error volatility for the dividend-to-price ratio almost exactly mimics the error volatility for stock returns. This is not surprising because of the consistently strong negative correlation between them. Fourthly, the error volatility of bond returns varies a lot over time. It reaches its maximum of 6% per month at the end of the sample. Its error correlation with the dividend-to-price ratio is almost exactly the negative of its error correlation with stocks.

Table 5.4 shows some interesting results for the error covariance matrix. Firstly, it indicates that time-variation in error volatilities and error correlations are much more persistent than variation in intercepts and slopes. The posterior mean of the persistence parameter of error volatility for the stock return equation is 0.96. This implies a half-life of almost 1 1/2 years. The other error volatilities are even more persistent. Secondly, there is hardly any excess kurtosis left after taking time-variation in volatility into account. For example, parameter τ_1^2 shows that the kurtosis of the error term in the stock equation is very close to 3 instead of more than 10 as in table 5.2. Hence, if one ignores time-variation in volatility, one would wrongly conclude that the distribution of error terms deviates substantially from normality.

Posterior mean $E(\rho_1|Y) = -0.43$ shows that there is a leverage effect in stock returns at the monthly horizon even if predictors are included. This means that three aspects play a role if there is a negative shock to stock returns. Firstly, a negative innovation leads to higher expected future stock returns due to the mean-reversion effect. Secondly, the negative innovation increases the error volatility for stocks. Thirdly, the negative shock increases the error volatility of the dividend-to-price ratio, since $e_{t,1}$ is a major component of the error for this ratio.[1]

The first effect is beneficial for long-term investors and implies that long-term investors should invest more in stocks than short term investors as shown in e.g. Campbell, Chan, and Viceira (2003). However, the second and third effect increase the risk

---

[1]The error for the dividend-to-price ratio in period t is a linear combination of $e_{t,1}$, $e_{t,2}$ and $e_{t,3}$. The almost perfect negative correlation between the error in the dividend-to-price ratio and the error in stock returns shows that the latter two components hardly matter for the error in the dividend-to-price ratio equation.

for long-term investors and therefore should reduce their stock allocations. The preceding suggests that ignoring time-variation in volatility leads to an overinvestment in stocks at long horizons. The second effect is considered in isolation in Chacko and Viceira (2005), who conclude that it leads to a modest negative hedge term for long-term investors. The third effect is ignored in the literature. There are no papers that consider all three effects jointly as we do.

We conclude that there is a lot of time-variation in error volatility and correlation and that this time-variation is very persistent and relevant for long-term investors.[1]

5.4.2 Term structure of risk and portfolio weights

Figure 5.3 reports the term structures of risk for excess log stock returns for both the time-varying and time-constant model. It shows the annualized predictive volatility of future cumulative stock returns. We obtain these figures by simulating stock returns from the predictive distribution of future stock returns. In the upper panel we draw time-varying parameters from its unconditional posterior distribution, draw the time-constant parameters from its posterior distribution and we set the dividend-to-price ratio equal to its historical average. In the bottom panel, we simulate the time-varying parameters from the posterior distribution in December 2008, we simulate the time-constant parameters from its posterior distribution and we set the dividend-to-price ratio equal to its December 2008 value.

The term structures of risk takes parameter uncertainty, state uncertainty and uncertainty due to the error term into account. An investor faces state uncertainty, since she does not know the exact values of the time-varying parameters (states) and only knows their posterior distribution at any point in time.

The figure shows that the term structure of risk for the time-varying specification varies a lot over time. It can either be upward sloping (upper panel) or downward

---

[1]A note on methodology. It is well-known in the time-series literature (e.g. Breusch and Pagan (1979)) that one can rewrite an AR(1) model with random coefficients as an AR(1) with heteroscedastic errors. In such a setting, the extra variation due to the random coefficient only leads to a heteroscedastic error term, but does not affect the mean. However, in our setting, the time-varying parameters are autocorrelated over time, i.e. if for example the autocorrelation of stock returns is high today, it will also be (relatively) high tomorrow. Therefore, the time-variation in say the slope parameter in for example an AR(1) does not only affect the conditional variance but also the conditional mean. Therefore, such a model cannot simply be rewritten as an heteroscedastic AR(1).

Figure 5.3: Annualized predictive standard deviation of excess log stock returns: time-varying and time-constant specifications
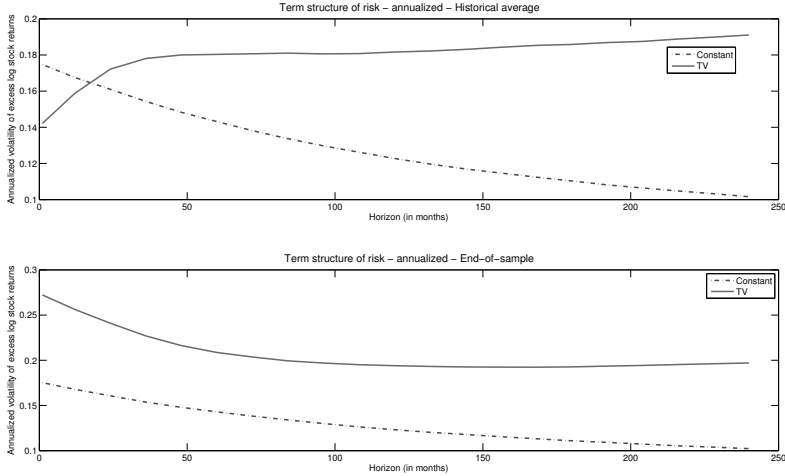
The figure shows the annualized predictive standard deviation of excess log stock returns for the time-varying and time-constant specifications. Time-varying parameters are either drawn from its unconditional distribution (upper panel) or from its posterior distribution at the end of the sample (lower panel). The predictor variable is either set to its historical average (upper panel) or to its end-of-sample value (lower panel). Results are based on 100,000 retained draws.

sloping (bottom panel). Let us firstly consider the term structure in the bottom panel. The error volatility at the 1-month horizon is very high (27%). Figure 5.2 already showed that at the end of the sample, the error volatility for stock returns is much higher than its mean. At medium horizons, the annualized predictive volatility is however much lower, because (i) mean-reversion in stock returns makes stocks safer in the long-run and (ii) the error volatility reverts back to its (lower) long-run mean. At horizons of 15 years or more, the term structure is slightly upward sloping due to the effect of parameter uncertainty. At the 20-year horizon, annualized volatility is almost 20%.

If time-varying parameters are drawn from its unconditional distribution (upper

panel), we see a different picture. Predictive volatility starts low and increases up to 19% at the end of the 20-year horizon. Due to the combined effect of parameter uncertainty and time-variation in parameters, the term structure of risk is upward sloping for all horizons. At a 20-year horizon, predictive volatility is around 20% in both panels.

In both panels, the term structures deviate strongly from their time-constant counterpart. The latter is always strongly downward sloping. Apparently, if one ignores model instability, mean-reversion strongly dominates parameter uncertainty. Especially at the end of the sample, this gives a false sense of security. Stocks are much riskier if one takes time-variation in parameters into account.

Next, we calculate stock and bond weights for buy-and-hold investors who want to maximize expected power utility at time t over terminal wealth at time $t + K$

$$\max_{w_s, w_b} E_t \left(\frac{W_{t+K}^{1-\gamma}}{1 - \gamma} \right), \tag{5.36}$$

where relative risk aversion $\gamma = 5$, where $w = (w_s, w_b)'$, where

$$W_{t+K} = \prod_{j=1}^{K} \left(1 + w' \exp(R_{tbill}\iota + x_{t+j}) + (1 - w'\iota) \exp(R_{tbill}\iota) \right) \tag{5.37}$$

and where ι is an $(n - k) \times 1$ vector of ones. We consider investors who either consider the time-varying or time-constant specification to calculate portfolio weights at both the end of the sample and when predictor variables are equal to its historical average. As above we draw time-varying parameters from its unconditional distribution in the latter case.

Figure 5.4 plots the results. Let us firstly consider the constant model. It is well-known - e.g. Campbell and Viceira (2002), Barberis (2000) - that the stock investment curve (plot of stock weights versus investment horizon) is upward sloping in settings without model instability. We find similar results both at the end of the sample and at the historical average.

If we take time-variation in model parameters into account, the stock investment curve can however be strongly downward sloping. This is the case at the historical average. This is not a surprising result given the term structure of risk we showed in the previous figure. At the end of the sample, the stock investment curve is initially upward sloping, but becomes downward sloping for longer horizons. We conclude that

Figure 5.4: Optimal stock and bond weights: time-varying and time-constant specifications
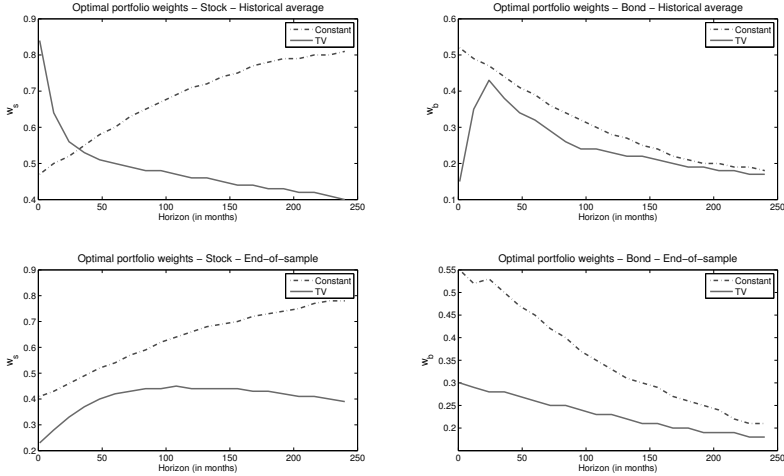
The figure shows the optimal stock weights (first column) and optimal bond weights (second column) for a buy-and-hold investor with $\gamma = 5$ who either uses the time-varying specification or the time-constant specification. Time-varying parameters are either drawn from its unconditional distribution (first row) or from its posterior distribution at the end of the sample (second row). The predictor variable is either equal to the historical average (first row) or to the end-of-sample value (second row). Results are based on 100,000 retained draws.

this curve can have all kind of shapes if time-variation in parameters is taken into account. The bond investment curve is downward sloping in general. Apparently, long-term (constant maturity) bonds are unattractive for long-term investors.

If we compare the results for the time-constant and time-varying specifications to each other, we see that portfolio weights differ a lot. For example, at the historical average, stock weights differ more than 40% at both short and long horizons. Ignoring model instability can therefore lead to huge investment mistakes.

In order to quantify investment mistakes, figure 5.5 plots the certainty equivalent returns (CERs) for buy-and-hold investors who either base their portfolio weights on the time-varying or time-constant specifications. In both cases, we calculate the CERs

Figure 5.5: Annualized certainty equivalent returns: time-varying and time-constant specifications
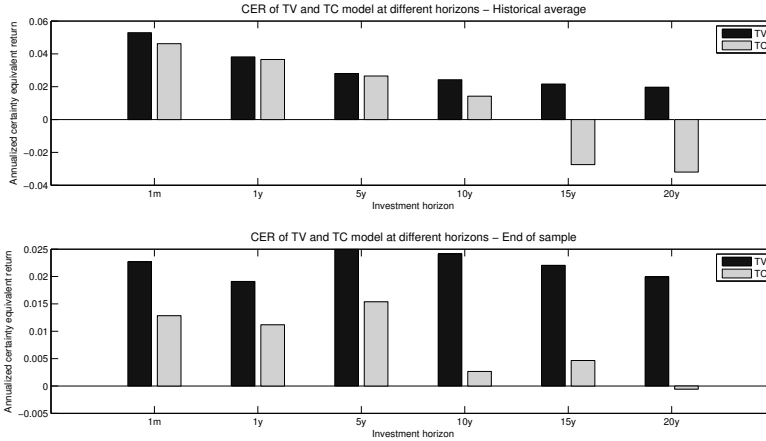
This figure shows the annualized certainty equivalent returns for buy-and-hold investors with $\gamma = 5$ who either base their portfolio weights on the time-varying specification or the time-constant specifications. The certainty equivalent returns are calculated using the time-varying specification. Time-varying parameters are either drawn from its unconditional distribution (first row) or from its posterior distribution at the end of the sample (second row) and the predictor variable is either equal to its historical average (first row) or to its end-of-sample value (second row). Results are based on 100,000 retained draws.

using the time-varying specification. The certainty equivalent return for the time-varying specification is therefore by construction the highest. We consider the difference in CERs between both specifications. If ignoring time-variation leads to a considerable reduction in CER for an investor, we can conclude that ignoring model instability is economically costly.

The figure shows some interesting results. Firstly, the average CER across horizons is much higher at the historical average. Apparently, even at medium to long horizons, the unfavorable initial state leads to a considerable reduction in the attractiveness of financial markets. Secondly, at long horizons, ignoring time-variation can be really costly for investors. Differences can be as large as 5% per year. This is not surprising, since the previous figure showed that investment mistakes can be as large as 40%.

Thirdly, if initial states deviate from their average values, it is also costly to ignore time-variation at short horizons. Even at the 1-month horizon, CERs are already reduced from more than 2% to 1%.

We conclude that it is economically important to take time variation into account. Ignoring time variation leads to a considerable underestimation of perceived risk, an overinvestment in the stock market and therefore a much lower performance. These effects are especially pronounced at long horizons and when state variables deviate from their historical averages.

5.4.3 A closer look at the time-variation of model components

The time-varying model shows the joint effect of time-variation in intercepts and slopes, time-variation in error volatility and error correlation and the non-normality of the error terms on term structures of risk, portfolio weights and portfolio performance. In this section, we are interested in assessing the importance of the individual components of the time-varying model.

One way to do this would be to use statistical criteria such as posterior model probabilities. We do not pursue this alternative here because of two reasons. Firstly, it is well-known that posterior model probabilities are very sensitive to prior assumptions. Especially if one tries to choose prior distributions as uninformative as possible (as we do), the prior choice can have unintended consequences for model probabilities. Secondly, the portfolio weights and their expected utility are what matters ultimately for long-term investors.

Our strategy is to choose the most general model that nests all submodels and compare the portfolio weights it gives to the portfolio weights for the submodels. These submodels are obtained by omitting some or all of the model components. We evaluate all portfolio weights using certainty equivalent returns calculated using the most general specification. The portfolio weights based on this most general model therefore have by definition the highest certainty equivalent return. What matters is whether other specifications lead to much lower certainty equivalent returns. If the omission of let's say time-varying slopes does not lead to a substantial loss of performance in a world where slopes indeed vary over time, then a risk-averse investor can safely ignore such an effect.

5. MODEL INSTABILITY AND LONG-TERM INVESTORS

Table 5.4 in section 5.4.1 shows the posterior means and standard deviation of the most important hyperparameters of seven specifications that differ in whether b is time-varying / time-constant, whether Ω is time-varying / time-constant and whether the distributions of the errors contain fat tails / are Gaussian.[1] Firstly, the table shows that ignoring stochastic volatility (column 1) leads to a large overestimation of the kurtosis of the error terms (τ^2). This is in line with results in section 5.4.1 and table 5.2. Secondly, the persistence parameters for volatility are very robust across specifications. This suggests that the modeling of the time-variation in the error covariance matrix is not sensitive to the exact specification of intercepts and slopes. Thirdly, ignoring stochastic volatility implies a large underestimation of the persistence of the time-varying slopes and intercepts. For example, $A_{b,(1,1)}$ is reduced from 0.91 to 0.32 if the error covariance matrix is constant. The reason is that ignoring stochastic volatility leads to many transitory movements in the time-varying slopes and intercepts due to outliers in periods when true volatility was actually very high. These transitory movements reduce the persistence of time-varying slopes and intercepts.

Table 5.5 shows the means and standard deviations of the posterior means of intercepts, slopes, error volatilities and error correlations over time. The table shows that the posterior means of time-varying intercepts and slopes are too variable if stochastic volatility is ignored. As above, this is caused by the presence of outliers in time-varying b_t if the error covariance matrix is constant. Furthermore, we clearly see that time-variation in intercepts and slopes is modest, while time-variation in error volatility and error correlation is substantial across specifications. Finally, the table indicates that the unconditional means of the parameters vary quite a bit across specifications. For example, the incorporation of time-variation in Ω leads to large differences. If the latter is incorporated, periods with large volatility are underweighted when estimating unconditional means.

Figure 5.6 plots the annualized CERs for specifications that include some or all three model components (time-varying intercepts/slopes, time-varying covariance matrix, excess kurtosis). In all subpanels, the first specification is the time-constant model and the fourth specification is the time-varying model. The second and third specifi-

---

[1]The posterior means of the hyperparameters for the time constant model are not reported in the table, since they are all 0.

Table 5.5: Means and standard deviations of the posterior means of the time-varying parameters

This table plots the means and standard deviations of the posterior means of the time-varying parameters over time for eight specifications. The specifications differ in whether b is time-varying (1) / time-constant (0), whether Ω is time-varying (1) / time-constant (0) and finally whether error terms have fat tails (1) / are gaussian (0). The first part of the table shows the means and standard deviations of the posterior mean of b_t over time. The second part gives the means and standard deviations of the posterior mean of Ω_t over time. The diagonal elements show the moments of the error volatilities and the off-diagonal elements show the moments of the error correlations.

TV b	0	0	0	0	1	1	1	1
TV Ω	0	0	1	1	0	0	1	1
Kurt	0	1	0	1	0	1	0	1
$b_{1,1}$	0.0368	0.0324	0.0280	0.0273	0.0405	0.0337	0.0454	0.0453
					(0.0085)	(0.0070)	(0.0042)	(0.0038)
$b_{1,2}$	0.0094	0.0072	0.0063	0.0061	0.0105	0.0076	0.0115	0.0114
					(0.0009)	(0.0017)	(0.0011)	(0.0011)
$b_{2,1}$	0.0046	0.0027	0.0070	0.0070	0.0092	0.0144	0.0078	0.0075
					(0.0006)	(0.0032)	(0.0008)	(0.0008)
$b_{2,2}$	0.0009	0.0003	0.0019	0.0019	0.0023	0.0038	0.0020	0.0018
					(0.0004)	(0.0004)	(0.0002)	(0.0003)
$b_{3,1}$	-0.0349	-0.0254	-0.0217	-0.0210	-0.0396	-0.0384	-0.0562	-0.0577
					(0.0114)	(0.0137)	(0.0106)	(0.0108)
$b_{3,2}$	0.9900	0.9939	0.9949	0.9951	0.9886	0.9898	0.9841	0.9838
					(0.0025)	(0.0019)	(0.0009)	(0.0010)
$\Omega_{1,1}$	0.0505	0.0507	0.0456	0.0458	0.0491	0.0500	0.0447	0.0449
			(0.0207)	(0.0203)			(0.0212)	(0.0205)
$\Omega_{1,2}$	0.1287	0.0951	0.1235	0.1249	0.1445	0.0919	0.1242	0.1266
			(0.2308)	(0.2415)			(0.2365)	(0.2402)
$\Omega_{1,3}$	-0.9827	-0.9825	-0.9806	-0.9810	-0.9999	-0.9999	-0.9969	-0.9970
			(0.0284)	(0.0278)			(0.0036)	(0.0034)
$\Omega_{2,2}$	0.0233	0.0243	0.0204	0.0205	0.0231	0.0246	0.0201	0.0204
			(0.0117)	(0.0117)			(0.0118)	(0.0118)
$\Omega_{2,3}$	-0.1333	-0.0963	-0.1244	-0.1258	-0.1428	-0.0933	-0.1240	-0.1258
			(0.2286)	(0.2393)			(0.2360)	(0.2397)
$\Omega_{3,3}$	0.0512	0.0517	0.0464	0.0465	0.0493	0.0498	0.0446	0.0448
			(0.0211)	(0.0207)			(0.0211)	(0.0203)

Figure 5.6: Annualized certainty equivalent returns: individual model components
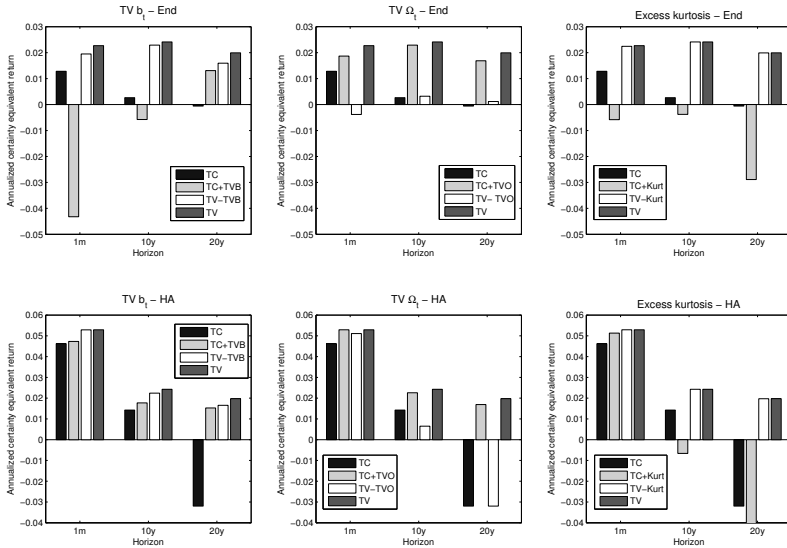
This figure shows the annualized certainty equivalent return for a buy-and-hold investor with $\gamma = 5$ who uses several specifications to calculate portfolio weights. These specifications differ in whether they allow for time-varying intercepts/slopes, a time-varying covariance matrix and excess kurtosis. The certainty equivalent returns are calculated using the time-varying specification. We report results for the 1 month horizon, the 10 years horizon and the 20 years horizon. Time-varying parameters are either drawn from its posterior distribution at the end of the sample (first row) or its unconditional distribution (second row). Results are based on 100,000 retained draws. The first and fourth column in every subpanel report the time-constant model and the time-varying model. The second column shows a specification where either the time-varying b_t, time-varying covariance matrix or excess kurtosis is added to the time-constant model. The third column reports the results for a setting in which either time-varying b_t's, time-varying covariance matrix or excess kurtosis is removed from the time-varying model.

cation respectively add the individual model component to the time-constant model or remove the individual model component from the fully time-varying model.

Let's firstly consider time-varying intercepts/slopes. The figure shows that it can be very costly (at the end-of-the-sample) to add time-varying intercepts/slopes to the time-

constant model. The incorporation of time-varying b_t's, while ignoring time-varying volatility leads to a misspecified model. On the other hand, removing time-varying intercepts/slopes from the fully time-varying model is not costly. It only leads to a modest loss in CERs at all horizons.

Secondly, we consider the time-variation of the error covariance matrix. The figure shows that its inclusion always improves performance irrespective of the horizon. In fact, just including the time-varying error covariance matrix while ignoring time-varying b_t's or excess kurtosis almost leads to the maximum performance. The omittance of the time-variation in the error covariance matrix however turns out to be costly in almost all cases. The exception is the one-month horizon when predictor variables are equal to its historical average.

Finally, let us look at the incorporation of excess kurtosis. Clearly, adding excess kurtosis to the time-constant model only hurts performance. The time-constant model contains fat tails, because time-varying volatility is ignored. Apparently, allowing for non-normal distributions while ignoring stochastic volatility leads to a misspecified model and deteriorated performance. The removal of excess kurtosis from the fully time-varying model hardly has an affect on performance. This is not surprising, since previous sections show that the error distribution is close to normal once time-varying volatility is incorporated.

We conclude that the incorporation of time-varying intercepts/slopes and fat tails hardly has a positive effect on performance and can therefore be safely omitted. However, it is extremely important to incorporate a time-varying error covariance matrix. Its omission drastically reduces certainty equivalence returns.

5.4.4 The persistence parameter

The persistence of time-varying parameters plays a crucial role in the portfolio formation of long-term investors. This section analyzes the importance of the persistence by considering two alternative specifications. The first specification sets all persistence parameters equal to the same value. We estimate this specification to *approximate* regime-switching models, since all parameters have equal persistence in such models. Note that this pooled model is nested in our most general model and is therefore a restricted version of this most general model. The aim of this section is to analyze the economic losses incurred by imposing these restrictions.

Figure 5.7: Time-series of the posterior means of the intercepts and slopes for the time-varying, pooled and random walk specifications
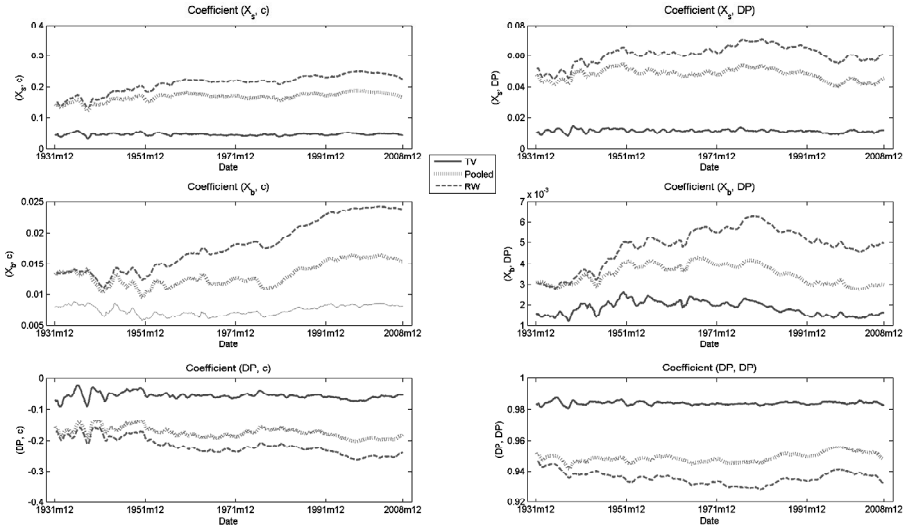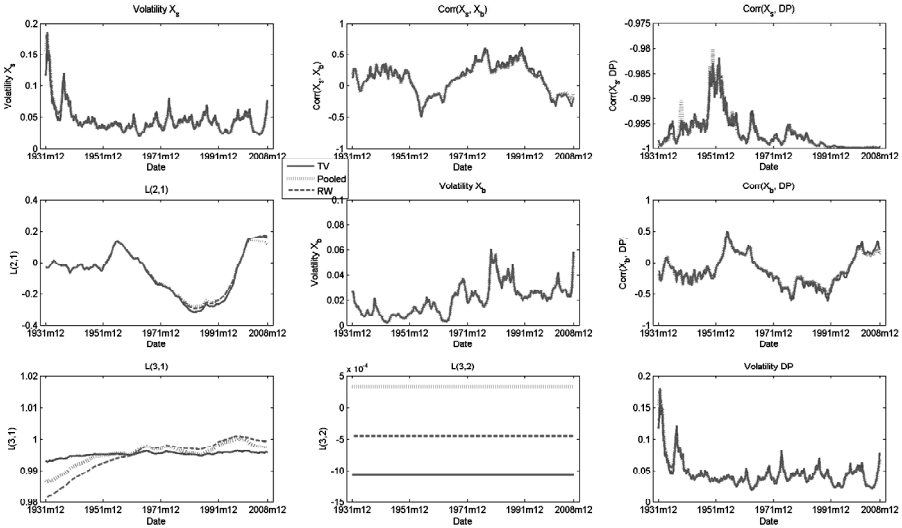

The figure shows the (smoothed) posterior means of the intercept and slopes for the time-varying model, the model in which the persistence parameters are equal for all time-varying parameters and the random walk specification. Results are based on 10,000 retained draws. Note that the equations are given in different rows.

The second specification sets all persistence parameters equal to 1. We consider this random walk specification, because it is a popular way to model time-varying parameter models (e.g. Primiceri (2005)). Note that it is not nested in our most general specification. However, the results in the previous section show that there is no evidence at all that the persistence parameters for intercepts and slopes can be set to 1. The aim is again to assess the economic losses incurred if one uses this alternative specification.[1]

Figure 5.7 reports the time-series of the posterior means of the intercepts and slope coefficients for three specifications: (i) the time-varying specification, (ii) the pooled

---

[1]In unreported results, we also consider alternative specifications that nest the random walk specification. These specifications do no impose the existence of the long-run mean of the time-varying parameters. Results are in line with the results in this paper, i.e. no evidence for persistent variation in intercepts/slopes.

Figure 5.8: Time-series of the posterior means of error volatilities, correlations and l_t for the time-varying, pooled and random walk specifications
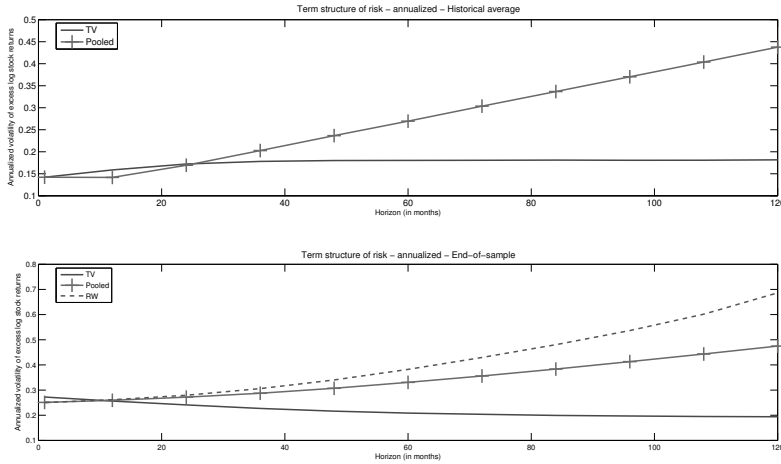

The figure shows the (smoothed) posterior means of the standard deviation of the residuals (diagonal), the correlation between the residuals (above the diagonal) and the l_t coefficients (below the diagonal) for the time-varying model, the model in which the persistence parameters are equal for all time-varying parameters and finally the random walk specification. Results are based on 10,000 retained draws. Note that equations are given in the different rows.

specification and (iii) the random walk specification. The figure indicates that the three specifications lead to very different posterior means. Both the pooled as well as the random walk model are quite different from the time-varying model. Note that even for the random walk specification, the time-variation in the parameters is still relatively modest. The differences across specifications suggest that "restricting" the persistence parameters to 1 or pooling the persistence parameters can lead to very different results.

Figure 5.8 plots time-series of the posterior means of error volatilities, correlations and parameters l_t for the three specifications. Remarkably, it is hardly possible to see any difference between the three models with the naked eye. There is some difference in $l_{t,(3,1)}$ and $l_{t,(3,2)}$, but this does not lead to any noticeable differences in the posterior

Figure 5.9: Annualized predictive standard deviation of excess log stock returns: time-varying, pooled and random walk specifications
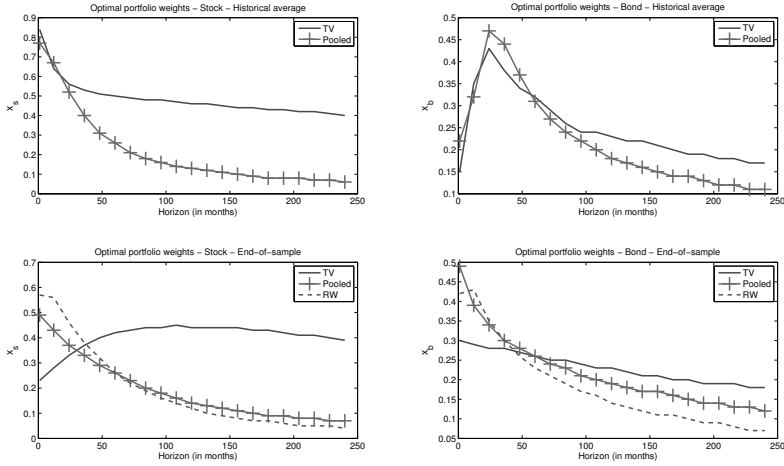

The figure shows the annualized predictive standard deviation of excess log stock returns for the time-varying specification, the specification in which the persistence parameters are pooled for all time-varying parameters and finally the random walk specification. Time-varying parameters are either drawn from its unconditional distribution (upper panel) or from its posterior distribution at the end of the sample (lower panel). The predictor variable is either equal to its historical average (first row) or to its end-of-sample value (second row). Results are based on 100,000 retained draws.

means of volatilities or correlations. We conclude that pooling persistence parameters or setting them to 1 is a viable alternative for the error covariance matrix.

Figure 5.9 plots the term structures of risk for the three specifications. Since the unconditional distribution of time-varying parameters is not defined for the random walk specification, we do not plot its term-structure at the historical average. The figure shows that restrictions lead to completely different term structures of risk. For horizons longer than 2 years, the term structures for the pooled and especially random walk specification increase extremely fast. At a 10 year horizon, the annualized predictive volatility is already more than 44%. For horizons longer than 10 years (not plotted), the predictive volatility reaches unrealistically high values. This suggests that restrictions

Figure 5.10: Optimal stock and bond weights: time-varying, pooled and random walk specifications

The figure shows the optimal stock weights (first column) and optimal bond weights (second column) for a buy-and-hold investor with $\gamma = 5$ who bases her portfolio weights on either the time-varying specification, the specification in which the persistence parameters are pooled for all time-varying parameters and finally the random walk specification. Time-varying parameters are either drawn from its unconditional distribution (first row) or from its posterior distribution at the end of the sample (second row). The predictor variable is either equal to its historical average (first row) or to its end-of-sample value (second row). Results are based on 100,000 retained draws.

should only be imposed for short horizons.

Figure 5.10 plots the portfolio weights for the three specifications. We again consider a buy-and-hold investor who maximizes expected power utility over final wealth with risk aversion parameter $\gamma = 5$. The figure shows that bond weights are very similar for all three models. However, the figure also shows that a risk-averse investor who either uses the pooled or random walk specification is much too conservative. Such an investor hardly invests in stocks at long horizons, since stock returns are much too risky in her eyes (see previous figure).

Finally, figure 5.11 plots the certainty equivalent returns for the three specifications.

197

Figure 5.11: Annualized certainty equivalent returns: time-varying, pooled and random walk specifications
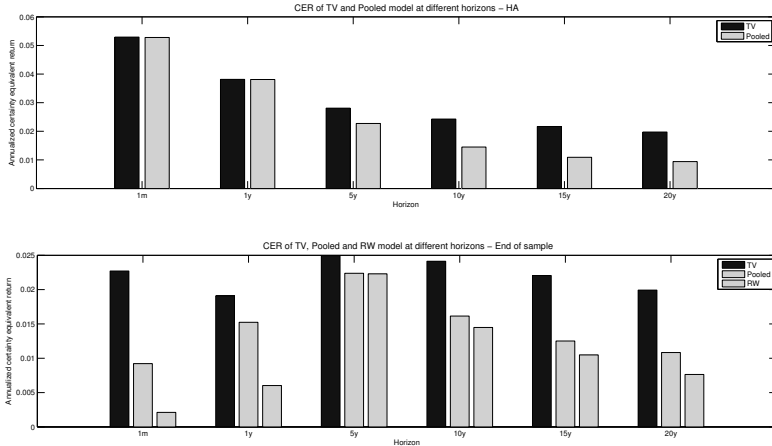
The figure shows the annualized certainty equivalent return for a buy-and-hold investor with $\gamma = 5$ who either bases her portfolio weights on the time-varying specification, the specification in which the persistence parameters are pooled for all time-varying parameters and finally the random walk specification. All certainty equivalent returns are calculated using the time-varying specification. Time-varying parameters are either drawn from its unconditional distribution (first row) or from its posterior distribution at the end of the sample (second row). The predictor variable is either equal to its historical average (first row) or to its end-of-sample value (second row). Results are based on 100,000 retained draws.

They are all evaluated under the time-varying model. Let us consider the situation at the historical average. The figure shows that an investor who uses a restricted specification hardly loses at very short horizons. Such an investor only starts to lose at horizons of 10 years or more. However, at the end of the sample, the situation is different. Here, investors lose at very short or very long horizons. The losses are still acceptable for medium horizons.

We conclude that it is economically important that persistence parameters are not restricted for different time-varying parameters. Restricting the parameters to be equal or imposing a random walk leads to an overestimation of risk, an underinvestment in

the stock market and therefore to a deteriorated performance at especially the longest horizons.

5.5 Robustness

Our main specification contains the dividend-to-price ratio as predictor for both stock and bond returns. One can argue that a different predictor for especially bond returns is more appropriate. In this section, we therefore consider the yield spread as an alternative predictor. For the sake of brevity, we only focus on the time-series of the posterior means of the time-varying parameters.

Figure 5.12 plots the posterior means of the intercepts and slopes over time. It reports results for the constant and time-varying model. The figure largely confirms the results of section 5.4.1. Firstly, there is modest time-variation in both intercepts

Figure 5.12: Time-series of the posterior means of the intercepts and slopes for Y_{spr} model
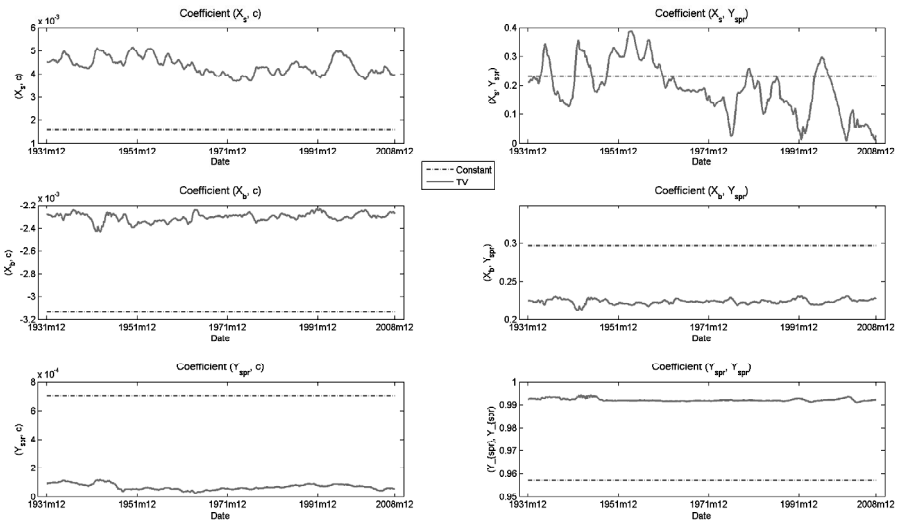
The figure shows the (smoothed) posterior means of the intercept and slopes for the time-varying and time-constant Y_{spr} - model. Results are based on 10,000 retained draws. Note that the equations are given in the different rows.

Figure 5.13: Time-series of the posterior means of error volatilities, correlations and l_t for the Y_{spr} model
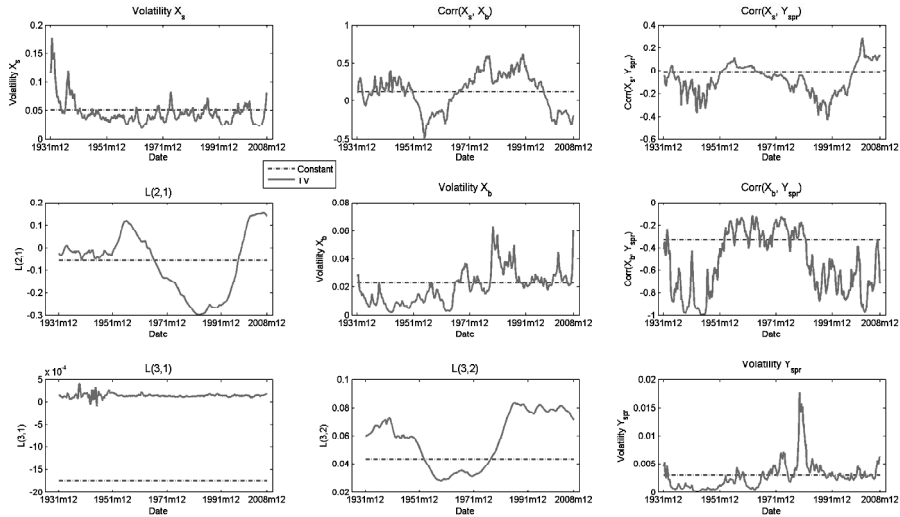
The figure shows the (smoothed) posterior means of the standard deviation of the residuals (diagonal), the correlation between the residuals (above the diagonal) and the l_t coefficients (below the diagonal) for the time-varying and time-constant Y_{spr} - specifications. Results are based on 10,000 retained draws. Note that the equations are given in the different rows.

and slopes in especially the stock return equation. The posterior mean of the yield spread coefficient varies between 0 and 0.40. Secondly, the time-variation is not very persistent and therefore not important for long-term investors. The diagonal elements of A_{slope} are all around 0.85 which implies a half-life of less than half a year. Thirdly, the average coefficients differ quite a lot from the constant model. The difference is especially large for the lagged yield spread in the yield spread equation, 0.99 versus 0.96. This change in persistence is very important for long-term investors.

In order to understand the large change in persistence, we need to consider time-variation in the error covariance matrix. Figure 5.13 shows the posterior means of the error volatilities, error correlations and l_t coefficients for both the time-constant and time-varying model. The figure confirms results that there is considerable time-variation in correlation and volatility. Firstly, the most remarkable result is the large

change in error volatility for the yield spread around 1980. Volatility increased to 0.015 which is much higher than its average value of 0.0028. This change is related to the change in the Fed policy in this period when Volcker was appointed as chairman of the Federal Reserve in August 1979. The previous figure shows that ignoring time-variation in volatility leads to a large underestimation of the average persistence of the yield spread. Secondly, the correlations between on one hand the error in the yield spread and on the other hand the errors in the stock and bond return equations vary a lot over time. The correlation for stocks is in general negative, but there are periods in which this correlation is positive. This implies that there is mean-reversion in stocks on average, but that there are periods in which stocks show mean-aversion. The correlation for bonds is negative, but the strength of this correlation changes a lot over time. Long-term bonds strongly mean-revert in the 1930s and 1940s, but hardly mean-revert in for example the 1960s and 1970s. Thirdly, the error volatility of stock and bond returns and the correlation between these errors is similar to figure 5.2. A change in predictor does not affect these parameters. This is not surprising given the low R^2 values of these regressions. Finally, the excess kurtosis in the error terms is considerably reduced if time-variation in volatility is incorporated. The posterior means of τ_1^2, τ_2^2 and τ_3^2 are only 0.10, 0.00 and 0.44. Apparently, if one does not take the large change in volatility around 1980 into account, one would wrongly conclude that the yield spread contains extremely high excess kurtosis as in table 5.2. In other words, once the time-varying error covariance matrix is taken into account, there is hardly any excess kurtosis left.

We conclude that it is very important to take the time-variation in the error covariance matrix into account. Ignoring the variation leads to a large underestimation of the persistence of the yield-spread and a large overestimation of the excess kurtosis of the error terms. The time-variation in intercepts and slopes is however not persistent enough for long-term investors to consider. This confirms results of previous sections.

An alternative robustness check would have been to analyze whether the parameters governing the transition equations (5.10) are themselves time-varying. We could analyze this issue by extending the most general specification to allow for time-variation in these parameters and then test whether these parameters can be pooled over time. We do not pursue this alternative here because of several reasons. Firstly, we expect that such "time-varying time-variation" (if present at all) only has a second-order impact on portfolios. It seems unlikely that time-variation in the time-variation of intercepts and

slopes is important given the results above. It could be important for the error covariance matrix, but it seems unlikely that it is as important as the direct time-variation in the error covariance matrix. Secondly, we would need to set-up transition equations for the intercepts, slopes and volatilities of the parameters in equations (5.10). This would increase the number of parameters considerably and would considerably increase the risk of overfitting. Finally, it would not solve the issue. One could wonder whether the parameters that govern the additional transition equations in the extended model are again time-varying etcetera.

5.6   Conclusion

This chapter analyzes the effects of time-variation in model parameters on long-term investors. Our most general specification allows for time-varying intercepts and slope coefficients, time-varying error volatility and time-varying error correlation. It also allows for non-normal error distributions.

We find that the persistence of time-varying parameters plays a decisive role in the importance of different time-varying components. The time-variation in intercepts and slopes does not turn out to be persistent enough and is therefore not relevant for long-term investors and hardly has an impact on portfolio allocations. Time-variation in error correlations and especially error volatility is however very persistent and very relevant for long-term investors due to the large impact on asset allocations. The normality assumption for the error term is valid as long as the error covariance matrix is allowed to vary over time. In case one ignores this time-variation, one would wrongly include that fat tails are important. Our preferred specification includes time-variation of the error covariance matrix, but ignores time-varying intercepts / slopes and ignores fat tails.

Our analysis can be extended in several directions. Firstly, we could consider a larger model that contains multiple predictor variables and multiple asset returns. Such a model contains a large number of parameters and its estimation therefore either requires restrictions or tight prior distributions. A second extension is to calculate fully dynamic strategies instead of straightforward buy-and-hold strategies. A full dynamic strategy includes learning about all parameters in the model. However, such a specification cannot be solved using the current state-of-the-art numerical techniques,

because of the large number of state variables. An interesting alternative would be to ignore learning about hyperparameters, but to explicitly consider learning about the time-varying parameters. This limits the number of state variables considerable and might be feasible using techniques such as Brandt, Goyal, Santa-Clara, and Stroud (2005). Thirdly, since volatility affects the asset allocations of investors, it would be interesting to analyze whether it also affects expected returns directly by using volatility as a (latent) predictor of stock returns. This would however considerably complicate the MCMC algorithm used in this chapter.

5.7 Appendix: Posterior distribution and MCMC algorithm

In this appendix, we write down posterior distributions and explain the MCMC algorithm we use to estimate the model.

Firstly, we consider how to draw time-varying parameters b_t, their unconditional means μ_b, their transition matrix A_b and the covariance matrix of its innovations Q. In order to do so, we reparameterize our model slightly

$$
\begin{aligned}
y_t &= X_t \mu_b + X_t \widetilde{b}_t + u_t \qquad\qquad (5.38)\\
\widetilde{\ln \sigma^2}_{t+1} &= A_s \widetilde{\ln \sigma^2}_t + \xi_t\\
\mu_b &= \mu_b\\
\widetilde{b}_{t+1} &= A_b \widetilde{b}_t + \eta_t.
\end{aligned}
$$

The properties of the innovations are explained in section 5.3.1 and the initial conditions are specified in equations (5.14) and (5.22). The equations form a linear Gaussian state space model where the first two equations are the observation equations and the last two equations are the transition equations. We condition on (i) A_b, Q (transition equation), (ii) λ_t, l_t, σ_t $\forall t$ (the covariance matrix of u_t) and (iii) A_s, S, μ_s and ρ (remaining terms second observation equation and correlation first and second observation equation). Therefore, we can use the standard Kalman filter - smoother technique to draw μ_b and $\widetilde{b}_{t+1}$ $\forall$ t from $p(\mu_b, \widetilde{b}|Y, A_b, A_s, Q, \text{S}, \lambda, l, \sigma, \rho)$. We use the Kalman filter - smoother technique explained in Durbin and Koopman (2002).

5. MODEL INSTABILITY AND LONG-TERM INVESTORS

We include the second equation because ξ_t is correlated with u_t due to the leverage effect and cannot be ignored when drawing $\widetilde{b}_t$ and μ_b conditional on (among others) σ_t^2. Therefore, we need to include it as an *observation* equation that is correlated to the first observation equation with correlation coefficient ρ. Note that in order to draw $\widetilde{\ln \sigma^2}_t$ itself we also need to consider this same equation as a *transition* equation below.

If the initial condition (5.14) would not depend in a non-linear way on A_b and Q, we could simply use an inverse Wishart distribution to draw $p(Q|Y, A_b, \widetilde{b})$ and a normal distribution to draw $p(A_b|Y, Q, \widetilde{b})$. Instead, we use a Metropolis-Hastings step where the inverse Wishart and normal distributions are used as proposal densities. For Q the proposal density is

$$\iota(Q^*) = iWishart\left(\sum_{t=1}^{T} \left(\eta_t \eta_t' \right) + W_Q, T + df_Q \right) \tag{5.39}$$

and the acceptance probability is

$$\alpha_b = \min\left\{ 1, \frac{|\Sigma_b^*|^{-\frac{1}{2}}}{|\Sigma_b|^{-\frac{1}{2}}} \exp\left(-\frac{1}{2}\widetilde{b}_1'\Sigma_b^*\widetilde{b}_1 + \frac{1}{2}\widetilde{b}_1'\Sigma_b\widetilde{b}_1 \right) \right\}, \tag{5.40}$$

where Σ_b^* and Σ_b are the unconditional covariance matrices - see the explanation below equation (5.14) - based on respectively the newly proposed draw Q^* and the draw from the previous iteration.

Define $\widetilde{B}$ as the $T \times nK$ matrix whose t^{th} row is equal to $\widetilde{b}_{t+1}'$ and define $\widetilde{B_{-1}}$ as the $T \times nK$ matrix whose t^{th} row is equal to $\widetilde{b}_t'$. Let a_b be the diagonal of A_b. Its proposal density is

$$\iota(a_b^*) = N(\hat{a}_b, M_b^{*-1}), \tag{5.41}$$

where

$$
\begin{aligned}
M_b^* &= X_{\widetilde{b}}' \left(Q^{-1} \otimes I_T \right) X_{\widetilde{b}} + v_b^{-1} I_{nK} \\
\hat{a}_b &= M_b^{-1} \left(X_{\widetilde{b}}'(Q^{-1} \otimes I_T)y_{\widetilde{b}} + \iota_{nK} v_b^{-1} m_b \right),
\end{aligned} \tag{5.42}
$$

where $y_{\widetilde{b}}$ vertically stacks the columns of $\widetilde{B}$, where

$$X_{\widetilde{b}} = \begin{pmatrix} \widetilde{b}_{1,-1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \widetilde{b}_{nK,-1} \end{pmatrix} \tag{5.43}$$

and where $\widetilde{b}_{i,-1}$ is the ith column of $\widetilde{B}_{-1}$. This is similar to a GLS regression. The acceptance probability is again α_b with Σ_b^* depending on the newly proposed draw A_b^*.

Secondly, we look at drawing time-varying parameters l_t, their unconditional means μ_l, transition matrix A_l and the covariance matrix of the innovations R. The system of equations can be rewritten as

$$L_t(y_t - X_t b_t) \equiv L_t \hat{y}_t = \Lambda_t \Sigma_t e_t. \tag{5.44}$$

We condition on b_t, Σ_t and Λ_t $\forall$ t and can therefore treat them as given in this step. Since matrix L_t is a lower triangular matrix with ones on the diagonal, we can rewrite the previous equation as

$$\hat{y}_t = T_t l_t + \Lambda_t \Sigma_t e_t, \tag{5.45}$$

where

$$T_t = \begin{pmatrix} 0 & \dots & \dots & 0 \\ -\hat{y}_{1,t} & 0 & \dots & 0 \\ 0 & -\hat{y}_{[1:2],t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -\hat{y}_{[1:n-1],t} \end{pmatrix}. \tag{5.46}$$

Since the dependent variable of the observation equation - $\hat{y}_t$ - also occurs on the right-hand-side (RHS) in T_t, this system has a nonlinear Gaussian state space representation. However, since (i) we impose that R is blockdiagonal (ii) the system has a triangular structure and (iii) the dependent variable in one equation does not show up on the RHS of the same equation, we can apply the Kalman filter - smoother technique equation by equation. Hence, for equation j with $j = 2, ..., n$, we consider the following linear Gaussian state space model

$$
\begin{aligned}
\hat{y}_{t,j} &= -\hat{y}_{t,[1:j-1]}\mu_{l,\{j\}} - \hat{y}_{t,[1:j-1]}l_{t,\{j\}} + \lambda_{t,j}^{\frac{1}{2}}\sigma_{j,t}e_{t,j} \\
\widetilde{\ln \sigma^2}_{j,t+1} &= A_{l,(j,j)}\widetilde{\ln \sigma^2}_{j,t} + \xi_{t,j} \\
\mu_{l,\{j\}} &= \mu_{l,\{j\}} \\
\widetilde{l}_{t+1,\{j\}} &= A_{l,\{j\}}\widetilde{l}_{t,\{j\}} + \zeta_t,
\end{aligned}
\tag{5.47}
$$

where $\{j\}$ refers to the elements of the vectors/matrices that belong to the jth equation. The properties of the error terms are explained in section 5.3.1 and the initial conditions are given in equations (5.15) and (5.23). The first two equations are the

observations equations and the last two are the transition equations. As above, the second equation is included, because the correlation between $e_{t,j}$ and $\xi_{t,j}$ is equal to ρ_j. We can use the Kalman filter - smoother technique to draw $\mu_{l,\{j\}}$ and $\widetilde{l}_{t,\{j\}}$ $\forall$ t from $p(\mu_{l,\{j\}}, \widetilde{l_{\{j\}}} | Y, A_{s,(j,j)}, S_{j,j}, \sigma_j, \lambda_j, A_{l,\{j\}}, R_{\{j\}})$.

We cannot simply use Gibbs steps to draw $R_{\{j\}}$ and $A_{l,\{j\}}$ for $j = 2, ..., n$, because the unconditional variance $\Sigma_{l,\{j\}}$ depends in a non-linear way on these parameters. We use Metropolis-Hastings steps instead where we use inverse Wishart and normal distributions as proposal densities for $p(R_{2:3,2:3} | Y, A_{l,\{3\}}, \widetilde{l}_{\{3\}})$, $p(A_{l,\{2\}} | Y, R_{1,1}, \widetilde{l}_{\{2\}})$ and $p(A_{l,\{3\}} | Y, R_{2:3,2:3}, \widetilde{l}_{\{3\}})$. Since these steps are almost similar to drawing A_b and Q above, we do not explicitly write them down. We do not need a Metropolis-Hastings step when drawing $R_{1,1}$ for $j = 2$. In this case, we can use a Gibbs step by drawing from the inverse gamma distribution

$$p(R_{1,1} | Y, A_{l,\{1,1\}}, \widetilde{l}_{\{2\}}) = iGamma\left([(1 - A_{l,\{1,1\}}^2)\widetilde{l}_{1,1}^2] + \sum_{t=1}^{T}(\zeta_t\zeta_t') + W_{R1}, T + 1 + df_{R1}\right).$$
(5.48)

Thirdly, we explain how to simulate the time-varying parameters $\ln \sigma_t^2$ and λ_t $\forall t$, the unconditional mean μ_s, transition matrix A_s, covariance matrix S and correlation coefficient ρ. We rewrite our model slightly

$$L_t(y_t - X_t b_t) \equiv y_t^* = \Lambda_t \Sigma_t e_t.$$
(5.49)

We condition on l_t and b_t and hence treat them as given. Note that we can treat the different equations separately, since (i) the elements of e_t are independent of each other and (ii) we impose a diagonal structure for covariance matrix S.

The observation equations are non-linear in the diagonal elements of Λ_t and Σ_t which means that we cannot use the linear Gaussian state space model without any further modification. Following Omori, Chib, Shephard, and Nakajima (2007) we consider the following transformation of the dependent variable instead

$$y_{t,j}^{**} = \ln y_{t,j}^{*2}, \forall t, j.$$
(5.50)

For $j = 1, ..., n$ this gives the following state space model

$$
\begin{aligned}
y_{t,j}^{**} &= \mu_{s,j} + \widetilde{\ln \sigma^2}_{t,j} + \ln(\lambda_{t,j}) + \ln(e_{t,j}^2) \quad\quad (5.51) \\
\mu_{s,j} &= \mu_{s,j} \\
\widetilde{\ln \sigma^2}_{t+1,j} &= A_{s,(j,j)} \widetilde{\ln \sigma^2}_{t,j} + \xi_{t,j} \\
\ln(\lambda_{t+1,j}) &= -(1/2)\tau_j^2 + u_{\tau,j},
\end{aligned}
$$

where

$$
u_{\tau,j} \sim N(0, \tau_j^2) \quad\quad (5.52)
$$

is independent of the other error terms. The properties of the other innovations are given in section 5.3.1 and the initial conditions are specified in equations (5.16) and (5.24). The first equation is the observation equation and the remaining equations are the transition equations. The error in the observation equation - $\ln(e_{t,j}^2)$ - and the innovation in the transition equation - ξ_t - are dependent, since the correlation between $e_{t,j}$ and $\xi_{t,j}$ is ρ_j.

The state space system is linear, but non-Gaussian, since the error in the observation equation has a log χ^2 distribution. In order to be able to use linear Gaussian state space techniques, Omori, Chib, Shephard, and Nakajima (2007) propose to approximate the log chi-squared distribution using a mixture of 10 normal distributions. Their method is an extension of Kim, Shepherd, and Chib (1998) by allowing for the leverage effect, i.e. dependence between $\ln(e_{t,j}^2)$ and $\xi_{t,j}$. It allows us to draw from $p(\ln \sigma_j^2, \ln(\lambda_j)|Y, b, l, A_{s,(j,j)}, S_{j,j}, \tau_j^2, \rho_j)$. Please refer to Omori, Chib, Shephard, and Nakajima (2007) for more details.

In order to draw θ_j and ψ_j (which we transform to ρ_j and $S_{j,j}$), we use results in Jacquier, Polson, and Rossi (2004). Let $r_{t,j} = (e_{t,j}, \xi_{t,j})$ for equation j. The posterior distribution of Σ_j^* is proportional to

$$
p(\Sigma_j^*|Y, A_{s,(j,j)}, \ln \sigma_j^2) \propto p(\Sigma_j^*) p(\ln \sigma_{j,1}^2 | A_{s,(j,j)}, S_{j,j}) |\Sigma_j^*|^{-T/2} \exp\left(-\frac{1}{2} tr\left(\Sigma_j^{*-1} UU_j\right)\right),
\quad\quad (5.53)
$$

where $UU_j = \sum_t r_{t,j} r_{t,j}'$.

Define $a_{(k,l),j}$ as the $(k,l)^{th}$ element of UU_j. Furthermore, let $a_{22.1,j} = a_{(2,2),j} - a_{(1,2),j}^2/a_{(1,1),j}$ and let $\widehat{\psi}^j = a_{(1,2),j}/a_{(1,1),j}$. Ignoring the initial condition, it is easy to

show (see Jacquier, Polson, and Rossi (2004)) that

$$p(\psi_j|Y, \theta_j, b, \lambda_{t,j}, \widetilde{\ln \sigma^2}_j, A_{s,(j,j)}) \quad = \quad N\left(\widetilde{\psi_j}, \frac{\theta_j}{a_{(1,1),j} + p}\right) \tag{5.54}$$

$$p(\theta_j|Y, b, \lambda_{t,j}, \widetilde{\ln \sigma^2}_j, A_{s,(j,j)}) \quad = \quad iGamma(a_{22.1,j} + W_{\theta,j}, T + df_\theta), \tag{5.55}$$

where

$$\widetilde{\psi_j} = \frac{a_{(1,1),j}\widehat{\psi^j}}{a_{(1,1),j} + p} \tag{5.56}$$

These are the conditional distributions used in Jacquier, Polson, and Rossi (2004). These are not equal to the conditional posteriors in our setting due to the presence of the initial conditions. We use these two equations as proposal densities in a Metropolis-Hastings step with the following acceptance probability

$$\alpha_s = \min\left\{1, \frac{|\Sigma_s^*|^{-\frac{1}{2}}}{|\Sigma_s|^{-\frac{1}{2}}} \exp\left(-\frac{1}{2}\widetilde{\ln \sigma^2}_{1,j}' \Sigma_s^* \widetilde{\ln \sigma^2}_{1,j} + \frac{1}{2}\widetilde{\ln \sigma^2}_{1,j}' \Sigma_s \widetilde{\ln \sigma^2}_{1,j}\right)\right\}, \tag{5.57}$$

where $\Sigma_{s,(j,j)}^*$ depends on the newly drawn ψ_j^* and θ_j^*.

The step to draw $A_{s,(j,j)}$ for equations $j = 1, ..., n$ is a relatively straightforward Metropolis-Hastings step. We need to take the correlation between $\xi_{t,j}$ and $e_{t,j}$ into account. Therefore, we obtain the following auxiliary equation for $\widetilde{\ln \sigma^2}_{t+1,j}$

$$\widetilde{\ln \sigma^{2*}}_{t+1,j} \equiv \widetilde{\ln \sigma^2}_{t+1,j} - \Delta \times e_t = A_{s,(j,j)}\widetilde{\ln \sigma^2}_{t,j} + \xi_t^*, \tag{5.58}$$

where innovation $\xi_t^* \sim N(0, S_{j,j}(1 - \rho_j^2))$ (due to the conditioning) and where

$$\Delta = \rho_j\sqrt{S_{j,j}}. \tag{5.59}$$

It is straightforward to show that we can use the following proposal density for $A_{s,(j,j)}^*$

$$\iota(A_{s,(j,j)}^*) = N(\hat{a}_j, M_j^{-1}) \tag{5.60}$$

with

$$M_j \quad = \quad \left(S_{j,j}(1 - \rho_j^2)\right)^{-1}\left(\sum_{t=1}^{T}(\widetilde{\ln \sigma^2}_{t,j})^2\right) + v_s^{-1} \tag{5.61}$$

$$\hat{a}_j \quad = \quad M_j^{-1}\left(\left(S_{j,j}(1 - \rho_j^2)\right)^{-1}\sum_{t=1}^{T}(\widetilde{\ln \sigma^2}_{t,j}\widetilde{\ln \sigma^2}_{t+1,j}) + v_s^{-1}m_s\right) \tag{5.62}$$

and the same acceptance probability as in equation (5.57) where $\Sigma_{s,(j,j)}^*$ depends on the newly drawn value for $A_{s,(j,j)}^*$.

Finally, we look at drawing τ_j^2. The kernel of the conditional posterior distribution for τ_j^2 is

$$p(\tau_j^2|Y,\lambda_j) \propto (\tau_j^2)^{-\frac{T+1.5}{2}} \exp\left(-\frac{\tau_j^2}{2}\right) \prod_{t=1}^{T} \exp\left(\frac{1}{2\tau_j^2}(\ln\lambda_{t,j} + 0.5\tau_j^2)^2\right). \qquad (5.63)$$

This is a non-standard distribution, but we simulate from this distribution for $j = 1,....,n$ using a Metropolis-Hastings step. More precisely, we simulate $\ln\tau_j^2$ by finding the mode m_j of the above posterior distribution in every iteration and by subsequently using a student-t distribution as proposal density with mean equal to the log of the mode, variance equal to 1.1 times the negative inverse of the hessian matrix of the log kernel at the mode and degrees of freedom equal to 8. The acceptance probability is calculated in the usual way using both the proposal and the kernel (taking the Jacobian of the transformation into account), refer to for example Bauwens, Lubrano, and Richard (1999), page 89. Finally, we transform $\ln\tau_j^2$ to obtain τ_j^2.[1]

The acceptance probabilities for all Metropolis-Hastings steps are larger than 80% in all cases. In the empirical section, we retain 10,000 iterations after a burn-in period of 5,000 iterations. Increasing the number of iterations does not significantly impact results. We draw 10 path asset return paths per iteration, 100,000 paths in total, to calculate predictive distributions and portfolio weights.

---

[1]The acceptance probability using $\ln\tau_j^2$ is approximately 97%. If we would simulate τ_j^2 directly, the acceptance probability would decrease to 67%.

6

Summary and conclusion

The main objective of this thesis is to investigate whether the promises of the (long-term) strategic asset allocation literature hold in settings that include parameter uncertainty, model uncertainty and model instability. A side objective is to analyze whether we can sort thousands of stocks in robust myopic portfolios with high expected returns and low risk.

We answer the research questions as follows. Firstly, we document that a panel data model that combines firm characteristics with industry effects can explain a large part of the cross-section of stock returns. The constructed portfolios are not particularly risky, are stable over time (especially using long investment horizons) and lead to a (risk-adjusted) average return of up to 2% per month. Secondly, although a naive implementation of strategic portfolios fails out-of-sample, the use of a shrinkage prior improves performance considerably and allows long-term investors to time the market. The hedge component of dynamic strategies hardly adds value though. Thirdly, we develop refined Bayesian Model Averaging techniques to incorporate model uncertainty in systems of equations and find that its impact is large, especially at long horizons. Including model uncertainty makes stocks riskier in the long-run in crisis periods and leads to lower optimal allocations to stocks. Finally, we develop a time-varying parameter model including fat tailed error distributions and the leverage effect and find that only changes in the elements of the error covariance matrix are variable and persistent enough to be relevant for long-term investors. Changes in intercepts and slopes can be safely ignored.

6. SUMMARY AND CONCLUSION

We provide a further, more detailed, summary of the thesis in the following 13 statements. These statements also include the main lessons derived from this thesis.

1. Standard asset pricing models cannot explain differences in expected returns based on multiple firm characteristics and industry effects jointly.

The portfolios, generated using a panel data model with 11 firm characteristics and industry-specific effects, are a challenge for standard asset pricing models. The long-short portfolios have average returns of up to 2% per month, do not contain a very clear factor structure and risk-adjustment hardly has an impact either using the Fama-French Carhart model with/without time-varying factor loadings. New research in asset pricing is needed to explain the cross-sectional differences in returns across the industry dimension.

2. Considering longer investment horizons can be beneficial, even when using short-term performance measures.

Even when evaluating the performance of long-short portfolios by using short-term performance measures such as Sharpe ratios, the specifications that forecast stock returns for longer forecasting horizons lead to the best performing portfolios. These portfolios have high average (risk-adjusted) returns, are well-diversified and have only low turnover.

3. After controlling for many observed characteristics, there is still considerable heterogeneity left across firms.

Although the panel data model is able to explain a large part of the cross-sectional variation in stock returns, the presence and importance of (fixed) individual effects suggests that considerable heterogeneity across firms is left that cannot be explained by the firm characteristics. This suggests that new econometric estimation techniques should be developed that are able to efficiently extract information from individual effects, while still allowing for correlation between firm characteristics and individual effects.

4. With a bit of skepticism long-term investors can benefit from market timing.

The finding that a naive implementation of strategic asset allocations leads to very

bad performance does not mean that long-term investors should not time the market. Skeptical investors - reflected in the use of a skeptical shrinkage prior - can benefit considerably from market timing. Their skepticism allows them to benefit from changes in market conditions, but saves them from large declines in their portfolio value. Hence, we do not share the negative view of DeMiguel, Garlappi, and Uppal (2009) that data-based methods to calculate (strategic) asset allocations are not able to outperform the $1/N$ allocation out-of-sample.

5. *Investment strategies but also forecasts should not be evaluated by quadratic loss functions, but by asymmetric utility/loss functions.*

Investors do not necessarily prefer specifications that lead to the best forecasts in terms of mean squared errors. They are risk-averse, evaluate big losses and gains differently and therefore value specifications that are able to avoid big losses. This fact should be reflected in both the evaluation of investment strategies as well as the evaluation of forecasts. Standard practice in forecast evaluation is to use symmetric loss functions instead. A notable exception is Christoffersen and Diebold (1997).

6. *Although optimal long-term strategies are much more difficult to calculate than short-term strategies, they often give the same results.*

Optimal long-term dynamic strategies are much more difficult to calculate than myopic strategies, since the former involve dynamic programming, many state variables and thousands of simulated paths. Still, there is hardly a difference in their out-of-sample performance. The main reason is estimation error. Estimated myopic and estimated dynamic strategies approximate the true optimal dynamic strategy equally well. Another explanation can be found in the last two chapters. In case model uncertainty and model instability are incorporated in the decision process, optimal long-term equity allocations can be very similar to short-term allocations or occasionally even lower.

7. *Even though short horizons suggest that parameter and model uncertainty and model instability are irrelevant, these concepts are very important for investment decision making at longer horizons.*

The incorporation of parameter uncertainty, model uncertainty and to a lesser degree model instability hardly has an impact on short-term asset allocations. However, since

(i) an unfavorable parameter set / model persists for the full investment horizon and (ii) (unexpected) time-variation in (especially) error volatility is very persistent, these effects become important at longer horizons.

8. The (perceived) riskiness of stocks changes considerably over time and should be taken into account by long-term investors.

The (error) volatility of stock returns changes considerably and persistently over time. However, even in settings where volatility is constant, the volatility of the predictive distribution of stock returns still changes a lot over time due to model uncertainty. One of its major components - the volatility of the conditional mean of the different model forecasts - is low when all models deliver similar predictions such as in the 1960s or 1970s, but high in periods when the models yield diverse predictions such as in 1929 or 2008. Ignoring the increases in volatility leads to a substantial overinvestment in equity.

9. The incorporation of uncertainty is important in decision making, but not all uncertainty is relevant.

Parameter uncertainty, model uncertainty and the time-variation in the error volatilities are all very important at long horizons. However, at short horizons these effects (especially the first two) can be safely ignored. Furthermore, the low persistence of time-variation in intercepts and slopes implies that this time-variation can be ignored for all horizons.

10. The stock market is not a better place for long-term investors than for short-term investors

The incorporation of model uncertainty can make stocks substantially riskier in the long-run than in the short-run and can lead to lower equity allocations for long-term investors. This finding is related to a recent paper by Pastor and Stambaugh (2010). In their setting they allow for expected returns that are not perfectly correlated with a linear combination of a small set of predictors. They find that this fact leads to returns that have a much larger variance in the long-run. In contrast to them we find that stocks can either be riskier/safer in the long-run than in the short-run. Crucial is the fact that the importance of model uncertainty changes over time. The incorporation of model instability strengthens these conclusions.

11. Long-term investors want inflation-indexed bonds.

If long-term investors recognize that the true parameters and true model are unknown, the predictive distribution of stock returns can in the most extreme cases have high variance, negative skewness and (extremely) high kurtosis due to model uncertainty. In such a setting, long-term investors prefer the real riskfree long-term asset - the inflation-indexed bond - over the optimal combination of T-bills, stock and nominal bonds, since the certainty equivalent of this optimal combination is lower than 1.

12. The disagreement about stock return predictors can for a large part be attributed to the forecast horizon.

Since there is only minor time-variation - especially compared to the amount of noise - in slope coefficients on the dividend-to-price ratio and the yield spread in the stock return equation, we conjecture that the sample period is not the most important driver of the disagreement in academia about viable stock return predictors. The forecast horizon most likely plays an even bigger role. While at short horizons only predictors of stock returns are important, predictors of predictors become important as well at longer forecast horizons.

13. The incorporation of relatively new asset classes in the investment menu is not attractive for long-term investors given the limited information we have about them.

Since we only have a short data-set available to estimate models for new asset classes (e.g. commodities), parameter uncertainty is expected to be huge for such models. Because parameter uncertainty increases predictive volatility considerably in the long-run, long-term investors will hardly invest in new asset classes unless they provide really large benefits compared to more traditional asset classes such as stocks and bonds. New asset classes only become attractive for long-term investors if sufficient data is available.

Finally, we consider a brief outlook for future research. Even with more than 80 years of data, uncertainty still plays a major role for long-term investors. The promising finding in the early 2000s that stocks mean-revert and are therefore safer in the long-run is more than offset by the parameter and model uncertainty effect. This implies that without any extra information about future stock returns, long-term investors cannot

benefit from mean-reversion and therefore should not invest more in the stock market than short-term investors. Therefore, we need other ways to impose structure/add prior information to econometric models and reduce parameter and/or model uncertainty. A promising avenue of research is to use economic theory to impose restrictions on models that are (approximately) true and in this way improve forecasts and reduce forecast uncertainty. Present value models seem a promising line of research.

References

Andersson, Michael, and Sune Karlsson, 2008, Bayesian forecast combination for var models, Chapter in *Bayesian Econometrics*, page 501-524. 118

Ang, Andrew, and Dennis Kristensen, 2009, Testing conditional factor models, *Working paper*. 17

Ang, Andrew, Jun Liu, and Krista Schwarz, 2010, Using stocks or portfolios in tests of factor models., *Working paper*. 10, 16

Avramov, Doran, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458. 4, 58, 93, 103, 104, 105, 109, 112, 117

Avramov, D., and T. Chordia, 2006, Asset Pricing Models and Financial Market Anomalies, *Review of Financial Studies* 19, 1001–1040. 11, 16, 34

Bai, J., 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* 71, 135–171. 23

———, and S. Ng, 2002, Determining the Number of Factors in Approximate Factor Models, *Econometrica* 70, 191–221. 23

Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *The Journal of Finance* 55, 225–264. 2, 6, 59, 60, 99, 102, 105, 109, 110, 129, 164, 171, 186

Barsky, Robert, and Bradford De Long, 1993, Why does the stock market fluctuate?, *Quarterly Journal of Economics* 2, 291–311. 121

Bauer, Rob, Bart Diris, Boris Pavlov, and Peter Schotman, 2011, Firm characteristics, industry, horizon and time effects, in the cross-section of expected stock returns, *Working paper*. 9

Bauwens, Luc, Michel Lubrano, and Jean-François Richard, 1999, *Bayesian Inference in Dynamic Econometric Models* (Oxford University Press). 95, 117, 151, 157, 161, 209

Berger, James, and William Strawdermann, 1996, Choice of hierarchical priors: admissibility in estimation of normal means, *Annals of Statistics* 24, 931–951. 49

Black, Fischer, and Robert Littermann, 1992, Global portfolio optimization, *Financial Analyst Journal* 48, 28–43.

Bodie, Z., A. Kane, and A. Marcus, 2007, *Investments* (Irwin/McGraw-Hill) 7th ed.

Brandt, Michael, 2010, Portfolio choice problems, *Handbook of Financial Econometrics, Volume 1: Tools and Techniques* pp. 269–336. 102, 164

———, Amit Goyal, Pedro Santa-Clara, and Jonathan Stroud, 2005, A simulation approach to dynamic portfolio choice with an application to learning about return predictability, *The Review of Financial Studies* 18, 831–873. 2, 51, 59, 60, 67, 68, 78, 79, 92, 93, 203

Branger, Nicole, Beate Breuer, and Christian Schlag, 2010, Discrete-time implementation of continuous-time portfolio strategies, *European Journal of Finance* 16, 137–152. 56

Brennan, M.J., T. Chordia, and A. Subrahmanyam, 1998, Alternative Factor Specifications, Security Characteristics and the Cross-Section of Expected Stock Returns, *Journal of Financial Economics* 49, 345–373. 10, 11, 16, 25, 30

Brennan, M.J., and A. Subrahmanyam, 1995, Investment Analysis and Price Formation in Security Markets, *Journal of Financial Economics* 38, 361–382. 25

Breusch, Trevor, and Adrian Pagan, 1979, A simple test for heteroscedasticity and random coefficient variation, *Econometrica* 47, 1287–1294. 184

Campbell, John, 1987, Stock return and the term structure, *Journal of Financial Economics* 18, 373–399. 52

———, 1995, Some lessons from the yield curve, *Journal of Economic Perspectives* 9, 129–152. 53, 167

———, Yeung Chan, and Luis Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of Financial Economics* 67, 41–80. 2, 53, 55, 56, 98, 102, 105, 109, 111, 129, 164, 171, 183

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press). 10

Campbell, John, and Robert Shiller, 1998, Valuation ratios and the long-run stock market outlook, *Journal of Portfolio Management* 24, 11–26. 52, 53, 167

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. 102

Campbell, John, and Luis Viceira, 2001, Who should buy long-term bonds, *American Economic Review* 91, 99–127. 71

———, 2002, *Strategic Asset Allocation* (Oxford University Press). 1, 59, 64, 67, 102, 111, 164, 186

———, 2005, The term structure of the risk-return trade-off, *Financial Analysts Journal* 61, 34–44. 129

Campbell, J.Y., and T. Vuolteenaho, 2004, Bad Beta, Good Beta, *American Economic Review* 94, 1249–1275. 10

Chacko, George, and Luis Viceira, 2005, Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets, *The Review of Financial Studies* 18, 1369–1402. 184

REFERENCES

Chan, L.K.C., N. Jegadeesh, and J. Lakonishok, 1996, Momentum Strategies, *Journal of Finance* 51, 1681–1713. 25

Christoffersen, Peter, and Francis Diebold, 1997, Optimal prediction under asymmetric loss, *Econometric Theory* 13, 808–817. 213

Christoffersen, Peter, Kris Jacobs, and Gregory Vainberg, 2008, Forward-looking betas, *Working paper*. 17

Cochrane, J.H., 2005, *Asset Pricing* (Princeton University Press). 25

Cochrane, John, 2007a, Portfolio theory, *Working paper*.

Cochrane, J.H., 2007b, The Dog That Did Not Bark: A Defense of Return Predictability, *Review of Financial Studies* forthcoming. 167

Conrad, J., and G. Kaul, 1998, An Anatomy of Trading Strategies, *Review of Financial Studies* 11, 489–520. 39

Cremers, Martijn, 2002, Stock return predictability: A bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249. 58, 93, 103, 109

Daniel, K., and S. Titman, 1997, Evidence on the Characteristics of Cross Sectional Variation in Stock Returns, *Journal of Finance* 52, 1–33. 10, 25

Davis, J.L., E.F. Fama, and K.R. French, 2000, Characteristics, Covariances and Average Returns: 1929 to 1997, *Journal of Finance* 55, 389–406. 10

De Bondt, W.F.M., and R.J. Thaler, 1985, Does the Stock Market Overreact?, *Journal of Finance* 40, 793–805. 10

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How efficient is the $1/n$ portfolio strategy, *The Review of Financial Studies* 22, 1915–1953. 3, 71, 213

Diebold, Francis, and Roberto Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13, 253–63. 81

Diris, Bart, 2011a, Model instability and long-term investors, *Working paper*. 163

———, 2011b, Model uncertainty for long-term investors, *Working paper*. 101

———, Franz Palm, and Peter Schotman, 2011, Long-term strategic asset allocation: an out-of-sample evaluation, *Working paper*. 47

Driscoll, J.C., and A.C. Kraay, 1998, Consistent Matrix Estimation with Spatially Dependent Panel Data, *Review of Economics and Statistics* 80, 549–580. 19, 21

Durbin, James, and Siem-Jan Koopman, 2002, A simple and efficient simulations smoother for state space time series analysis, *Biometrika* 89, 603–615. 173, 178, 179, 203

Fama, E.F., 1998, Market Efficiency, Long-Term Returns, and Behavioral Finance, *Journal of Financial Economics* 49, 283–306. 34

Fama, Eugene, and Kenneth French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49. 53, 167

Fama, E.F., and K.R. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427–465. 24, 25

———, 1995, Size And Book-to-Market Factors in Earnings and Returns, *Journal of Finance* 50, 131–156. 10

———, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55–84. 10

———, 1997, Industry Costs of Equity, *Journal of Financial Economics* 43, 153–193. 10

Fama, Eugene, and Kenneth French, 2008, Dissecting anomalies, *Journal of Finance* 63, 1653–1678. 10, 26

Fama, Eugene, and G.William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146. 52

Fernandez, Carmen, Eduardo Ley, and Mark Steel, 2001, Benchmark priors for bayesian model averaging, *Journal of Econometrics* 100, 381–427. 106, 113

Fernandez-Villaverde, Jesus, and Juan Francisco Rubio-Ramirez, 2004, Comparing dynamic equilibrium models to data: a bayesian approach., *Journal of Econometrics* 123, 153–187. 109

Frankel, R., and C.M.C. Lee, 1998, Accounting Valuation, Market Expectation, and Cross-Sectional Stock Returns, *Journal of Accounting and Economics* 25, 283–319. 25

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. 109

George, Edward, Shawn Ni, and Dongchu Sun, 2008, Bayesian stochastic search for var model restrictions, *Journal of Econometrics* 142, 553–580. 117

Geweke, John, 2005, *Contemporary Bayesian Econometrics and Statistics* (Wiley). 95

Ghysels, Eric, and Eric Jacquier, 2007, Market beta dynamics and portfolio efficiency, *Working paper*. 17

Godsill, 2001, On the relationship between markov chain monte carlo methods for model uncertainty, *Journal of Computational and Graphical Statistics* 10, 230–248. 106, 116, 159, 160

Goncalves, Sylvia, 2010, The moving blocks bootstrap for panel linear regression models with individual fixed effects, *Forthcoming Econometric Theory*. 19

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. 2, 52, 102, 106, 107, 166

Green, Peter, 1995, Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika* 82, 711–732. 160

REFERENCES

Greenaway-McGrevy, Ryan, Chirok Han, and Donggyu Sul, 2010, Asymptotic distribution of factor augmented estimators for panel regression, *Forthcoming Journal of Econometrics.* 23, 24

Guidolin, Massimo, and Allan Timmermann, 2007, Asset allocation under multivariate regime switching, *Journal of Economic Dynamics and Control* 31, 3503–3544. 2, 4, 165

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. 159

Haugen, R.A., and N. Baker, 1996, Commonality in the Determinants of Expected Stock Returns, *Journal of Financial Economics* 41, 401–439. 10, 11

Hjort, Nils, and Gerda Claeskens, 2003, Frequentist model average estimators, *Journal of the American Statistical Association* 98, 879–899. 109

Hodrick, R.J., and Z. Zhang, 2001, Evaluating the Specification Errors of Asset Pricing Models, *Journal of Financial Economics* 62, 327–376. 10

Hoeting, J.A., D. Madigan, A. Raftery, and C. Volinsky, 1999, Bayesian Model Averaging: A Tutorial, *Statistical Science* 14, 382–417. 109

Hoevenaars, Roy, Roderick Molenaar, Peter Schotman, and Tom Steenkamp, 2007, Strategic asset allocation for longterm investors: Parameter uncertainty and prior information, *Working paper.* 129, 137

Jacquier, Eric, Nicholas Polson, and Peter Rossi, 2004, Bayesian analysis of stochastic volatility models with fat tails and correlated errors, *Journal of Econometrics* 122, 185–212. 170, 173, 177, 207, 208

Jacquier, Eric, Sheridan Titman, and Atakan Yaçın, 2010, Predicting systematic risk: Implications from growth options, *Journal of Empirical Finance* 17, 991–1005. 17

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1684. 50, 71, 76

Jagannathan, R., and Z. Wang, 1996, The Conditional CAPM and the Cross Section of Expected Returns, *Journal of Finance* 51, 3–53. 34

Jegadeesh, N., and S. Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65–91. 25

———, 2001, Profitability of Momentum Strategies: An Evaluation of Alternative Explanations, *Journal of Finance* 56, 699–720.

———, 2002, Cross-Sectional and Time Series Determinants of Momentum Returns, *Review of Financial Studies* 15, 143–157. 39, 40

Johannes, Michael, Arthur Korteweg, and Nicholas Polson, 2011, Sequential learning, predictive regressions and optimal portfolio returns, *Working paper.* 165, 166

Jorion, Philippe, 1986, Bayes-stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21.

Jurek, Jakub, and Luis Viceira, 2010, Optimal value and growth tilts in long-horizon portfolios, *Forthcoming Review of Finance.* 53, 55, 56, 59, 67, 68, 84, 85, 98

Kandel, Shmuel, and Robert Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *Journal of Finance* 51, 385. 113

Kapetanios, George, and Hashem Pesaran, 2007, Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset return, Chapter in *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis* by Garry Phillips and Elias Tzavalis. 21, 23, 40

Kim, Sangjoon, Neil Shepherd, and Siddhartha Chib, 1998, Stochastic volatility: Likelihood inference and comparison with arch models, *Review of Economic Studies* 65, 361–393. 178, 207

Kleibergen, Frank, 2010, Reality checks for and of factor pricing, *Working paper.* 16, 17

Koijen, Ralph, Theo Nijman, and Bas Werker, 2010, When can life-cycle investors benefit from time-varying bond risk premia, *Review of Financial Studies* 23, 741–780. 67, 68, 96

Koski, J.L., and R. Michaely, 2000, Prices, Liquidity and the Information Content of Trades, *Review of Financial Studies* 13, 659–696. 25

Lakonishok, J., A. Shleifer, and R. Vishny, 1994, Contrarian Investment, Extrapolation and Risk, *Journal of Finance* 49, 1541–1578. 25

Lewellen, J., 2004, Predicting Returns with Financial Ratios, *Journal of Financial Economics* 74, 209–235. 11

Ley, Eduardo, and Mark Steel, 2007, Jointness in bayesian variable selection with applications to growth regression, *Journal of Macroeconomics* 29, 476–493. 125, 127

———, 2009, On the effect of prior assumptions in bayesian model averaging with applications to growth regression, *Journal of Applied Econometrics* Forthcoming. 112, 149, 150

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. 151

Litterman, Rober, 1986, Forecasting with bayesian vector autoregressions - five years of experience, *Journal of Business and Economic Statistics* 4, 25–38. 64

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. 14, 118

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

REFERENCES

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

———, 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Michaud, Richard, 1989, The markowitz optimization enigma: Is "optimized" optimal?, *Financial Analysts Journal* 45, 31–42. 3

Moskowitz, T.J., and M. Grinblatt, 1999, Do Industries Explain Momentum?, *Journal of Finance* 54, 1249–1290. 11, 18

Newey, Whitney, and Kenneth West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–08. 81

———, 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, 631–53. 81

Ni, Shawn, and Dongchu Sun, 2003, Noninformative priors and frequentist risks of bayesian estimators of vector-autoregressive models, *Journal of Econometrics* 115, 159–197. 49, 62, 63, 95

Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima, 2007, Stochastic volatility with leverage: Fast and efficient likelihood inference, *Journal of Econometrics* 140, 425–449. 170, 172, 173, 178, 206, 207

Pastor, Lubos, and Robert Stambaugh, 2000, Comparing asset pricing models: an investment perspective, *Journal of Financial Economics* 56, 335–381.

———, 2010, Are stocks really less volatile in the long run?, *Working paper*. 104, 130, 136, 214

Pesaran, Hashem, 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica* 74, 967–1012. 21, 22, 23, 24

———, Ron Smith, and Kyung So Im, 1996, Dynamic linear models for heterogeneous panels, Chapter in *The Econometrics of Panel Data* by Laszlo Mattyas and Patrick Sevestre. 20

Pesaran, Hashem, and Takashi Yamagata, 2008, Testing slope homogeneity in large panels, *Journal of Econometrics* 142, 50–93. 20

Pettenuzzo, Davide, and Allan Timmermann, 2010, Predictability of stock returns and asset allocation under structural breaks, *Forthcoming Journal of Econometrics*. 2, 4, 165

Phillips, Peter, and Donggyu Sul, 2007, Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence, *Journal of Econometrics* 137, 162–188. 19

Primiceri, Giorgio, 2005, Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* 72, 821–853. 170, 173, 175, 176, 178, 194

Rosenberg, B., K. Reid, and R. Lanstein, 1985, Persuasive Evidence of Market Inefficiency, *Journal of Portfolio Management* 11, 9–17. 25

Rouwenhorst, K.G., 1998, International Momentum Strategies, *Journal of Finance* 53, 267–284. 25, 32

Sims, Christopher, and Harald Uhlig, 1991, Understanding unit rooters: A helicopter view, *Econometrica* 59, 1591–1599. 64, 108

Stambaugh, Robert, 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421. 64

Stock, James, and Mark Watson, 2002, Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* 97, 1167–1179. 23

Stoll, H., 1978, The Pricing of Dealer Services: An Empirical Study of NASDAQ Stocks, *Journal of Finance* 33, 1152–1173. 25

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. 159

van Binsbergen, Jules, and Michael Brandt, 2007, Solving dynamic portfolio choice problems by recursing on optimized portfolio weights or on the value function, *Computational Economics* 29, 355–367. 67, 68, 99, 100

Vogelsang, Timothy, 2008, Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects, *Working paper*. 19

Vuolteenaho, T., 2002, What Drives Firm Level Stock Returns?, *Journal of Finance* 57, 233–267. 25

Wachter, Jessica, and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market?, *Journal of Econometrics* 148, 162–178. 50, 58, 113

Wansbeek, T.J., and A. Kapteyn, 1989, Estimation of the Error Components Model with Incomplete Panels, *Journal of Econometrics* 41, 341–361. 19

Wright, Jonathan, 2008, Bayesian model averaging and exchange rate forecasts, *Journal of Econometrics* 146, 329–341. 58, 104, 118, 119

Xia, Y., 2001, Learning about Predictability: the Effect of Parameter Uncertainty on Dynamic Asset Allocation, *Journal of Finance* 56, 205–246. 2

Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (Wiley). 94

Nederlandse samenvatting

Deze sectie geeft de Nederlandse samenvatting van de dissertatie getiteld: "Strategische beleggingskeuze: het effect van onzekerheid op portefeuillekeuze".

Personen en instituten (zoals pensioenfondsen) beleggen hun vermogen in financiële producten om een lange-termijn doel te bereiken. Personen sparen bijvoorbeeld zelf voor hun pensioen of de opleiding van hun kinderen. Pensioenfondsen beleggen om hun pensioendeelnemers een pensioen uit te kunnen keren. Deze fondsen willen een veilig pensioen garanderen tegen de laagst mogelijke kosten (contributies). Aan de hand van een set preferenties kiezen beleggers hoe ze hun vermogen verdelen over beleggingsklassen zoals aandelen, obligaties, korte-termijn deposito's, vastgoed, handelsgoederen en hedge fondsen. "Strategic asset allocation" (strategische beleggingskeuze) is de keuze hoe beleggers dienen te investeren in deze brede klassen om hun lange-termijn doel te bereiken. De beleggingen variëren over de tijd door veranderingen in beleggingskansen, de investeringshorizon en lange-termijn macro-economische risicofactoren zoals inflatie en de rente.

Ruim 40 jaar geleden liet Merton (1969, 1971) al zien dat lange-termijn beleggers hedge-portefeuilles dienen aan te houden die toekomstige veranderingen in beleggingskansen anticiperen. Aan het eind van de jaren '90, begin 2000 was er een enorme opkomst in de populariteit van de strategische beleggingskeuze literatuur door de vondst dat aandelenrendementen mogelijk voorspelbaar zijn (bijvoorbeeld door de dividend-prijs ratio en rentes) en door de grote vooruitgang in computerkracht.

Campbell en Viceira (2002) geven een overzicht van de stand van de literatuur rond 2000. Het feit dat de dividend-prijs ratio mogelijk aandelenrendementen voorspelt leidt tot zogenaamde "mean-reversion" (terugkering naar het gemiddelde) in rendementen. Dit betekent dat een lager dan verwacht rendement wordt gevolgd door een hoger

dan verwacht toekomstig rendement. Deze negatieve autocorrelatie in aandelenrendementen maakt aandelen veiliger (kijkend naar de variantie) en dus aantrekkelijker op de lange termijn. De mean-reversion is ook een belangrijke component van de hedge portefeuille van lange-termijn beleggers. Dergelijke beleggers willen investeren in een financieel product waarvan het rendement hoog is wanneer verwachte toekomstige aandelenrendementenlaag zijn. Aandelen zelf blijken zo'n soort product te zijn. Een andere belangrijke overweging voor lange termijn beleggers is het inflatie- en renterisico. De risico-vrije investering voor een lange termijn belegger is een lange-termijn inflatie geïndexeerde staatsobligatie. Aangezien T-bills meerdere malen moeten worden overgesloten, zijn deze niet de risico-vrije investering. Ook het reële rendement van nominale lange-termijn obligaties is niet veilig, aangezien deze gevoelig zijn voor inflatierisico wat onaantrekkelijk is op de lange termijn. Lange-termijn beleggers moeten deze risico's meenemen in hun hedge portefeuille.

De bovenstaande resultaten zijn verkregen met behulp van eenvoudige modellen waarin de ware parameters en het ware model bekend worden verondersteld. Elke vorm van modelmisspecificatie heeft echter een grote invloed op de samenstelling van de berekende portefeuilles. In de jaren 2000 is daarom een nieuwe tak van de literatuur naar voren gekomen die de kwaliteit van de modellen en de gevoeligheid van de resultaten voor veranderingen in parameters analyseert. Barberis (2000) neemt parameter onzekerheid mee in het beslisproces van beleggers. Hij vindt dat aandelen nog altijd aantrekkelijker zijn op de lange termijn dan op de korte termijn, hoewel het verschil in beleggingen tussen lange en korte termijn verkleind wordt door het erkennen van parameter onzekerheid. Xia (2001) en Brandt, Goyal, Santa-Clara en Stroud (2005) beschouwen het leren over de voorspelbaarheid van aandelen en vinden dat het negeren van leren mogelijk leidt tot portefeuilles met een lager nut voor de belegger. Guidolin en Timmermann (2007) en Pattenuzo en Timmermann (2010) bekijken het effect van model instabiliteit door te kijken naar zogenaamde "regime-switching" modellen en concluderen dat het negeren van structurele breuken ook tot substantiële nutskosten kan leiden.

Er zijn verschillende redenen om sceptisch te zijn over de voorspellingen van de strategische beleggingskeuze literatuur. Ten eerste, de optimale portefeuilles blijken extreem, onrealistisch en erg gevoelig te zijn voor veranderingen in voorspelvariabelen. In Campbell, Chan en Viceira (2003) varieren de aandelengewichten bijvoorbeeld

tussen de - 1000% en 1000% en jaarlijkse veranderingen van 500% zijn niet ongewoon. Ten tweede, Goyal en Welch (2008) documenteren de slechte "out-of-sample" voorspelbaarheid van aandelenrendementen door te laten zien dat het historisch gemiddelde van rendementen even goed voorspelt als alle bekende voorspelvariabelen. Deze vondst leidt tot twijfels over de "mean-reversion" van aandelenrendementen en de tijdsvariatie van optimale portefeuilles. Ten derde, strategische beleggingskeuze is nog gecompliceerder dan myopische beleggingskeuze door de aanwezigheid van de hedge component. Terwijl de myopische portefeuilles slechts aangetast worden door schattingsfouten in de myopische component, worden strategische portefeuilles aangetast door fouten in zowel de myopische als de hedge component.

Een rijke literatuur documenteert de slechte prestaties van myopische portefeuilles. De reden waarom korte termijn portefeuilles zo slecht presteren is zogenaamde "error maximization" (fout maximalisatie). De "input" van portefeuille optimalisatie technieken (gemiddeldes, varianties etc.) worden geschat met fouten en portefeuille optimalisatie technieken geven te veel (te weinig) gewicht aan financiële producten met hoge (lage) rendementen. Deze producten hebben waarschijnlijk de grootste schattingsfouten zoals beweerd in Michaud (1989). DeMiguel, Garlappi en Uppal (2009) analyseren 14 verschillende modellen voor het uitrekenen van de inputs van mean-variance optimisatie technieken en laten zien dat geen enkele van deze 14 consistent beter presteert dan een simpele 1/N regel (zelfde gewicht voor alle producten). Aangezien strategische portefeuilles gevoeliger voor fouten zijn, zullen zij waarschijnlijk zelfs slechter presteren.

De geciteerde literatuur suggereert dat het onduidelijk is of de potentiële winsten van strategische portefeuille keuze kunnen worden gerealiseerd in de praktijk. Enerzijds suggereren de slechte out-of-sample resultaten van eenvoudige myopische portefeuilles dat dit niet het geval is, maar anderzijds kunnen recente inzichten zoals het meenemen van parameter onzekerheid in beslissingen de prestaties misschien aanzienlijk verbeteren. Dit overzicht suggereert ook dat het vormen van myopische portefeuilles bestaande uit vele individuele aandelen leidt tot zeer instabiele portefeuilles.

Het hoofddoel van deze dissertatie is te onderzoeken of de beloftes van de strategische portefeuille literatuur waar te maken zijn in realistische set-ups waarin mogelijke parameter onzekerheid, model onzekerheid en model instabiliteit worden meegenomen. Een tweede doel is te onderzoeken of we robuuste myopische portefeuilles kunnen vormen (d.w.z. portefeuilles die stabiel zijn door de tijd) bestaande uit bijna 2,000 indi-

viduele aandelen die hoge verwachte rendementen met minimaal risico kunnen gene-
reren. Deze dissertatie beantwoordt de volgende onderzoeksvragen:

- Hoe kunnen we een panel data model opzetten om individuele aandelenrende-
  menten met behulp van meerdere bedrijfskarakteristieken uit te leggen en kunnen
  we dit model gebruiken om robuuste portefeuilles te construeren?

- Kunnen de mogelijke winsten van de strategische beleggingskeuze literatuur wor-
  den gerealiseerd in een out-of-sample test en hoe kunnen we de prestaties ver-
  beteren?

- Hoe kunnen we een methodologie ontwikkelen om model onzekerheid met be-
  trekking tot lange-termijn voorspellingen mee te nemen en wat is de impact van
  model onzekerheid op lange-termijn beleggers?

- Hoe kunnen we een model met tijdsvariërende parameters opzetten en schatten
  en wat voor een soort tijdsvariatie moet een lange-termijn belegger meenemen in
  zijn analyse?

Als eerste analyseren we hoe we een model kunnen opzetten om de cross-sectie van
individuele aandelenrendementen te modelleren gebruikmakend van meerdere bedrijfs-
karakteristieken. De standaard sorteermethode werkt niet in zo'n set-up en daarom
ontwikkelen we een panel data model dat in staat is meerdere karakteristieken tegelijk
mee te nemen. Ten tweede onderzoeken we de prestaties van strategische portefeuilles
out-of-sample. Hoewel de out-of-sample prestaties zeer relevant zijn voor lange-termijn
beleggers, is een dergelijke out-of-sample test nog niet gedaan. Ten derde documenteren
we de impact van model onzekerheid op de verdeling van toekomstige aandelenrende-
menten en op de beslissingen van lange-termijn beleggers. Hoewel verschillende
modellen kunnen leiden tot compleet andere voorspellingen van toekomstige rende-
menten, wordt modelonzekerheid bijna altijd genegeerd in de strategische portefeuille
literatuur. Ten vierde gebruiken we een model met tijdsvariërende parameters om
de impact van model onzekerheid op de verdeling van toekomstige rendementen en
op de portefeuillekeuzes van lange-termijn beleggers te analyseren en om inzichten te
krijgen in het belang van de tijdsvariatie van de verschillende parameters. Andere ar-
tikelen - zoals Pettenuzo en Timmermann (2010) en Guidolin en Timmermann (2007)

- gebruiken "regime-switching" modellen en kunnen daarom geen onderscheid maken tussen de belangrijkheid van de verschillende parameters.

We beantwoorden de onderzoeksvragen als volgt. Ten eerste, we documenteren dat een panel data model dat bedrijfskarakteristieken met industrie-effecten combineert een groot gedeelte van de cross-sectie van aandelenrendementen kan verklaren. De gevormde portefeuilles zijn niet erg risicovol, zijn stabiel over de tijd (vooral wanneer gebruik wordt gemaakt van een lange beleggingshorizon) en leiden tot (risico-gecorrigeerde) gemiddelde rendementen van rond de 2% per maand. Ten tweede, hoewel een naïeve implementatie van strategische portefeuilles niet werkt out-of-sample, verbetert het gebruik van zogenaamde "shrinkage" prioren de prestaties aanzienlijk en staat dit lange-termijn beleggers toe om de markt te timen. De hedge component van de dynamische strategieën voegt echter amper waarde toe. Ten derde ontwikkelen we "Bayesian Model Averaging" (Bayesiaanse model middel technieken) om model onzekerheid mee te nemen in een systeem van vergelijkingen en vinden we dat de impact groot is, vooral voor een lange horizon. Het meenemen van model onzekerheid maakt aandelenrendementen risicovol op de lange termijn in crisis periodes en leidt tot een lagere optimale belegging in aandelen. Tenslotte ontwikkelen we een model met tijdsvariërende parameters dat zowel distributies met dikke staarten als het "leverage" (hefboom) effect aankan en vinden we dat alleen veranderingen in de covariantiematrix van de fouttermen variabel en persistent genoeg zijn om relevant te zijn voor lange-termijn beleggers. Veranderingen in constanten en hellingscoefficiënten mogen echter rustig genegeerd worden.

We geven een verdere, meer gedetailleerde samenvatting van de dissertatie in de volgende 13 stellingen. Deze stellingen bevatten ook de belangrijkste lessen van deze dissertatie.

1. Standaard asset pricing modellen kunnen de verschillen in verwachte rendementen - gebruik makend van zowel verscheidene bedrijfskarakteristieken als industrie-effecten - niet verklaren.

Portefeuilles, gevormd door gebruik te maken van een panel data model met 11 bedrijfskarakteristieken en industrie-specifieke effecten, zijn een uitdaging voor standaard asset pricing modellen. De portefeuilles hebben een gemiddeld rendement van 2% per maand, hebben geen duidelijke factor structuur en risico-correcties met behulp van het Fama-French Carhart model met of zonder tijds-variërende factorladingen heeft amper

invloed. Nieuw onderzoek in asset pricing modellen is nodig om de cross-sectionele verschillen in rendementen uit te leggen .

2. Het beschouwen van een lange horizon kan een goede invloed hebben op beleggingen, zelfs wanneer we alleen de korte-termijn prestaties meten.
Zelfs wanneer we de prestaties van portefeuilles evalueren met korte-termijn prestatie maatstaven zoals een Sharpe ratio, geven specificaties die aandelenrendementen voorspellen voor een langere horizon de best presterende portefeuilles. Deze portefeuilles hebben hoge gemiddelde (risico-gecorrigeerde) rendementen, zijn goed gediversifieerd en hebben slechts lage transactiekosten.

3. Er is nog altijd heel veel heterogeniteit tussen bedrijven in verwachte rendementen, zelfs wanneer we corrigeren voor vele bekende karakteristieken.
Hoewel panel data modellen in staat zijn om een groot gedeelte van de cross-sectionele variatie in aandelenrendementen uit te leggen, suggereert de aanwezigheid en de belangrijkheid van vaste individuele effecten dat voldoende heterogeniteit tussen bedrijven over is die niet kan worden uitgelegd door alleen gebruik te maken van de verschillen in bedrijfskarakteristieken. Dit suggereert dat nieuwe econometrische schattingstechnieken ontwikkeld moeten worden die in staat zijn om op een efficiënte manier informatie uit individuele effecten te halen, terwijl deze technieken ook nog correlaties tussen bedrijfskarakteristieken en individuele effecten moeten toestaan.

4. Met een beetje scepsis kunnen lange-termijn beleggers profiteren van het timen van de markt.
De vondst dat een naïeve implementatie van strategische beleggingskeuze leidt tot slechte prestaties betekent niet dat lange-termijn beleggers geen market timing moeten doen. Sceptische beleggers - gebruikmakend van sceptische shrinkage prioren - kunnen veel baat hebben bij market timing. Hun scepsis staat hen toe te profiteren van veranderingen in markt condities, maar redt hen ook van zeer grote reducties in hun portefeuille waardes. Met andere woorden, we delen de negatieve visie van DeMiguel, Garlappi en Uppal (2007) niet dat op data gebaseerde methodes om portefeuilles te berekenen niet in staat zijn om de eenvoudige 1/N regel te kunnen verslaan out-of-sample.

5. Beleggingsstrategieën maar ook voorspellingen moeten niet geëvalueerd worden met kwadratische verliesfuncties, maar met asymmetrische nut/verlies functies.

Beleggers hebben niet noodzakelijkerwijs een voorkeur voor specificicaties die leiden tot de beste voorspellingen wat betreft gemiddelde kwadratische fouten. Ze vermijden risico, evalueren grote verliezen en winsten verschillend en waarderen daarom specificaties die in staat zijn om grote verliezen te voorkomen. Dit feit moet zowel in de evaluatie van beleggingsstrategieeën als de evaluatie van voorspellingen naar voren komen. Het is echter standaard in de literatuur om gebruik te maken van symmetrische verliesfuncties. Een belangrijke uitzondering is Christoffersen en Dieboldt (1997).

6. Hoewel de optimale lange-termijn strategieën veel moeilijker te berekenen zijn dan korte-termijn strategieën, geven ze vaak dezelfde resultaten.

De optimale lange-termijn dynamische strategieën zijn veel moeilijker uit te rekenen dan myopische strategieën, omdat men voor de eerste dynamisch programmeren, veel variabelen en duizenden gesimuleerde paden nodig heeft. Toch is er bijna geen verschil in hun out-of-sample prestaties. De hoofdreden is schattingsfouten. Geschatte myopische en geschatte dynamische strategieën benaderen de echte optimale dynamische strategie even goed. Een andere uitleg kan worden gevonden in de laatste twee hoofdstukken. Wanneer model onzekerheid en model instabiliteit meegenomen worden in het beslisproces, kunnen optimale lange-termijn beleggingen in aandelen bijna gelijk of soms zelfs lager zijn dan korte-termijn beleggingen.

7. Ook al suggeren resultaten voor een korte horizon dat parameter en model onzekerheid en model instabiliteit niet belangrijk zijn, toch zijn deze concepten erg belangrijk voor portefeuille beslissingen voor een langere horizon.

Het meenemen van parameter onzekerheid, model onzekerheid en in minder mate model instabiliteit heeft bijna geen invloed op korte-termijn beleggingen. Echter, omdat (i) een ongunstige parameter set of model blijft voortduren voor de hele beleggingshorizon en (ii) (onverwachte) tijdsvariatie in (vooral) fout volatiliteit zeer persistent is, worden deze effecten belangrijker voor een langere beleggingshorizon.

8. Het geschatte risico van aandelen verandert behoorlijk over de tijd en moet in ogen-schouw genomen worden door lange-termijn beleggers.

De volatiliteit van de fouten in aandelenrendementen verandert behoorlijk en persis-tent gedurende de tijd. Zelfs in situaties waarin volatiliteit constant is, verandert de volatiliteit van de verdeling van toekomstige aandelenrendementen behoorlijk door model onzekerheid. Een van zijn belangrijke componenten - de volatiliteit van de con-ditionele gemiddeldes van de verschillende modelvoorspellingen - is laag wanneer alle modellen dezelfde voorspellingen afleveren zoals in de jaren '60 en '70, maar hoog in periodes wanneer de modellen juist verschillende voorspellingen leveren zoals in 1929 en 2008. Het negeren van deze stijging in volatiliteit leidt tot een te hoge belegging in aandelen.

9. Het meenemen van onzekerheid is belangrijk in het maken van beslissingen, maar niet alle onzekerheid is relevant.

Parameter onzekerheid, model onzekerheid en de tijdsvariatie in de fout volatiliteit zijn allemaal erg belangrijk voor een lange horizon. Echter, voor een korte horizon kunnen deze effecten (vooral de eerste 2) worden genegeerd. Verder, de lage persistentie van de tijds-variatie in intercepten en hellingshoeken betekent dat deze tijdsvariatie kan worden genegeerd voor zowel een korte als een langere beleggingshorizon.

10. De aandelenmarkt is niet een beter oord voor lange-termijn beleggers dan voor korte-termijn beleggers.

Het meenemen van model onzekerheid kan aandelen veel risicovoller maken op de lange-termijn dan op de korte-termijn en kan leiden tot lagere aandelenbeleggingen voor lange-termijn beleggers. Deze vondst is ook gedaan in een recent artikel van Pastor en Stambaugh (2010). In hun set-up nemen ze mee dat verwachte rendementen niet perfect gecorreleerd hoeven te zijn met een lineaire combinatie van een klein aantal voorspellers. Ze vinden dat dit feit leidt tot rendementen die een veel hogere variantie hebben op de lange termijn. In tegenstelling tot hun vinden we dat aandelen zowel risi-covoller als veiliger kunnen zijn op de lange termijn dan op de korte termijn. Cruciaal is het feit dat de invloed van model onzekerheid verandert over de tijd. Het meenemen van model instabiliteit versterkt deze conclusies.

11. Lange-termijn beleggers willen inflatie-geïndexeerde overheidsobligaties.

Als lange-termijn beleggers zich realiseren dat de echte parameters en het echte model onbekend zijn, dan heeft hun verdeling van toekomstige aandelenrendementen in de meest extreme gevallen een hogere variantie, een negatieve scheefheid en extreme dikstaartigheid door model onzekerheid. In zo'n setup preferen lange-termijn beleggers het echte risicovrije lange-termijn financieel product - inflatie-geïndexeerde staatsobligaties - boven de optimale combinatie van T-bills, aandelen en nominale obligaties, omdat het zogenaamde "certainty equivalent" van deze optimale combinatie lager is dan 1.

12. De onenigheid over de verschillende voorspellers van aandelenrendementen kan voor een groot gedeelte worden toegeschreven aan de voorspelhorizon.

Aangezien er slechts beperkte tijdsvariatie - vooral vergeleken met de hoeveelheid schattingsfouten - in de hellingscoefficienten van de dividend - prijs ratio en de yield spreiding is in de vergelijking voor aandelenrendementen, speculeren we dat de schattingsperiode niet de belangrijkste reden is voor onenigheid in de academische wereld over mogelijke voorspellers van aandelenrendementen. De voorspelhorizon speelt waarschijnlijk zelfs een grotere rol. Terwijl voor een korte horizon alleen voorspellers van aandelenrendementen belangrijk zijn, worden voorspellers van voorspellers belangrijke voorspellers voor een langere voorspelhorizon.

13. Het meenemen van relatief nieuwe productklassen in het keuzemenu voor beleggers is niet aantrekkelijk voor lange-termijn beleggers door de beperkte informatie die we hebben over hen.

Aangezien we slechts een korte data-set beschikbaar hebben om modellen voor nieuwe financiële production te schatten, verwachten we dat parameter onzekerheid een belangrijke rol zal spelen voor zulke modellen. Omdat parameter onzekerheid de voorspelonzekerheid flink laat toenemen op de lange termijn, zullen lange-termijn investeerders bijna niet investeren in nieuwe producten tenzij ze zeer grote baten opleveren in vergelijking met meer traditionele klassen zoals aandelen en obligaties. Nieuwe beleggingsklassen worden alleen aantrekkelijk voor lange termijn beleggers als er voldoende data beschikbaar is.

Tenslotte een kleine vooruitblik op toekomstig onderzoek. Zelfs met meer dan 80 jaar data, speelt onzekerheid nog altijd een zeer belangrijke rol voor lange-termijn beleggers. De veelbelovende vondst rond 2000 dat aandelen teruggaan naar een gemiddelde en daarom veiliger zijn op de lange termijn wordt meer dan teniet gedaan door het parameter en model onzekerheid effect. Dit betekent dat zonder extra informatie over toekomstige aandelenrendementen, lange-termijn beleggers niet kunnen profiteren van de "mean-reversion" en daarom niet meer geld moeten beleggen in de aandelenmarkt dan korte-termijn beleggers. Daarom hebben we andere manieren nodig om structuur op te leggen / prior informatie toe te voegen aan econometrische modellen om parameter en model onzekerheid te beperken. Een veelbelovende onderzoeksrichting is om economische theorie te gebruiken om restricties op te leggen aan modellen die (bij benadering) waar zijn en op deze manier voorspellingen te verbeteren en voorspelonzekerheid te verminderen. Contante waarde modellen lijken een interessante richting van toekomstig onderzoek te zijn.

Curriculum Vitae

Bart Diris was born on December 4, 1983 in Born, The Netherlands. He completed a Bachelor of Science (2002-2005) and Master of Science (2005-2006) in Econometrics and Operations Research at Maastricht University. During his Bachelor studies, Bart spent a semester abroad at the University of California at Berkeley. In August 2006, he obtained his master's degree with honors. His master thesis was awarded with the best thesis award of the Faculty of Economics and Business Administration of Maastricht University.

He joined the Finance and Quantitative Economics departments of Maastricht University in September 2006 as a PhD candidate. His research was supported by the Maastricht Research School of Economics of Technology and Organization (METEOR), The Network for Studies on Pensions, Aging and Retirement (Netspar) and the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). Bart's work was or will be presented at various universities and international conferences, among them the European Meeting of the Econometric Society in Milan in 2008, the European Finance Association meeting in Frankfurt in 2010, a seminar at Harvard University in 2009 and the annual conference of The Society for Financial Econometrics (SoFiE) in Chicago in 2011. He conducted part of his research as a visiting PhD student at The Institute for Financial Research (SIFR) in Stockholm (January - March 2009) and as a visiting research fellow at Harvard University (September - December 2009). Bart finished his PhD research in January 2011.

In February 2011, Bart started as a (trainee) Assistant Professor at the Econometric Institute at Erasmus University Rotterdam. Here, he is responsible for the Master of Science course Portfolio Management in the Quantitative Finance master. His research interests are financial econometrics, in particular strategic asset allocation, asset pricing and time-series econometrics.