Forecasting Financial Time Series Using Model Averaging

ISBN: 978 90 5170 914 8

Cover design: Crasborn Graphics Designers bno, Valkenburg a.d. Geul

This book is no. 415 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Forecasting Financial Time Series Using Model Averaging

Voorspellen van financiële tijdreeksen met behulp van model wegingen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the rector magnificus

Prof.dr. S.W.J. Lamberts

and in accordance with the decision of Doctoral Board.

The public defence shall be held on

Friday, 23 November 2007 at 11.00 hrs

by

Francesco Ravazzolo
born in Padova, Italy

Erasmus Universiteit Rotterdam

Doctoral Committee

Promoters:	Prof.dr. H.K. van Dijk Prof.dr. M.J.C.M. Verbeek
Other members:	Prof.dr. M. Billio Dr. R. Paap Prof.dr. D.J.C. van Dijk

Preface

Four years ago I started my Ph.D. at Erasmus University and Tinbergen Institute. I am of course thrilled to have reached the end of my project, to which several people have contributed by providing enthusiastic guidance and support along the way.

First of all, I would like to thank my promoters, Herman K. van Dijk and Marno Verbeek. In the last four years Herman and Marno stimulated my research ideas, shaped my thinking and extensively taught me accuracy. But at the same time they gave me the possibility to apply what I was learning, working with different people on several projects. I thank them again to have believed in very informative priors on my scientific skills. I also thank the members of the inner doctoral committee, Monica Billio, Richard Paap, Dick van Dijk, as well as the additional members of the plenary committee Philip Hans Franses, Massimo Guidolin and Domenico Sartore both for supporting the project at different stages and for reading this manuscript.

Parts of this thesis are based on joint work with several colleagues. In this respect I thank my co-authors Chen Zhou, Philip Hans Franses, Christian Huurman, Richard Paap, Michiel de Pooter, Dick van Dijk, Herman K. van Dijk and Marno Verbeek for the pleasant cooperation, helpful discussion and tremendous day-to-day support.

One of the key ingredient to finish this thesis has been perseverance. The atmosphere at Erasmus helped to keep it. Several people contributed to it, in particular I would like to mention Chen, Christian, Jeroen, Joop, Michiel, Sandra, the lunch group, TI Knudde players, Smitse fellows, and the staff of Econometric and Tinbergen Institute.

Finally, I thank my parents Vanarosa and Ettore, my sister Fabiola, Marilena, Roberto and all friends for their support. The last and most important "thank you" goes to Claudia, who coped in these four years with econometric programs, long distance, flights and much more.

Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

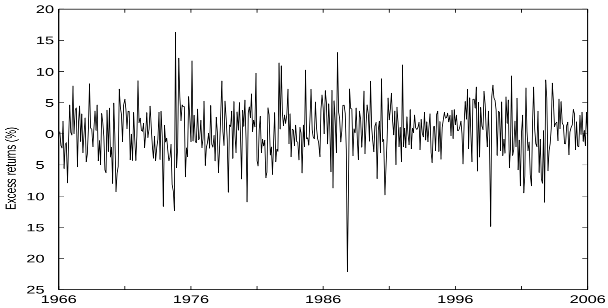
Introduction

1.1 Motivation

Accurate forecasting of such future events as the next World Cup triumph of Italy or a stock market crash in the USA constitute a fascinating challenge for theoretical and applied econometric research. For a long time it was thought that the best approach to forecast some future event is to estimate a parametric model on particular data and then use the estimated model for forecasting. Different parametric models give different forecasts and when wide set of forecasts is available, the "art" of the economic consultant is usually to accept the best one and discard the other ones. But, discarded forecasts may have some independent valuable information and including them in the forecasting process may provide more accurate results. Bates and Granger (1969) introduced in economics the idea of combinations of forecasts of different models as a simple and effective way to obtain improvements in forecast accuracy compared to those of individual models. The success of forecast combinations is related to the evidence that a decision maker in almost all cases cannot identify ex ante the exact true process, but different models play a complementary role in approximating it. Moreover, data may be subject to structural instability, in the sense that the relevance of predictors varies over time, and a simple strategy to cope with future unknown instability is to average models that forecast in a different way. In simple terms, believing in a single model may be dangerous, and addressing model uncertainty by averaging different models in making forecasts may be very beneficial.

The literature on forecast combinations is extensive, see for example Clemen (1989) and Timmermann (2006). Despite this, research has often focused on macroeconomic applications and little attention is given to combine forecasts of financial time series. In macroeconomic applications equal weight schemes or other simple averaging schemes that do not require estimates of model weights provide accurate results. But financial

Figure 1.1: S&P500 Excess returns



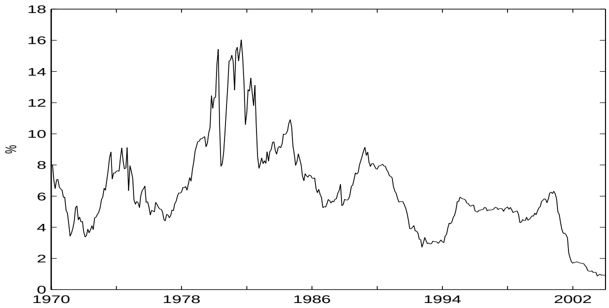
Note: The figure presents the excess returns on the S&P500 over the sample 1966:1-2005:12.

time series have particular and different stylized facts. For example, stock index returns are characterized by such features as low predictability and structural instability, see e.g. Figure 1.1, making averaging methods that do not model these features inadequate. Therefore, averaging schemes that model the time variation in the underlying data generating process and assign almost zero probability to a subset of the individual models that perform poorly, may improve the forecasting performance.

Model averaging approaches can be classified in two broad categories: frequentist and Bayesian model averaging. The two classes of schemes are conceptually different. Frequentist combinations combine individual forecasts by minimizing some loss function, Bayesian model averaging average individual models by assigning posterior probabilities to each specification and then compute combined forecasts. It is difficult to compare and choose between Bayesian and frequentist averaging techniques. And there is not a given rule indicating that one method should be preferred to the other one. Bayesian model averaging incorporates two sources of uncertainty, parameter uncertainty as well as model uncertainty, coping more extensively with the uncertainty of future events. Frequentist methods are often very easy to be implemented and they do require only to estimate weights and not to derive a complete density of interest.

In this thesis we focus on forecasting financial time series using model averaging schemes as a way to produce optimal forecasts. We derive and discuss in simulation exercises and empirical applications model averaging techniques that can reproduce stylized facts of financial time series, such as low predictability and time-varying patterns. In particular, we start by revaluating the use of time varying weight schemes. We show that when individual forecasts contain different information these schemes provide very accurate results as they add individual forecasts. Moreover, they scheme capture in-sample structural instability, in the sense that the relevance of predictors varies over time, providing superior forecasts. In the following chapter, we proposes a new Bayesian model averaging method that not only models in-sample structural instability, but also forecast

Figure 1.2: US 3-month zero-coupon yield



Note: The figure presents the end-of-month unsmoothed 3-month coupond yields constructed using the Fama and Bliss (1987) bootstrap method. Sample period is Janaury 1970 - December 2003.

out-of sample instability. Our Bayesian model averaging method has not a limitation on the number of competitive models, meaning that it can combine a huge set of individual models. We show that the averaging strategies outperform individual models in forecasting S&P 500 excess returns. Then, we move to the analysis of model averaging when individual forecasts are strongly homogenous, and competitive forecasting models are possible non-nested. As typical example we choose the term structure of US interest rates. Figure 1.2 plots the end-of-month unsmoothed 3-month coupond yields. Again we show that individual models play a complementary role in approximating the data generating process and mitigating model uncertainty by model averaging leads to substantial gains in forecasting performance.

We emphasize that model averaging is not a "magic" methodology which solves *a priori* problems of poorly forecasting. Averaging techniques have an essential requirement: individual models have to fit data. Therefore, we advice decision makers that the first step to compute optimal forecasts is to implement sound individual models. As empirical application, we show the role of weather forecasts in predicting electricity prices and we discuss whether those variables can be used to improve standard electricity price forecasting models.

1.2 Outline

In this section we provide a general outline of the thesis and its contributions to previous research. A detailed outline with references to related literature can be found in the introduction of each chapter. The thesis is partitioned in six chapters. In Chapter 2 we focus on the use of time varying model weight combinations. This chapter is based on Ravazzolo *et al.* (2007b). We extend the research on forecast combinations by investigating the relative merits of different combining procedures in a fairly simple context. More specifically, we evaluate the performance of eight different forecast combination methods in simulation examples where the data generating process is subject to low predictability, high noise and possible non-stationary. We suppose that a set of different models play complementary roles in generating data and the true process cannot be identified *a priori*. We compare two simple methodologies, which do not require parameter estimates, to two OLS weight regressions, and to a more advanced time varying weight scheme. And we also consider three Bayesian model averaging schemes in the analysis. Precisely, we include two Bayesian schemes based on marginal and predictive likelihoods respectively, and third one, which is a novel Bayesian scheme that allows for weight instability. We use an adequate long out-of-sample period to evaluate strategies and in particular their reactions to shifts in the data generating process. We show that when the heterogeneity among forecasts of individual models is strong due to the fact that individual models contain different information, simple and "standard" Bayesian averaging strategies perform poorly. On contrary, unconstrained OLS and (Bayesian) time varying weights provide more accurate results as they add individual forecasts. Furthermore, when the structural instability is high, the time varying weight schemes estimate the instability giving the most accurate forecasts. As an illustration, we consider forecasting the returns on the S&P 500 index. We apply two different approaches in forecasting: a model based on a set of financial and macroeconomic variables that should have explanatory power, and the "Halloween indicator" based on the popular market saying "Sell in May and go away". Then, we compare individual forecasts to forecast combinations. We find that the time varying weight schemes give the best forecasts in term of symmetric loss functions. As an investor is more interested to the economic value of stock return forecasting models than their precisions, we test our findings in an active short-term investment exercise, with an investment horizon of one month. The time varying weight schemes also provide the highest economic gains.

In Chapter 3, we extend the analysis in the previous chapter to a new Bayesian averaging scheme that models structural instability carefully. This chapter is based on Ravazzolo *et al.* (2007a). We develop a return forecasting methodology that allows for instability in the relationship between stock returns and predictor variables, for model uncertainty, and for parameter estimation uncertainty simultaneously. On the one hand, the predictive regression specification that we put forward allows for occasional structural breaks of random magnitude in the regression parameters. On the other hand, we allow for uncertainty about the inclusion of the forecasting variables in the model and about the parameter values by employing Bayesian model averaging. We consider an empirical application to predicting monthly US excess stock returns using a set of 11 financial and macro-economic predictor variables. We find that over the period 1966-2005, several structural breaks occurred in the relations between the excess stock returns and predictor variables such as the dividend yield and interest rates. These changes appear to be caused by important events such as the oil crisis, changes in monetary policy, the October 1987 stock market crash, and the internet bubble at the end of the 1990s. Although incorporating the different sources of uncertainty does not lead to large improvements in the statistical accuracy of excess return forecasts, their economic value in asset allocation decisions is considerable. We find that a typical investor would be willing to pay up to several hundreds of basis points annually to switch from a passive buy-and-hold strategy to an active strategy based on a return forecasting model that allows for model and parameter uncertainty as well as structural breaks in the regression parameters. The active strategy that incorporates all three sources of uncertainty performs considerably better than strategies based on more restricted return forecasting models.

In Chapter 4 we focus on forecasting the term structure of U.S. interest rates. This chapter is based on de Pooter *et al.* (2007). In spite of the powerful advances in term structure modelling and forecasting we feel that a number of issues regarding estimation and forecasting have so far been left nearly unaddressed in the interest rate literature. We try to fill in some of these gaps by investigating parameter uncertainty and, in particular, model uncertainty. Especially for VAR and affine models, which are highly parameterized if one attempts to model the whole term structure, parameter uncertainty is likely to be substantial and needs to be accounted for. We estimate each model and generate forecasts by applying frequentist maximum likelihood techniques as well as Bayesian sampling techniques to gauge the effects of explicitly taking into account parameter uncertainty. Regarding model uncertainty, when looking at the historical time series of (U.S.) interest rates we can easily identify subperiods across which yield dynamics are quite different. Likely reasons are, for example, the reigns of different FED Chairmen, most notably that of Paul Volcker, or the strong decline in interest rate levels accompanied by a pronounced widening of spreads in the early 90's and after the burst of the Internet bubble. It will be unlikely that any individual model is capable of consistently producing accurate forecasts in each of these subperiods. The forecasting performance of the models we consider in this study does indeed vary substantially across subperiods. In these situations, combining forecasts yields diversification gains and can therefore be attractive relative to relying on forecasts from a single model. Moreover, even if it would be possible to identify the best individual model, creating forecast combinations can be useful to scale down the magnitude of forecast errors. In addition to these two focal points, we also further examine the use of macroeconomic diffusion indices in term structure models. We analyze each model both with and without macrofactors to assess the benefits of adding macroeconomic information for shorter as well as longer forecast horizons. Using an out-sample period of 1994-2003 we show that the predictive ability of individual models varies over time considerably, which a prime example being the Nelson and Siegel (1987) model. We show by analyzing two five-year subperiods that models which incorporate macroeconomic variables seem more appropriate in subperiods during which the uncertainty about the future path of interest rates is substantial. This is especially the case for the early 2000s with the pronounced drop in interest rates and the widening of spreads. Models without macro information do particularly well in subperiods where the term structure has a more stable pattern such as in the early 90s with models outperforming the random walk RMSPE by sometimes well over 30%. That different models forecast well in different subperiods confirms ex-post that alternative model specifications play a complementary role in approximating the data generating process, which provides a strong claim for using forecast combination techniques as opposed to believing in a single model. Our forecast combination results confirm this conjecture. We show that combining forecasts, in particular when using a weighting method that is based on relative historical performance results and Bayesian inference on individual models, results in highly accurate forecasts. The gains in terms of forecast performance are substantial, especially for longer maturities, and are consistent over time.

In Chapter 5 we attempt to shed more light on forecasting performance of stochastic day-ahead price models. This chapter is based on Huurman *et al.* (2007). We examine six stochastic price models to forecast day-ahead prices of the two most active power exchanges in the world: the Nordic Power Exchange and the Amsterdam Power Exchange. Three of these forecasting models include weather forecasts. Firstly, considering that operators make decisions today on tomorrow's electricity, the real weather of tomorrow is unknown at that moment, and the only available information of weather comes from the weather forecasts. Therefore, we use weather forecasts as predictors, which are more appropriate than real weather. The empirical study agrees with this intuition. Secondly, we consider a set of weather variables which capture significant and interpretable supply and demand effects. Thirdly, since we find that the influence of the weather forecasts on the electricity prices is non-linear, we use non-linear transformations of the weather forecasts in our new models. Finally, we implement specific models for different power markets due to their heterogeneity in weather conditions and production plants. We find that an extended ARMA model, which includes nonlinear combinations of next-day weather forecasts, yields the best forecasting results for predicting one day-ahead power prices. This model has some predictability power to anticipate prices jumps. Intuitively, adverse climate conditions often lead to sharp increases in demand resulting in supply shortages in electricity. We investigate carefully the relation between prices and weather. We find that the weather forecasts influence the electricity prices via the demand as well as the supply side, and when production is less related to weather, which is the case for Amsterdam Power Exchange, the weather forecasts play only a minor role. We also show that a GARCH specification extended with weather forecast variables provides accurate forecasts. This result contradicts earlier findings that 'standard' GARCH models would predict electricity prices poorly.

To sum up, the research finds an increase of forecasting power of financial time series when parameter uncertainty, model uncertainty and optimal decision making are included. Although the implementation of these techniques is not often straightforward and it depends on the exercise that it is studied, the predictive gains are statistically and economically significant over different applications. In particular, in two studies we show that incorporating model averaging and modelling instability improves stock index forecast performances. In the third study we find that assessing model uncertainty and adding macroeconomic factors leads to substantial gains in predicting the US term structure. In the fourth study we document the explanatory power of weather to predict day-ahead power spot prices and the improvements in forecast accuracy when weather forecast variables are included in popular forecasting electricity price models. Several extensions are worth to be considered, and they such as possible limitations are detailed discussed in each chapter. Among them, the new algorithm developed in Chapter 3 may be implement in a multivariate framework, by for example forecasting international stock markets and allocating portfolios on them. Moreover, the forecasting techniques may be applied to the analysis of density forecasts. Market operators, such as financial investors or central bank decision makers, are becoming always more interested to the complete distributions of the assets of interests to implement their decisions.

Chapter 2

Predictive gains from forecast combinations using time varying model weights

2.1 Introduction

When a wide set of forecasts of some future economic event is available, decision makers usually attempt to discover which is the best forecast, then accept this and discard the other ones. However, the discarded forecasts may have some independent valuable information and including them in the forecasting process may provide more accurate results. An important explanation is related to the fundamental assumption that in most cases one cannot identify *a priori* the exact true economic process or the forecasting model that generates smaller forecast errors than its competitors. An alternative reasonable assumption appears to be one where different models may play a - possibly temporary - complementary role in approximating the data generating process. Furthermore, perhaps due to the presence of private information such as forecasters' subjective judgements or differences in modelling approaches, it may not be possible to pool the underlying information sets and construct a 'super' model that nests each of the underlying forecasting models. In these situations, forecast combinations are viewed as a simple and effective way to obtain improvements in forecast accuracy.

Since the seminal article of Bates and Granger (1969) several papers have shown that combinations of forecasts can outperform individual forecasts in terms of symmetric loss functions. For example, Stock and Watson (2004) find that forecast combinations to predict output growth in seven countries generally perform better than forecasts based on single models. Marcellino (2004) has extended this analysis to a large European data set with broadly the same conclusion. However, several alternative combination schemes are available and it is not clear which is the best scheme, either in a classical or a

Bayesian framework. For example, Hendry and Clements (2004) and Timmermann (2006) show that simple combinations[1] often give better performance than more sophisticated approaches. Further, using a frequentist approach, Granger and Ramanathan (1984) propose the use of coefficient regression methods, Hansen (2007) introduces a Mallows' criterion, which can be minimized to select the empirical model weights, and Terui and van Dijk (2002) generalize the least squares model weights by reformulating the linear regression model as a state space specification where the weights are assumed to follow a random walk process. Guidolin and Timmermann (2007) propose a different time varying weight combination scheme where weights have regime switching dynamics. Stock and Watson (2004) and Timmermann (2006) use the inverse mean square prediction error (MSPE) over a set of the most recent observations to compute model weights. In a Bayesian framework, Madigan and Raftery (1994) revitalize the concept of Bayesian model averaging (BMA) and apply it in an empirical application dealing with Occam's Window. Recent applications suggest its relevance for macroeconomics (Fernández *et al.*, 2001 and Sala-i-Martin *et al.*, 2004). Strachan and van Dijk (2007) compute impulse response paths and effects of policy measures using BMA in the context of a large set of vector autoregressive models. Geweke and Whiteman (2006) apply BMA using predictive and not marginal likelihoods.

This Chapter contributes to the research on forecast combinations by investigating the relative merits of eight combination schemes in simulation exercises where the data generating process is subject to low predictability, structural instability, in the sense that the relevance of forecasting factors varies over time, and fat tails. Sensitivity of results with respect to misspecification of the number of included predictors and the number of included models is explored.

The different combination schemes are summarized as two simple schemes, which do not require parameter estimates; two schemes that involve OLS weight regressions, and a more advanced time varying weight scheme due to Terui and van Dijk (2002). Next, we include two Bayesian model averaging schemes: the original one first proposed in an empirical application by Madigan and Raftery (1994), and a more recent one in terms of predictive densities given by Geweke and Whiteman (2006)[2]. Finally, we propose a new Bayesian scheme which allows for parameter uncertainty, model uncertainty and time varying model weights simultaneously.

---

[1]In this Chapter simple combinations are defined as combinations that do not require estimating parameters; arithmetic averages constitute a simple example. Complex combinations are defined as combinations that rely on estimating weights that depend on the full variance-covariance matrix and, possibly, allow for time varying weights.

[2]Alternative BMA's exist such as MC^3, or frequentist approaches that share similar features as BACE or thick modelling; but we omit them to simplify the analysis.

As in Aiolfi and Timmermann (2006) we use an adequate long out-of-sample period to evaluate the forecasting performance of the different combination schemes.
Our results indicate that when correlation among forecasts of individual models is low, simple and Bayesian averaging strategies using marginal likelihoods perform poorly, while unconstrained OLS and time varying model weight schemes provide more accurate results. Moreover, when structural instability is high, we explain asymptotically and in a simulation experiment that the time varying combination schemes give the most accurate forecasts.

A second contribution of this Chapter is to provide an empirical illustration, where we consider forecasting the returns on the S&P 500 index by combining individual forecasts from two competing models. The first one assumes that a set of financial and macroeconomic variables have explanatory power, the second one is based on the popular market saying "Sell in May and go away", also known as the "Halloween indicator", see for example Bouman and Jacobsen (2002). Low predictability of stock market return data is well documented, see for example Marquering and Verbeek (2004) and so is structural instability in this context, see for example Pesaran and Timmermann (2002) and Ravazzolo *et al.* (2007a). We confirm these results, and show that the two models, taken individually, perform poorly and in a differential way over time. We continue by applying model averaging and find that the two time varying weight schemes that we apply give the best forecasts in term of symmetric loss functions, confirming the results of the simulation exercises. Moreover, as an investor is more interested in the economic value of a forecasting model than in its forecast accuracy, we test our findings in an active short-term investment exercise, with an investment horizon of one month. Again, the time varying weight schemes provide the highest economic gains.

This Chapter proceeds as follows. In Section 2.2 we describe the eight different forecast combination schemes. In Section 2.3 we report results from simulation exercises in predicting future unknown point values. In Section 2.4 we give results from an empirical application to US stock returns and show that forecast combinations give economic gains. Section 2.5 concludes. In the Appendices some technical details are presented.

2.2 Forecast combinations

Two schemes are based on simple constant weights; three are frequentist approaches based on estimated (time varying) model weights; two are "known" Bayesian averaging schemes, the final one is a new Bayesian averaging scheme that allows for time varying weights. We note that the vast majority of studies on forecast combination deals with point forecasts, and we also focus on this.

We start with a brief description of the basic set up of the simulation experiments. Suppose two time series $y_1 = \{y_{s,1}\}_{s=1}^{S}$ and $y_2 = \{y_{s,2}\}_{s=1}^{S}$ are generated from the following models:

$$y_{s,1} = \alpha_1 + x_{s,1}^{'}\beta_1 + \epsilon_{s,1} \tag{2.1}$$

$$y_{s,2} = \alpha_2 + x_{s,2}^{'}\beta_2 + \epsilon_{s,2} \tag{2.2}$$

where $x_{s,1}$ and $x_{s,2}$ are $(k_1 \times 1)$ and $(k_2 \times 1)$ vectors of predictor variables respectively, where α_1, α_2 are two scalar parameters and β_1, β_2 are $(k_1 \times 1)$ and $(k_2 \times 1)$ vectors of parameters, and where $\epsilon_{s,1}$ and $\epsilon_{s,2}$, $s = 1, ..., S$, are two zero mean i.i.d. disturbances with variances σ_1^2 and σ_2^2, respectively. The simulated data generating process (DGP) is a linear combination of the previous two models:

$$y_s = y_{s,1}c_{s,1} + y_{s,2}c_{s,2}, \tag{2.3}$$

where $c_{s,1}$ and $c_{s,2}$ are two possibly time varying scalars. We refer to $c_{s,1}$ and $c_{s,2}$ as DGP weights.

Equations (2.1) and (2.2) are estimated over the sample period $[1, ..., T]$ with $T < S$ to compute two independent one-step ahead forecasts $\widehat{y}_{T+1,1}$ and $\widehat{y}_{T+1,2}$, combined to compute a forecast of y_{T+1}. We let $\widehat{y}_{T+1} = g(\widehat{y}_{T+1,1}, \widehat{y}_{T+1,2}, w_{T+1})$ be the combined point forecast as a function of the underlying single forecasts $\widehat{y}_{T+1,1}$ and $\widehat{y}_{T+1,2}$, the forecast combination scheme g, and the vector of the parameters of the combination w_{T+1}[3]. The values of the optimal combination $\widehat{w}_{T+1}$ solve the problem:

$$\min_{w_{T+1}} E[L(e_{T+1}(w_{T+1}))|\widehat{y}_{T+1,1}, \widehat{y}_{T+1,2}], \tag{2.4}$$

where $e_{T+1} = y_{T+1} - g(\widehat{y}_{T+1,1}, \widehat{y}_{T+1,2}, \widehat{w}_{T+1})$ is the forecast error from the combination, and where L is the loss function, which for simplicity we assume to depend only on the forecast error. We emphasize that the vector $\widehat{w}_{T+1}$ is not necessarily an estimate of the vector $[c_{T+1,1}, c_{T+1,2}]'$, but refers to estimated weights that minimize the loss function. The general class of combination schemes in (2.4) comprises non-linear as well as time-varying methods.

In most cases there is no closed form solution of equation (2.4), but analytical results may be computed imposing restrictions on the loss function and making distributional restrictions on the forecast errors. Often it is simply assumed that the objective function is the mean squared error (MSE) loss function:

$$L(e_{T+1}(w_{T+1})) = \theta(\widehat{y}_{T+1} - y_{T+1})^2 \quad \theta > 0. \tag{2.5}$$

[3]Note that w_{T+1} may also be a vector of constants.

For this case the combined forecasts choose a combination of the individual forecasts that best approximates the conditional expectation, $E(y_{T+1}|\widehat{y}_{T+1})$. In the five frequentist approaches that we apply we assume the MSE loss function and we fix $\theta = 1$. Different distributional restrictions, for example assuming a time varying θ imply different estimation techniques in equation (2.4).

As a next step we expand the sample period with the observation y_{T+1} and we compute new individual and combination forecasts for the value y_{T+2}. We repeat the procedure to compute H forecasts where $T + H = S$.

2.2.1 Simple combination schemes

Following Timmermann (2006) we define simple combination schemes as cases that do no require estimating (many) parameters, in particular do not require estimating the full variance-covariance matrix. Moreover, these schemes are distinguished by the restriction that the weight coefficients add up to unity.

The forecasts on y_{T+1} given by simple combination schemes can be written as:

$$\widehat{y}^{(j)}_{T+1} = \widehat{y}_{T+1,1}\widehat{w}^{(j)}_{T+1,1} + \widehat{y}_{T+1,2}\widehat{w}^{(j)}_{T+1,2}, \tag{2.6}$$

where $(\widehat{w}^{(j)}_{T+1,1}, \widehat{w}^{(j)}_{T+1,2})$, $j = 1, 2$, are computed following schemes 1 and 2 below.

Scheme 1: Equal weights

$$\widehat{w}^{(1)}_i = 1/n \tag{2.7}$$

where $i = 1, 2$. Extension to more general case with n individual models is straightforward. Equal weights are optimal in situations when the individual forecast errors have the same variance and identical pair-wise correlations, see Timmermann (2006).

Scheme 2: Inverse Mean Square Prediction Error (MSPE) weights

Scheme 2 derives weights from the models' relative inverse MSPE performances computed over a window of the previous υ periods, see Timmermann (2006). Estimation errors in combination weights tend to be particularly large due to the difficulties in precisely estimating the covariance matrix of the forecast error. One answer to this problem is to ignore correlation across forecast errors and making combination weights that reflect performance of each individual model relative to the performance of the average model. The MSPE at time T over the previous υ forecasts for model $i = 1, 2$ is defined as:

$$MSPE^{\upsilon}_{T,i} = \frac{\sum_{j=0}^{\upsilon-1}(\widehat{y}_{T-j,i} - y_{T-j})^2}{\upsilon} \tag{2.8}$$

The weights are computed as:

$$\widehat{w}^{(2)}_{T+1,i} = \frac{(1/MSPE^{v}_{T,i})}{\sum_{j=1}^{2}(1/MSPE^{v}_{T,j})} \tag{2.9}$$

2.2.2 Estimated weight combination schemes

The next three combination schemes estimate the weights in regression form, add a constant term, and do not impose that the weights add to 1.

Scheme 3: Constant OLS weights

The weights are equal to the OLS estimators of the weights (w_0, w_1, w_2) in equation:

$$y_t = w_0 + \widehat{y}_{t,1}w_1 + \widehat{y}_{t,2}w_2 + u_t; \quad u_t \sim N(0, s^2) \tag{2.10}$$

where $t = 1, .., T$, and w_0 is a constant term[4]. The estimation of the weights, while attractive in the sense of minimizing forecast errors, introduces parameter estimation errors. Therefore, one may estimate weights for the first forecast and then fix these as constant over the remaining out-of-sample period.

The forecast on y_{T+1} given by the estimated combination scheme is given as:

$$\widehat{y}^{(3)}_{T+1} = \widehat{w}^{(3)}_0 + \widehat{y}_{T+1,1}\widehat{w}^{(3)}_1 + \widehat{y}_{T+1,2}\widehat{w}^{(3)}_2 \tag{2.11}$$

where $(\widehat{w}^{(3)}_0, \widehat{w}^{(3)}_1, \widehat{w}^{(3)}_2)$ are the OLS estimates of the parameters (w_0, w_1, w_2) in (2.10). To compute the following $H - 1$ forecasts, the same estimated weights $(\widehat{w}^{(3)}_0, \widehat{w}^{(3)}_1, \widehat{w}^{(3)}_2)$ are applied.

Scheme 4: Recursive OLS weights

The estimated weights are equal to the recursive OLS estimators of the weights in (2.10). The estimated weights are updated every time when a new observation becomes available.

Scheme 5: Time varying weights

When the conditional distribution of $(y_{T+1}, \widehat{y}_{T+1})$ varies over time, it may be effective to let the combination weights also change over time. Terui and van Dijk (2002) have proposed a method that extends the OLS weight combination. The weights satisfy the following recursions:

$$y_t = w_{t,0} + \widehat{y}_{t,1}w_{t,1} + \widehat{y}_{t,2}w_{t,2} + u_t; \quad u_t \sim N(0, s^2) \tag{2.12}$$

[4]Granger and Ramanathan (1984) explain that the constant term must be added to avoid biased forecasts. They also conclude that this strategy is often more accurate than restricted OLS weights.

$$w_t = w_{t-1} + \xi_t; \quad \xi_t \sim N(0, \Sigma) \tag{2.13}$$

where $w_t = [w_{t,0}, w_{t,1}, w_{t,2}]'$; $t = 1, .., T$; and Σ is a diagonal matrix. The weights are time varying and follow a random walk process. The time varying weight combination may be interpreted as a state space model, where (2.12) is the measurement equation which defines the distribution of y_t, and where (2.13) is the state equation which defines the distribution of the weights for every t. The Kalman filter algorithm can be applied to compute the estimators $\widehat{w}^{(5)}_{t|t-1}$. Appendix A gives details of the computation and explains the difference with the recursive OLS estimator.
The forecasts on y_{T+1} given by schemes 4 and 5 are:

$$\widehat{y}^{(j)}_{T+1} = \widehat{w}^{(j)}_{T+1,0} + \widehat{y}_{T+1,1}\widehat{w}^{(j)}_{T+1,1} + \widehat{y}_{T+1,2}\widehat{w}^{(j)}_{T+1,2} \tag{2.14}$$

where $j = 4, 5$.

2.2.3 Bayesian model averaging

Bayesian approaches have been widely used to construct forecast combinations, see for example Leamer (1978), Hodges (1987), Draper (1995), Min and Zellner (1993), and Strachan and van Dijk (2007). In this approach one does not estimate regression weights and uses those to compute forecasts, but one derives the posterior probability for any individual model and combines these. The predictive density accounts then for model uncertainty by averaging over the probabilities of individual models. Since the output is a complete density, point prediction (for example by taking the mean), distribution and quantile forecasts can be easily derived.
We choose three BMA schemes: the original one proposed in an empirical application by Madigan and Raftery (1994), a more recent one discussed in Geweke and Whiteman (2006), and a new one to be introduced below.

Scheme 6: BMA using marginal likelihood
The predictive density of y_{T+1} given the data up to time T, F_T, is computed by averaging over the conditional predictive densities given the individual models with the posterior probabilities of these models as weights:

$$p(y_{T+1}|F_T) = \sum_{i=1}^{n} P(m_i|F_T)p(y_{T+1}|F_T, m_i) \tag{2.15}$$

where n is the number of individual models; $p(y_{T+1}|F_T, m_i)$ is the conditional predictive density given F_T and model m_i; $P(m_i|F_T)$ is the posterior probability for model m_i. The conditional predictive density given F_T and model m_i is defined as:

$$p(y_{T+1}|F_T, m_i) = \int p(y_{T+1}|\theta_i, F_T, m_i)p(\theta_i|F_T, m_i)d\theta_i \tag{2.16}$$

where $p(y_{T+1}|\theta_i, F_T, m_i)$ is the conditional predictive density of y_{T+1} given F_T, the model parameters $\theta_i = (\alpha_i, \beta_i, \sigma_i^2)'$, and model m_i in (2.1) or (2.2); $p(\theta_i|F_T, m_i)$ is the posterior density for parameter θ_i. The posterior probability for model m_i is:

$$P(m_i|F_T) = \frac{p(y|m_i)p(m_i)}{\sum_{j=1}^{n} p(y|m_i)p(m_j)} \tag{2.17}$$

where $y = \{y_t\}_{t=1}^{T}$; $p(m_i)$ is the prior density for model m_i; and $p(y|m_i)$ is the marginal likelihood for model (m_i) given by:

$$p(y|m_i) = \int p(\theta_i|F_T, m_i)p(\theta_i)d\theta_i, \tag{2.18}$$

$p(\theta_i)$ is the prior density for the parameter θ_i. The integral in equation (2.18) can be evaluated analytically in the case of linear models, but not for more complex forms. Chib (1995), for example, has derived a method to compute the expression also for nonlinear examples. Proper priors for θ_i are usually applied, otherwise the Bartlett paradox may hold and models with less parameters preferred. The point forecast is computed by taking the mean of the predictive density in (2.15).

We note that an alternative Bayesian procedure to compute model weights is presented below under scheme 8.

Scheme 7: BMA using predictive likelihood

Geweke and Whiteman (2006) propose a BMA based on the idea that a model is good as its predictions. The predictive density of y_{T+1} conditional on F_T has the same form as equation (2.15), but the posterior density of model m_i conditional on F_T is now computed as:

$$P(m_i|F_T) = \frac{p(y_T|F_{T-1}, m_i)p(m_i)}{\sum_{j=1}^{n} p(y_T|F_{T-1}, m_j)p(m_j)} \tag{2.19}$$

where $p(y_T|F_{T-1}, m_i)$ is the predictive likelihood for model m_i, e.g. the density derived by substituting the realized y_T in the predictive density of y_T conditional on F_{T-1} given model m_i. We compute the predictive density for month T using information until month $T-1$ and we evaluate the *realized* value for time T using the same density. The resulting probability is then applied to compute the weight for model m_i in constructing the forecast for $T+1$ made at time T[5]. Similar to scheme 6, the point forecast is computed by

[5]Eklund and Karlsson (2007) evaluate the fit of the predictive density over some more observations, by means of the predictive likelihood, and then update the probability density for the forecasts. The latter approach results in weights which are based more on the fit of the model, even when using out-of-sample data, than on the probability of out-of-sample realized values. Our approach incorporates the uncertainty that future out-of-sample values may differ from historical out-of-sample realizations. It would be more

taking the mean of the predictive density in (2.15).

Scheme 8: BMA using time varying model weights
We present a new combination scheme that extends the time varying weight scheme 5 by adding parameter uncertainty and model uncertainty. We reformulate equations (2.12) and (2.13) by substituting the means of the conditional predictive densities $p(y_T|F_{T-1}, m_i)$ given models m_i, $i = 1, 2$ for the point forecasts $\widehat{y}_{T,i}$. Then we apply Bayesian inference using Gibbs sampling to estimate w_t; for details we refer to Appendix C. The result is a set of posterior densities for the model weights given the data F_T, $p(w_{T+1,i}|F_T)$. These posterior densities are used to average over the conditional predictive densities given F_T and model m_i

$$p(y_{T+1}|F_T) = p(w_{T+1,0}|F_T) + \sum_{i=1}^{n} p(w_{T+1,i}|F_T)p(y_T|F_{T-1}, m_i) \tag{2.20}$$

in order to derive the predictive density of y_{T+1} given F_T. The point forecast is computed by taking the mean of the predictive density in (2.20).
Scheme 8 allows for parameter uncertainty by applying Bayesian analysis to individual models m_i, for model uncertainty by combining the conditional predictive densities given F_T and model m_i, and for time varying patterns by assuming a pattern for model weights as in (2.13). It also extends scheme 5 by providing a density forecast and not only a point forecast. Thus, for instance, forecasting and policy measures with respect to risk management can be performed in a more flexible way.

We emphasize that special cases of this proposed scheme may be constructed as Bayesian versions of schemes 3 and 4. More details are presented in Appendix C.

2.3 Simulation exercises

In this section we describe ten simulation exercises to evaluate the eight forecast combination schemes presented in Section 2.2. In exercises I-III the correlation between predictors varies from low to high; in exercises IV-VII misspecification with respect to the number of included predictors and number of included models is explored; exercises VIII-IX deal with structural change; finally exercise X considers the case of fat tailed generated data patterns.

natural to compute the predictive likelihoods as product of the predictive likelihood made for last υ successive forecasts. Some computational problems may arise because any predictive likelihood is in the interval [0,1]. Then, it might be difficult to work with possible small numbers, or if only one predictive value of the υ averaged is close to zero the weight on the respective model will be zero independently by performances in the other periods.

Table 2.1: Simulation design of exercises I-X

	EXERCISES						
PARAMETERS	I,VIII,IX	II	III	IV,V	VI	VII	X
μ_{x_1}	1.00	1.00	1.00	1.00	1.00	1.00	1.00
μ_{x_2}	1.00	1.00	1.00	1.00	1.00	1.00	1.00
μ_{x_3}	-	-	-	-	1.00	1.00	-
μ_{x_4}	-	-	-	-	-	1.00	-
μ_{x_5}	-	-	-	-	-	1.00	
μ_{x_6}	-	-	-	1.00	-	-	-
$\sigma^2_{x_1}$	2.00	2.00	2.00	2.00	2.00	2.00	2.00
$\sigma^2_{x_2}$	2.00	2.00	2.00	2.00	2.00	2.00	2.00
$\sigma^2_{x_3}$	-	-	-	-	2.00	2.00	2.00
$\sigma^2_{x_4}$	-	-	-	-	-	2.00	-
$\sigma^2_{x_5}$	-	-	-	-	-	2.00	-
$\sigma^2_{x_6}$	-	-	-	2.00	-	-	-
ϱ_{x_1,x_2}	0.00	1.00	1.80	0.00	0.00	0.00	0.00
ν	-	-	-	-	-	-	4
α_1, β_1	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_2, β_2	1.00	1.00	1.00	1.00	1.00	1.00	1.00
α_3, β_3	-	-	-	-	1.00	1.00	-
α_4, β_4	-	-	-	-	-	1.00	-
α_5, β_5	-	-	-	-	-	1.00	-
β_6	-	-	-	1.00	-	-	-

Following previous notation, we simulate DGPs in a range of settings from equations (2.1)-(2.3). We fix $T = 240$ and $H = 120$, that is the genuine out-of-sample period has 120 one-step ahead forecasts. The last 60 observations of the in-sample period ($t = 181, .., 240$) are used as initial training period for the combination schemes. We repeat each exercise 1000 times. In all examples we assume that the predictor variables (x) are normally distributed with values for the means (μ), variances (σ^2) and covariances (ϱ) that are specified in Table 2.1. The disturbances are assumed to be i.i.d normal (0,1). We restrict the DGP weights $c_{s,1}$ and $c_{s,2}$ to add to 1 for any s in order to exclude shifts in the unconditional mean of the DGP. In exercises I-VII $\{c_{s,1}\}_{s=1}^{360}$ and $\{c_{s,2}\}_{s=1}^{360}$ are time invariant and the DGP is stationary. In exercises VIII-X time-variation is added. In Bayesian analysis we generally use diffuse proper priors for the model parameters.

For any simulation we compute the MSPE's of the individual forecasts and forecast combinations over the 120 "genuine" one-step ahead forecasts, and its decomposition in bias and variance of the forecast errors. In Table 2.2 we report the average of 1000 MSPE's, bias and variance of the forecasts. For completeness, we also give the same statistics for the correctly specified models (labelled as "correct" model), and the forecast combination where the vector $\widehat{w}_{T+1}$ is identical to $[c_1, c_2]$ (labelled as "given" weights).

2.3.1 Varying correlations between predictors

In exercises I-III a stationary DGP is simulated; c_1 and c_2 are plotted at the top-right corner in Figure 2.1: c_1 is set almost two times c_2. The difference in exercises relates to the degree of correlation between the individual forecasts.

Exercise I: zero correlation between predictor variables We first give some analytical results that may help the analysis. With the parameter values from Table 2.1, it is easy to derive that

$$y_{s,i} = 2 + \epsilon^*_{s,i} \qquad with\ \ \epsilon^*_{s,i} \sim iid\, N(0,3)$$

with $i = 1, 2$. Then,

$$y_s = 0.7y_{s,1} + 0.3y_{s,2} = 2 + 0.7\epsilon^*_{s,1} + 0.3\epsilon^*_{s,1} \tag{2.21}$$

Accordingly, the expected value of y_s is $E(y_s) = 2$ and its variance is $V(y_s) = 1.74$. We also notice that the coefficients of the variables $(x_{s,1}, x_{s,2})$ in the simulated DGP are $(\beta_1 c_{s,1}) = 0.7$ and $(\beta_2 c_{s,2}) = 0.3$ for any s.

By computing the probability limit of the OLS estimator $\widehat{\beta}_1$ in model (2.1) we find that $\widehat{\beta}_1$ is a consistent estimator of $(\beta_1 c_{s,1})$, its estimate is close to 0.7 for any $s = 181, .., 360$, and $\widehat{\beta}_2$ is a consistent estimator of $(\beta_2 c_{s,2})$, its estimate is close to 0.3 for any $s = 181, .., 360$. Moreover, both $(\widehat{\alpha_1}+\widehat{\beta}_1)$ and $(\widehat{\alpha_2}+\widehat{\beta}_2)$ add to 2 implying that the forecasts of the single models are unbiased, since $E(y_s) = 2$. In term of accuracy (MSPE), equation (2.1) does much better than equation (2.2), but the difference with the correct model, in which both (x_1, x_2) are included, is substantial. As the forecasts of both models are unbiased, the difference in accuracy is only due to the variance of the prediction errors. The variance of the prediction error of model (2.2) is more than double than that of the prediction error of model (2.1), reflecting the choice of (c_1, c_2)[6].

We find that the forecasts from the individual models and frequentist combination schemes can be approximated respectively as:

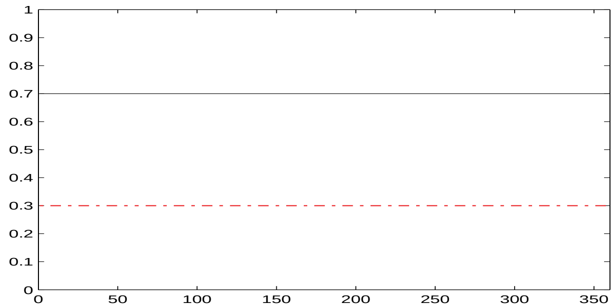
[6]We compute forecasts also by applying Bayesian inference (with diffuse priors). We do not report results because they are very similar to the previous ones.

Model 1	$\widehat{y}_{T+h,1} = 1.3 + 0.7x_{T+h,1}$
Model 2	$\widehat{y}_{T+h,1} = 1.7 + 0.3x_{T+h,2}$
True model	$\widehat{y}_{T+h} = 1 + 0.7x_{T+h,1} + 0.3x_{T+h,2}$
Given weights	$\widehat{y}^{(g)}_{T+h} = 1.42 + 0.49x_{T+h,1} + 0.09x_{T+h,2}$
Case 1	$\widehat{y}^{(1)}_{T+h} = 1.5 + 0.35x_{T+h,1} + 0.15x_{T+h,2}$
Case 2	$\widehat{y}^{(2)}_{T+h} = 1.42 + 0.49x_{T+h,1} + 0.09x_{T+h,2}$
Case 3	$\widehat{y}^{(3)}_{T+h} = 1 + 0.7x_{T+h,1} + 0.3x_{T+h,2}$
Case 4	$\widehat{y}^{(4)}_{T+h} = 1 + 0.7x_{T+h,1} + 0.3x_{T+h,2}$
Case 5	$\widehat{y}^{(5)}_{T+h} = 1 + 0.7x_{T+h,1} + 0.3x_{T+h,2}$

where $h = 1, .., 120$.
The estimators of β_1, β_2 are consistent for the products $(\beta_1 c_{s,1})$, $(\beta_2 c_{s,2})$. Therefore, estimating $(c_{s,1}, c_{s,2})$ both equal to 1 is the optimal solution to reduce the variance of the prediction errors. Combination schemes 3, 4 and 5 are the only methods to provide estimates of (c_1, c_2) equal to vectors of 1, providing the best statistics. Recursive and time-varying weight schemes, which allow for time varying estimates of (c_1, c_2), do not improve results compared to constant OLS weight scheme as (c_1, c_2) are time-invariant in the simulation. Other combination approaches (given weights, case 1 and 2) provide different estimates of (c_1, c_2), implying that the products $(\widehat{\beta}_1 \widehat{w}^j_1)$ and $(\widehat{\beta}_2 \widehat{w}^j_2)$ are not consistent estimator of $(\beta_1 c_1)$ and $(\beta_2 c_2)$. The forecasts given by those combination schemes are still unbiased but the variance of the prediction errors is higher. For example, assigning weights to single models based on the inverse of the MSPE well approximates the variance of the noises of the single models, $\epsilon^*_{s,1}$ and $\epsilon_s, 2^*$ respectively. Indeed, weight estimates of this scheme are very similar to the original values $c_1 = 0.7\iota$ where ι is a (120×1) vector of ones and $c_2 = 0.3\iota$ such as in the given weight combination. But this is not optimal in this exercise.
To sum up, model (2.1) predicts the part of the DGP related to $x_{s,1}$, model (2.2) predicts the part of the DGP related to $x_{s,2}$. Therefore, the optimal averaging strategy is adding with weight 1 the forecasts of the individual models and inserting a constant term to avoid biases. As Table 2.2 confirms, both the OLS weights and Terui and van Dijk (2002)'s time varying extension model this providing very accurate forecasts.

The Bayesian averaging scheme using marginal likelihood requires a different expla-

Figure 2.1: Exercise I (1)



Note: The figure presents the patterns of parameters c_1 (in solid line) and c_2 (in dotted line) in equation (2.3) in exercises I.

nation. What is important in Bayesian averaging is assigning the right probability to individual models. BMA based on marginal likelihood does not do this job well: it gives almost all the probability to model (2.1) and zero probability to model (2.2). The problem apparently relates to the use of un-normalized marginal likelihoods. To derive the marginal likelihood given by the individual models we compute the log marginal likelihood. Figure 2.2 plots the average of the log marginal likelihood for the two individual models for $s = 181, .., 360$ over the 1000 simulations. When we take the exponential to compute posterior weights the two numbers are not anymore comparable. And since (2.1) has higher log marginal likelihood all the probability is given to it. We note that more sophisticated ways of computing marginal likelihoods may exist, but we do not pursue this further. Instead we present a group of "simple to compute" Bayesian schemes under scheme eight.

BMA based on predictive likelihood gives on average probabilities similar to the original values 0.7 and 0.3. But its performance is not up to the level of estimated weight schemes. Bayesian results depend on the priors that we apply. We assume diffuse proper priors for model parameters, which imply parameter posterior means around OLS estimates (for derivation see, e.g., Koop, 2003, p. 37). The priors for $(\alpha_1, \alpha_2, \beta_1, \beta_2)$, however, could be chosen very informative around the true values 1. Then, averaging models (2.1) and (2.2) with predictive likelihoods would provide forecasts very similar to the correct model[7]. We think that it is in practice not easy to find such accurate priors and not all agents may agree on these precise priors, therefore we have applied diffuse priors that allow direct comparison to frequentist inference, but these diffuse priors apparently reduce forecast accuracy.

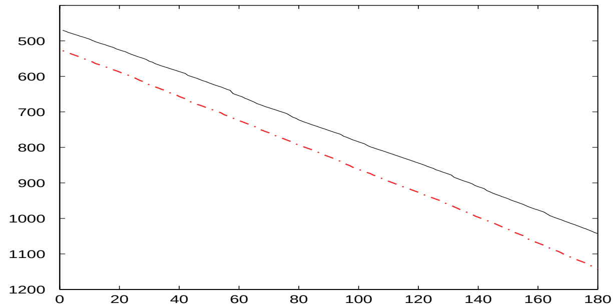
[7] Results for this exercise are available upon request.

Table 2.2: Results of simulation exercises

Exercises		Individual Models 1	2	3	4	5	Correct	Schemes Given	1	2	3	4	5	6	7	8
I-III: Varying correlations between predictors																
	MSPE	0.77	1.57	-	-	-	0.58	0.76	0.88	0.78	0.62	**0.61**	**0.61**	0.76	0.79	**0.61**
0.00	$BIAS^2$	0.01	0.01	-	-	-	0.00	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.00
	VAR	0.76	1.56	-	-	-	0.58	0.75	0.87	0.77	0.60	0.60	0.61	0.75	0.78	0.61
	MSPE	0.72	1.32	-	-	-	0.58	0.65	0.72	0.66	0.62	**0.60**	**0.60**	0.72	0.71	0.61
0.50	$BIAS^2$	0.01	0.01	-	-	-	0.00	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.00
	VAR	0.71	1.31	-	-	-	0.58	0.64	0.71	0.65	0.60	0.59	0.60	0.71	0.70	0.61
	MSPE	0.62	0.77	-	-	-	0.58	**0.58**	0.60	0.60	0.62	0.60	0.60	0.63	0.62	0.60
0.90	$BIAS^2$	0.01	0.01	-	-	-	0.00	0.00	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.00
	VAR	0.61	0.76	-	-	-	0.58	0.58	0.59	0.59	0.60	0.59	0.60	0.62	0.61	0.60
IV-VII: Misspecification																
	MSPE	0.76	1.57	-	-	-	0.58	0.76	0.88	0.77	0.63	**0.61**	**0.61**	0.76	0.79	**0.61**
Irrelevant variable	$BIAS^2$	0.01	0.01	-	-	-	0.00	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.00
	VAR	0.75	1.56	-	-	-	0.58	0.75	0.87	0.76	0.61	0.60	0.61	0.75	0.78	0.61
	MSPE	1.78	2.59	-	-	-	0.58	1.78	1.89	1.83	1.70	**1.66**	1.67	1.85	1.79	**1.66**
Omiited variable	$BIAS^2$	0.02	0.02	-	-	-	0.00	0.02	0.02	0.02	0.05	0.02	0.00	0.07	0.07	0.00
	VAR	1.76	2.57	-	-	-	0.58	1.76	1.87	1.81	1.65	1.64	1.67	1.78	1.72	1.66
	MSPE	0.67	1.21	1.36	-	-	0.46	0.69	0.87	0.76	0.52	0.53	**0.50**	0.66	0.71	**0.50**
3	$BIAS^2$	0.01	0.01	0.01	-	-	0.00	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.00
	VAR	0.66	1.20	1.35	-	-	0.46	0.68	0.86	0.75	0.50	0.52	0.50	0.66	0.70	0.50
	MSPE	0.57	0.89	1.05	1.07	1.07	0.36	0.59	0.81	0.74	0.44	0.44	**0.40**	0.57	0.61	**0.40**
5	$BIAS^2$	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.02	0.01	0.00	0.00	0.00	0.00
	VAR	0.57	0.88	1.04	1.06	1.06	0.36	0.59	0.81	0.73	0.42	0.43	0.40	0.57	0.61	0.40
VIII-IX: Structural changes																
	MSPE	1.78	0.99	-	-	-	1.01	1.00	1.11	1.08	1.06	0.76	0.74	1.70	1.01	**0.73**
1	$BIAS^2$	0.01	0.01	-	-	-	0.01	0.01	0.01	0.01	0.02	0.01	0.00	0.02	0.01	0.00
	VAR	1.77	0.98	-	-	-	1.00	0.99	1.10	1.07	1.04	0.75	0.74	1.68	1.00	0.73
	MSPE	0.86	1.26	-	-	-	0.55	0.77	0.81	0.81	0.96	0.68	0.60	0.93	0.80	**0.59**
2	$BIAS^2$	0.01	0.01	-	-	-	0.01	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.00
	VAR	0.85	1.25	-	-	-	0.54	0.76	0.80	0.80	0.94	0.67	0.60	0.93	0.80	0.59
X: Fat tails																
	MSPE	1.38	1.79	-	-	-	1.08	1.29	1.33	1.33	1.53	1.23	1.15	1.47	1.31	**1.13**
	$BIAS^2$	0.01	0.01	-	-	-	0.01	0.01	0.01	0.01	0.03	0.01	0.00	0.01	0.01	0.00
	VAR	1.37	1.77	-	-	-	1.07	1.28	1.32	1.32	1.50	1.22	1.15	1.46	1.30	1.13

Note: The table presents the mean square prediction error (MSPE), the square bias ($BIAS^2$), and the variance of the prediction errors (VAR) for the individual models and combination schemes, given in section 2, in exercises I-X.

Figure 2.2: Exercise I (2)



Note: The figure presents the log marginal likelihood given model 1 (in solid line) and the log marginal likelihood given model 2 (in dotted line) in exercise I.

The use of diffuse priors does not reduce the forecast accuracy of scheme 8 compared to that of schemes 3-5. In scheme 8 a Gibbs sampling procedure is applied to combine predictive densities of individual models. This Gibbs procedure is a Bayesian extension of scheme five. Results may be even more accurate when informative priors are applied.

Exercise II: medium correlation In the second exercise the correlation of the individual forecasts is increased and a medium positive (0.5) correlation between $x_{s,1}$ and $x_{s,2}$ is assumed for any s.

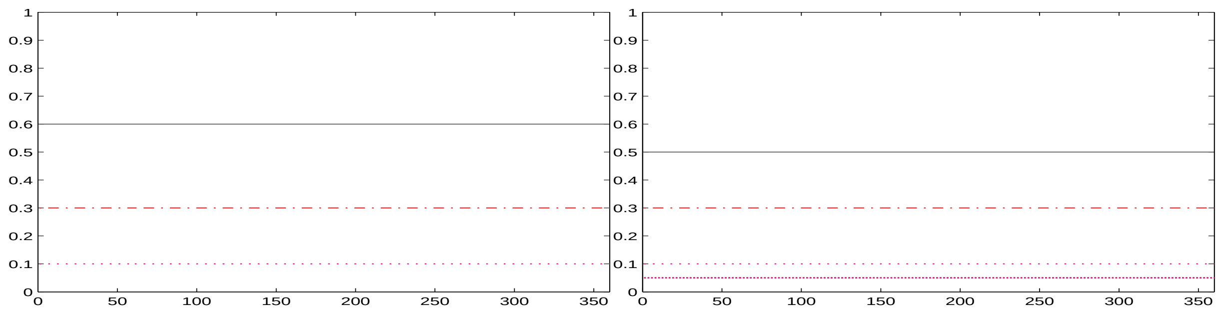
Model (2.1) performs better than model (2.2) due to the magnitude of the weights. Estimated weight schemes and Bayesian time varying weight scheme provide again better statistics than other averaging schemes, with results very similar to the correct model. However, simple combination schemes and BMA based on the predictive likelihood also give quite accurate forecasts. In this exercise model (2.1) and model (2.2) do not provide consistent estimate of $(\beta_1 c_{s,1})$ and $(\beta_2 c_{s,2})$, therefore weight estimates achieve this result. BMA based on marginal likelihood still selects only model (2.1).

Exercise III: high correlation In this exercise, the correlation of the individual forecasts is substantially increased (around 0.9). As in Timmermann (2006) in this framework equal weights are an appropriate choice. All the schemes forecast accurately and very similar to the correct model, since the individual models (2.1) and (2.2) give accurate and highly correlated results. Note that the time varying weight combinations are robust in this case.

2.3.2 Misspecification

In Exercises IV-VII the number of predictors and individual models varies. The DGP is still assumed stationary.

Figure 2.3: Exercise VI-VII



Note: The figures present in the panel a) the patterns of parameters c_1 (- line), c_2 (-. line), c_3 (.. line) in equation (2.3) in exercises VI, and in the panel b) also the parameters c_4 and c_5 (-- line) in exercise VII.

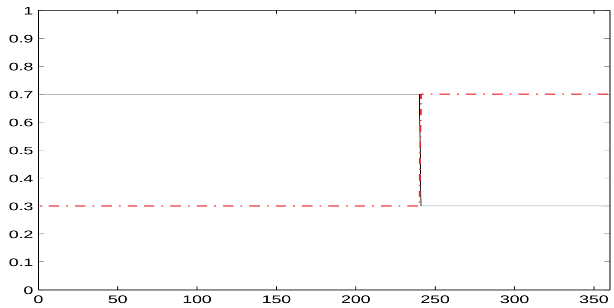
Exercise IV: included irrelevant variable In exercise IV an irrelevant variable (x_6) is included as additional regressor in model (2.1); its coefficient β_6 is given in Table 2.1. Due to the long series and the number of repetitions of the simulations β_6 is correctly estimated to be zero and results are very similar to exercise I.

Exercise V: omitted relevant variable In exercise V, a new variable, $x_{s,6}$, is added in the simulation of the DGP in equation (2.3). This variable is excluded in both models (2.1) and (2.2). All the forecasts are less accurate than in exercise I and the difference with the forecasts of correct model is substantial. However, estimated weight and Bayesian time varying weight schemes still give better statistics than individual models and other combination schemes. Results given by simple schemes and BMA schemes 6 and 7 are marginally worse than those of model (2.1).

Exercise VI-VII: 3 and 5 individual models The analysis is extended to include three and five individual models in the simulation exercise. Individual series are combined with weights given in Figure 2.3. In exercise VI $c_4 = \{c_{s,4}\}_{s=1}^{360}$ and $c_5 = \{c_{s,5}\}_{s=1}^{360}$ are vector of zeros.

In both examples, the estimated weight and Bayesian time varying weight schemes give the best forecasts. These schemes provide forecasts very similar to the correct model, and are the only ones to outperform the best individual model. Simple combination schemes do perform worse than the best individual model and Bayesian model averaging using marginal likelihoods. In the exercises where the misspecification of individual models is more substantial, allowing for parameter uncertainty is beneficial, even if parameter priors are not precise.

Figure 2.4: Exercise VIII



Note: The figure presents the patterns of parameters c_1 (in solid line) and c_2 (in dotted line) in equation (2.3) in exercises VIII.

2.3.3 Structural change

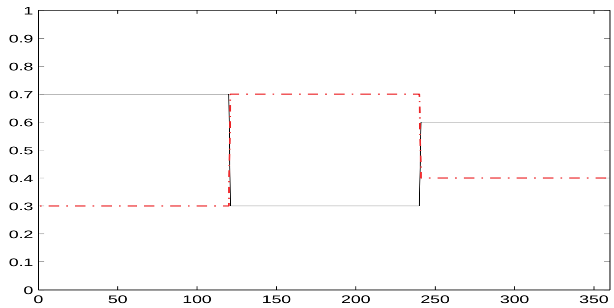
In the following two exercises, VIII and IX, the vectors c_1 and c_2 in equation (2.3) are subject to instability. For exercise VIII, Figure 2.4 shows that a shift happens at the beginning of the out-of-sample period. The weights assigned to models (2.1) and (2.2) are exactly reversed. In exercise IX, two shifts are plotted in Figure 2.5, at different times, with one of them in the in-sample period, and of opposite direction.

Exercise VIII: one shift The recursive OLS weight and (Bayesian) time-varying weight schemes dramatically outperform individual models, other combination schemes, and the correct model. The weight estimates of these three schemes capture the signal of instability, and react faster to it, partially reducing the inefficiency of parameter estimates of the individual models, which do not allow for instability in estimation. Rejecting instability may cause serious mistakes and, indeed, the correct model[8] gives marginally worse statistics than model (2.2). However, the instability, and therefore its signal, is quite moderate due to the fact that we have a unique break over the full sample. As Appendix A shows, this explains why recursive OLS and the Kalman Filter produce very similar weight estimates. Bayesian time varying weigh scheme 8 produces results very similar or marginally superior to scheme 5 again due to the use of diffuse priors.
BMA with predictive likelihood now provides quite accurate forecasts, even though it gives too high probability to model (2.2). BMA with marginal likelihood does not seem adequate even in this exercise. It assigns all the weight to model (2.1).

Exercise IX: two shifts The correct model gives the most accurate forecasts. The second shift partially correct the first one and moves the weight patterns close to their

[8]We remember that the “correct” model does not account for instability.

Figure 2.5: Exercise IX



Note: The figure presents the patterns of parameters c_1 (in solid line) and c_2 (in dotted line) in equation (2.3) in exercises IX.

in-sample average value. The time varying weight schemes provides the lowest statistics comparing to individual models and other averaging schemes. The instability is higher than in exercise VIII therefore the difference between the recursive OLS and the Kalman filter is evident, following the derivations in Appendix A. Simple combination schemes provide less accurate results. BMA based on predictive likelihoods copes with instability quite efficiently, but the diffuse type of priors chosen for individual model parameters reduce the forecast accuracy. Interestingly, the other BMA method initially assigns positive probability to both models, but when the number of observation increases, it converges to assign all the weight to model (2.1).

2.3.4 Fat tails

The DGP from exercise IX is changed by assuming fat tailed errors. In particular, $\epsilon_{s,1}$ and $\epsilon_{s,2}$ in (2.1)-(2.2) are assumed to be Student t distributed with mean, variance and ν degree of freedom in Table 2.1. The DGP weights are still as in Figure 2.5. All forecasts are less accurate than in exercise IX, but the results are qualitatively similar to the previous example. Again, the time varying weight schemes provide the lowest statistics among the averaging schemes and provides results very close to the correct model. Adding parameter uncertainty seems beneficial as scheme 8 gives marginally superior results that scheme 5. As in the previous case, several averaging schemes give more accurate forecasts than individual models, confirming that averaging is in our set up of experiments a simple and attractive way to cope with instability.

2.3.5 Summary of findings

The results in Table 2.2 indicate that it is not easy to find a general rule how to average individual models in an optimal way, and elements as the degree of correlation of the individual forecasts, data predictability, structural instability and model (mis)specification, play a strategic role in the process of combining forecasts of individual models. In particular, we find that in situations of low predictability and high noise, and almost no correlation of a limited set of individual forecasts, combination schemes that estimate model weights and their extension in a Bayesian framework give the most accurate forecasts. Intuitively, when individual forecasts contain complementary information, the best averaging strategy is to add this independent information. Simple combination schemes are not adequate schemes as they average individual models instead of adding with weight 1 the independent information of different models. Bayesian model averaging based on marginal likelihood has some computational problems due to the fact of deriving unnormalized marginal likelihoods for a relative small set of individual models. Bayesian model averaging based on predictive likelihood assigns precise weights to individual models, but using diffuse priors in model parameters as we do reduce the forecast accuracy.

If the DGP is also subject to structural instability, in the sense that the relevance of the predictors varies over time, time varying weight schemes give the highest predictive gains. Simple combination schemes and recursive OLS weight schemes do not learn (efficiently) from the signals of instability, and therefore do not react fast to it. Bayesian model averaging based on predictive likelihood copes better with instability, but inadequate priors can reduce forecast accuracy. Results are qualitative similar when the distribution has fatter tails than the standard normal case, and adding more sources of uncertainty as the Bayesian time varying weight scheme does seems to be beneficial.

2.4 Empirical illustration

We extend our study by investigating the forecasting performance and economic gains obtained by applying the eight forecast combination schemes to the case of US stock index returns, defined as the discretely compounded monthly return on the S&P 500 index in excess of the 1-month T-Bill rate, from January 1976 to December 2005, for a total of 360 observations; see Figure 2.6. We use two linear non-nested forecasting models. The first one is based on the idea that a set of financial and macroeconomic variables are potentially relevant factors for forecasting stock returns. Among others, Pesaran and Timmermann (1995), Cremers (2002), Marquering and Verbeek (2004) have shown that such variables have predictive power. We label this forecasting model "Leading factor" (LF). The second forecasting model is a simple linear regression model with a constant and a dummy for November-April. It is based on the popular market saying "Sell in May and go away", also known as the "Halloween indicator" (HI), and it based on the assumption that stock returns can be predicted simply by deterministic time patterns.

This suggests to buy stock in November and sell it in May. Bouman and Jacobsen (2002) show that this strategy has predictive power.

2.4.1 Data and evaluation

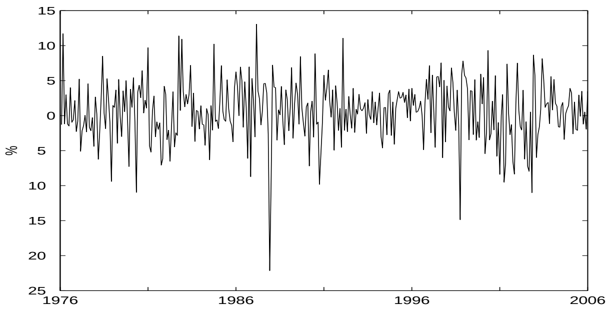
The source of the S&P 500 index is the CRSP database and the 1-month T-Bill rate is from Ibbotson and Associates. We include as predictors the S&P 500 index price-earnings ratio (PE), the S&P 500 index dividend yield (DY) defined as the ratio of dividends over the previous twelve months and the current stock price, the 3-month T-Bill rate ($I3$), the monthly change in the 3-month T-bill rate ($DI3$), the term spread (TS) defined as the difference between the 10-year T-bond rate and the 3-month T-bill rate, the credit spread (CS) defined as the difference between Moody's Baa and Aaa yields, the yield spread (YS) defined as the difference between the Federal funds rate and the 3-month T-bill rate, the annual inflation rate based on the producer price index (PPI) for finished goods (INF), the annual growth rate of industrial production (IP), the annual growth rate of the monetary base (MB), and the log monthly realized volatility of the S&P 500 index ($LVol$). The monthly realized volatility is computed using daily returns, where we follow French *et al.* (1987) and Marquering and Verbeek (2004) by assuming that daily returns are appropriately described by a first-order autoregressive process. In particular, we use the following estimate for realized volatility

$$\hat{\sigma}_t^2 = \sum_{t=1}^{N_s} (y_{i,t} - \bar{y}_t)^2 \left[1 + \frac{2}{N_t} \sum_{j=1}^{N_t - 1} (N_t - j)\hat{\phi}_t^j, \right]$$

where $y_{i,t}$ is the return on day i in month t which has N_t trading days, $\bar{y}_t$ is the average daily return in month t, and $\hat{\phi}_t$ denotes the first-order autocorrelation estimated using daily returns within month t. We take into account the typical publication lag of macroeconomic variables in order to avoid look-ahead bias. We therefore include inflation and the growth rates of industrial production and the monetary base with a two-month lag. As the financial variables are promptly available, these are included with a one-month lag. Finally, the "Halloween indicator" (HI) model is specified as a simple linear regression with a constant and a dummy for November-April.

We evaluate the statistical accuracy of the individual models and the eight forecast combinations schemes in terms of MSPE, and its decomposition in square bias and variance of the forecast errors. Again means of Bayesian predictive densities are computed for the BMA schemes. Moreover, as an investor is more interested in the economic value of a forecasting model than its precision, we test our conclusions in an active short-term investment exercise, with an investment horizon of one month. The investor's portfolio

Figure 2.6: S&P500 Excess returns



Note: The figure presents the excess returns on the S&P500 over the sample 1976:1-2005:12.

consists of a stock index and riskfree bonds only. At the start of month $T+1$, the investor decides upon the fraction of her portfolio to be invested in stocks $w_{p,T+1}$, based upon a forecast of the excess stock return y_{T+1}. The investor is assumed to maximize a mean-variance utility function

$$\max_{w_{T+1}} u(E_T(y_{p,T+1}), Var_T(y_{p,T+1})) \tag{2.22}$$

where $y_{p,T+1}$ is the return of the investor's portfolio return at time $T+1$, which is equal to

$$y_{p,T+1} = W_T((1 - w_{p,T+1})(y_{f,T+1}) + w_{p,T+1}(y_{f,T+1} + y_{T+1})) \tag{2.23}$$

where W_T denotes the wealth at time T, where y_{T+1} is the excess returns on S&P500, and where $y_{f,T+1}$ is the riskfree rate.

Without loss of generality we set initial wealth equal to one, $W_T = 1$. Further, we assume the following utility function:

$$E_T(y_{p,T+1}) - \frac{1}{2}\gamma Var_T(y_{p,T+1}) \tag{2.24}$$

where γ is the coefficient of relative risk aversion. Solving the maximization problem shows that the optimal portfolio weight for the investor is given by:

$$w^*_{p,T+1} = \frac{E_T(y_{T+1}) - r_{y,T+1}}{\gamma Var_T(y_{T+1})}. \tag{2.25}$$

If the expected excess return on the risky asset increases, it is optimal for the investor to increase her weight on the risky asset. The conditional variance $Var_T(y_{T+1})$, which represents a measure of the risk involved, is negatively related to this weight. We forecast $E_T(y_{T+1})$ with nine different approaches: two individual models, the 'leading factor' one (LF), and the 'Halloween indicator' one (HI), and the eight averaging schemes discussed in this Chapter. Each individual forecasting approach corresponds to an investment strategy which is defined in the same way. We approximate the conditional variance with the 60-month moving window average of the realized variances computed as above[9]. We also assume that short selling and borrowing at the riskfree rate are not allowed, therefore we restrict the portfolio weights to be between 0 and 1. For purposes of comparison we consider a passive investment strategy where the total wealth is invested in the risky market (RW).

We evaluate the different investment strategies by computing the average return, the standard deviation of the portfolio return, and the Sharpe ratio, defined as the ratio of the mean excess return on the (managed) portfolio and the standard deviation of the portfolio return. Since the Sharpe ratio overestimates risk in case of time varying volatility, we also compute the *ex post* utility levels - in order to estimate the economic value of the strategy - by substituting the realized return of the portfolios at time $T+1$ in (3.18)

$$U^*_{p,T+1} = y_{p,T+1} - \frac{1}{2}\gamma w^2_{p,T+1} Vol_{T+1} \tag{2.26}$$

where Vol_{T+1} denotes the ex post realized volatility of the risky return on month $T+1$. Total utility is then obtained as the sum of U^*_p across all H investment periods. The above approach enables us to compare alternative investment strategies by calculating the associated average utility levels.

Finally, as the portfolio weights in the active investment strategies change every month, the portfolio must be rebalanced accordingly. Hence, transaction costs play a non-trivial role and should be taken into account when evaluating the relative performance of different strategies. Rebalancing the portfolio at the start of month $T+1$ means that the weight invested in the risky asset is changed from w_T to w_{T+1}. We assume that transaction costs amount to a fixed percentage c on each traded dollar. Setting the initial wealth W_T equal to 1 for simplicity, transaction costs at time $T+1$ are defined as equal to

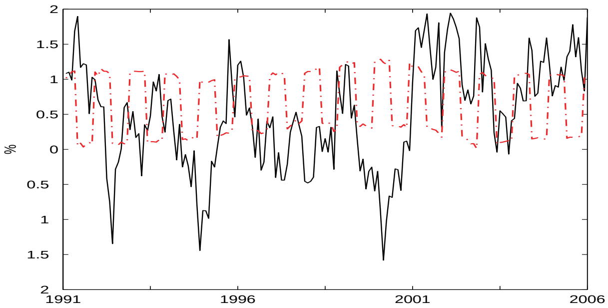
$$c_{T+1} = 2c|w_{T+1} - w_T| \tag{2.27}$$

where the multiplication by 2 follows from the fact that the investor rebalances her investments in both stocks and bonds. The net portfolio return is then given by $r_{T+1} - c_{T+1}$. We apply three scenarios with transaction costs of 0.1%, 0.5% and 1%[10]. Note that for the passive investment strategy where the total wealth is invested in the risky market the inclusion of transaction costs matters only in setting up the portfolio at time T_0.

[9] We also forecast the conditional variance $Var_T(y_{T+1})$ using an AR(1), an AR(12), an Heterogeneous Autoregressive (HAR) model similar to Corsi (2004), and an EGARCH model as in Marquering and Verbeek (2004). Results are qualitative similar. We prefer the 60-month moving window average because most investors use similar simple schemes, in particular at beginning of our sample period.

[10] We think that 10 basis points is an average transaction cost to buy a 1-month future on S&P500 or a 1-month future on 1-month Treasury Bill.

Figure 2.7: Individual forecasts



Note: The figure presents the forecasts on excess returns on the S&P500 given by the individual models 'Leading Indicator' (in solid line) and 'Halloween indicator' (in dotted line) over the sample 1996:1-2005:12.

2.4.2 Empirical Results

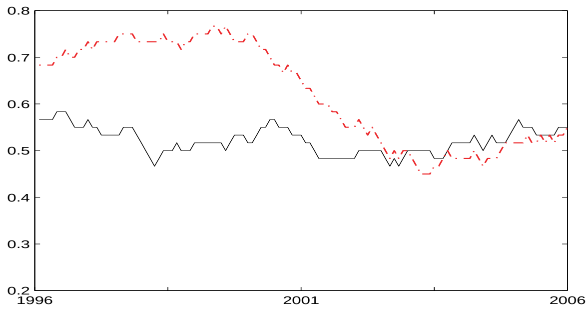
The analysis for the active investment strategies is implemented for the period from January 1996 until December 2005, involving 120 one month ahead excess stock return forecasts. The models are estimated recursively using an expanding window of observations. The period January 1991 to December 1995 is used to start up the forecast combination schemes. The investment strategies are implemented for three levels of relative risk aversion, $\gamma = 2, 5$ and 10. Before we analyze the performance of the different portfolios, we summarize the statistical accuracy of the excess return forecasts.

Statistical accuracy

The statistical accuracy of the individual models and forecast combination is evaluated by MSPE, and its decomposition in square bias and variance as in Section 2.3. Results are reported in Table 2.3. In the market column, labelled RW, we report the statistics of the Random Walk model.

We notice that both the individual models provide much lower evaluation criteria than the RW. In particular, the Halloween Indicator model has the lowest MSPE error and both the mean and the variance of the forecast errors are lower than for the other individual models. However, both series of forecasts have a quite different pattern than the very noise excess return series in Figure 2.7. The HI model has a seasonal pattern given by the particular strategy with a positive unconditional mean, and few negative forecasts only in 2002. The LF generates forecasts which are more volatile, and in particular too low at the end of 1990's and at beginning of 2000, and too high in 2001. In term of sign prediction the HI strategy performs very well in 90's. The 60 month moving average sign hit ratios, which are the proportions of correctly predicted signs of the excess return over the previous 60 months, shown in Figure 2.8, are higher than 0.7 and close to 0.8. But

Figure 2.8: 60 month moving average sign hit ratios



Note: The figure presents the 60 month moving average sign hit ratios given by the individual models 'Leading Indicator'(in solid line) and 'Halloween indicator' (in dotted line).

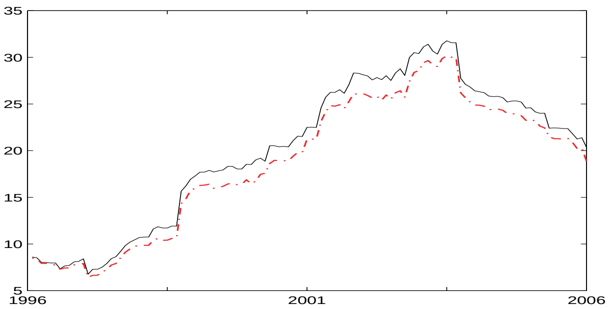
after 1998, the ratios start to deteriorate and stabilize at hit ratios around 0.5 for the final years of the sample period. The higher percentage of positive returns in 90's, and the almost always positive forecasts given by model HI may explain the result. The hit ratios given by the LF model are more stable and on average just above 0.5. In term of MSPE, Figure 2.9 show similar predictive patterns of the set of forecasts, but after middle of 1996 the HI model always provides lower mean square errors than the LF model.

When averaging schemes are applied, the results are intriguing; see the top of Table 3 for details. The MSPEs of schemes 1, 2, 3, 4, 6 are all higher than that of model HI. Moreover, constant OLS and recursive OLS schemes have a positive bias[11]. The time varying weight schemes, however, provide the best statistics. If we investigate the weight estimates, we find that there is an indication of a break in the weight for model HI in the training period at year 1995, moving from a lower value to values very stable around 1. At the same time, the weight on model LF decreases and stabilizes around -0.5. This confirms ex-post instability evidence in Figure 2.9 that model HI provides more accurate forecasts than the alternative model after 1996. The dramatic boom of stock prices at the end of 90's and well documented lower predictability of macroeconomic and financial indicators can explain this result. It may also indicate that strategy HI captures some seasonal stylized facts of stock index returns and assigning weight 1 to it is beneficial in term of forecasting performance.

The BMA with predictive likelihood also gives a marginal lower MSPE than the individual model HI. These results are similar to the ones from exercise IX, which shows that the BMA scheme 7 copes with possible instability better than simple combination schemes. Summarizing, the forecast statistics of the combination schemes are rather similar; the largest difference between schemes is less than 5%. However, because predictability of

[11]We emphasize that their bias is insignificant with respect to the MSPE, and it is less than 0.2% of the unconditional mean return.

Figure 2.9: 60 month moving average MSPEs



Note: The figure presents the 60 month moving average MSPE given by the individual models 'Leading Indicator'(in solid line) and 'Halloween indicator' (in dotted line).

stock market is very low, small improvements in MSPE may have substantial economic value. To investigate this we implement a portfolio exercise, reported n the next section.

Summarizing, the forecast statistics of the combination schemes are rather similar; the largest difference between schemes is less than 5%. Moreover, because predictability of stock market is very low, small improvements in MSPE may have substantial economic value. To investigate this we implement a portfolio exercise, reported n the next section.

Economic value

Panel B of Table 2.3 provides performance measures for the different investment strategies based on the ten different forecasting methods presented in the previous sections. Over the forecasting period, January 1996 to December 2005, the average return on the stock portfolio is 10%, the standard deviation is 16%, and the Sharpe ratio is 0.12. The strategies based on forecasting returns with one of the two individual models give lower mean returns for a moderately risk averse ($\gamma = 5$) investor, but also lower standard deviation, which results in a higher Sharpe ratio for the Halloween strategy. Accounting for possible time varying volatility, and evaluating strategies with the *ex-post* realized utility shows that the Halloween indicator performs better than the leading indicator and the market. The leading factor strategy gives very low mean portfolio returns, which implies a low Sharpe ratio and utility level.

Next, consider the strategies based on forecasting excess returns with the eight averaging schemes. Strategy 5 and 8, based on time varying model weights, give the highest mean returns among all the active strategies, among the lowest standard deviations, and the highest Sharpe ratios and utility levels. In particular, the Bayesian time varying weight scheme has marginally higher mean return but also standard deviation. Strategy 7, based on BMA with predictive likelihood, provides also marginally superior results in terms of portfolio measures than the strategy HI, but substantially lower than the pre-

Table 2.3: Empirical application - No transaction costs

Statistics	Individual Models RW	LF	HI	Strategies 1	2	3	4	5	6	7	8
				Panel A: Statistical accuracy							
MSPE	40.96	21.45	20.04	20.55	20.53	20.96	20.73	19.84	20.33	19.99	**19.83**
$BIAS^2$	0.00	0.00	0.02	0.01	0.01	0.36	0.10	0.01	0.03	0.03	0.01
VAR	40.96	21.45	20.02	20.54	20.52	20.61	20.63	19.84	20.30	19.97	19.82
				Panel B: Economic value							
Criteria				$\gamma = 2$							
Mean	9.94	5.24	9.09	6.59	6.60	10.09	9.24	11.04	9.51	9.84	11.09
St dev	15.63	12.82	13.69	13.93	14.01	15.61	14.48	13.43	14.00	14.41	13.46
SR	0.12	0.04	0.12	0.06	0.06	0.12	0.11	0.16	0.12	0.13	0.16
Utility	0.77	0.35	0.77	0.48	0.48	0.79	0.76	**0.97**	0.76	0.80	**0.97**
				$\gamma = 5$							
Mean	9.94	4.51	8.94	6.45	6.51	8.24	8.12	9.76	8.60	9.57	9.87
St dev	15.63	10.16	11.01	10.21	10.22	14.41	12.89	10.35	10.83	11.03	10.52
SR	0.12	0.03	0.14	0.08	0.08	0.09	0.10	0.17	0.13	0.16	0.17
Utility	0.44	0.15	0.65	0.39	0.39	0.35	0.59	0.82	0.54	0.68	**0.83**
				$\gamma = 10$							
Mean	9.94	3.99	7.16	5.07	5.14	6.21	6.23	8.05	6.82	6.98	8.15
St dev	15.63	7.30	6.91	6.12	6.11	9.75	9.25	7.21	7.37	6.35	7.31
SR	0.12	0.02	0.15	0.07	0.07	0.08	0.08	0.18	0.13	0.15	0.18
Utility	-0.11	0.15	0.56	0.34	0.35	0.25	0.44	**0.69**	0.43	0.53	**0.69**

Note: The table presents in panel A the mean square prediction error (MSPE), the square bias ($BIAS^2$), and the variance of the prediction errors (VAR) for the individual models and combination schemes in forecasting the S&P500 index; in panel B the average portfolio return and standard deviation (both in percentage points), the Sharpe ratio (SR), and utility for various level of risk aversion coefficients, γ.

Table 2.4: Empirical application - Transaction costs

	Individual Models			Strategies							
Statistics	RW	LF	HI	1	2	3	4	5	6	7	8
					$c = 0.1\%$						
Mean	9.93	4.10	8.62	6.01	6.06	8.06	7.86	9.38	8.16	9.15	9.51
St dev	15.63	10.12	10.98	10.18	10.20	14.40	12.89	10.34	10.81	11.02	10.50
SR	0.12	0.01	0.13	0.07	0.07	0.09	0.10	0.16	0.12	0.15	0.16
Utility	0.44	0.11	0.62	0.34	0.35	0.33	0.56	0.78	0.49	0.64	**0.79**
					$c = 0.5\%$						
Mean	9.89	2.44	7.37	4.26	4.30	7.34	6.83	7.87	6.43	7.47	8.03
St dev	15.62	10.01	10.90	10.11	10.13	14.36	12.90	10.30	10.76	10.98	10.46
SR	0.12	-0.03	0.10	0.02	0.02	0.08	0.07	0.12	0.08	0.10	0.12
Utility	0.44	-0.06	0.50	0.17	0.17	0.26	0.46	0.63	0.32	0.47	**0.64**
					$c = 1\%$						
Mean	9.84	0.38	5.79	2.07	2.09	6.43	5.54	5.98	4.26	5.36	6.19
St dev	15.62	9.92	10.85	10.04	10.07	14.33	12.92	10.29	10.74	10.97	10.46
SR	0.12	-0.09	0.06	-0.04	-0.04	0.06	0.04	0.07	0.02	0.05	0.07
Utility	0.43	-0.26	0.34	-0.05	-0.05	0.17	0.33	0.44	0.10	0.26	**0.46**

Note: The table presents the average portfolio return and standard deviation (both in percentage points), the Sharpe ratio (SR), and utility for various level of transaction costs c and coefficient of risk aversion $\gamma = 5$.

vious strategy. Again, more precise priors may be chosen, but we omit this "subjective" exercise. All other strategies have lower economic values, in particular, give lower mean portfolio returns. Results are qualitative similar for a risk seeking investor ($\gamma = 2$) and a risk averse investor ($\gamma = 10$). Moreover, adding transaction costs does not change the quality of the results, and even with substantial transaction costs of 100 basis points, strategies 5 and 8 give higher levels of utility compared to a random walk strategy of investment. We notice that their Sharpe ratios are lower, confirming that the Sharpe ratio may overestimates risk in case of time varying volatility.

To conclude, the results indicate than the individual models HI and LF provide different forecasts. Moreover, instability in the relation between realized excess returns and individual forecasts seems to be relevant. As in the simulation exercises, in the empirical example the time varying weight schemes give the highest predictive gains both in statistical measures and economic gains.

2.5 Conclusion

Investors often have a set of forecasts on asset returns available from different models. Such investors may attempt to discover which is the best forecasting model and use it to allocate their portfolios, or they may consider all forecasts and take decisions by averaging forecast information from the individual models. In this Chapter we explained in a simulation experiment that when data is subject to low predictability, low correlation among individual forecasts, and structural instability, the Terui and van Dijk (2002) time varying model weight scheme and its extension in a Bayesian framework to incorporate parameter uncertainty provides the most accurate forecasts compared to other frequentist and Bayesian model averaging (with diffuse priors on model parameters) schemes. We applied the different model averaging schemes also to forecast the index of US stock returns. As in the simulation exercise, stylized facts of stock index data are low predictability and possible structural instability. We considered two forecasting models that represent different views on predicting the US stock index. We have shown, firstly, that averaging strategies can give higher predictive gains than selecting the best model; secondly, that time varying model weights have higher statistical and economic values than other averaging schemes considered. An interesting topic for further research is to compare our results to other time varying weight combination schemes, such as regime switching, see e.g. Guidolin and Timmermann (2007), or schemes that carefully model breaks, see e.g. Ravazzolo *et al.* (2007a). Moreover, combination schemes can be applied to the analysis of density forecasts. Market operators, such as financial investors or central bank decision makers, are becoming increasingly interested in knowing the complete distribution of the assets of interests for purposes of risk management. The Bayesian time varying weight scheme that we put forward seems particular adequate in this context.

Appendix: Estimation details

2A Comparison of Recursive Least Squares and time varying model weight combinations

Recursive Least Squares model weights The model weights of the OLS averaging scheme 4 can be computed by Recursive Least Squares. Consider (2.10) and rewrite it as

$$y_t = z_t^{'} w^{(4)} + u_t; \quad u_t \sim N(0, s^2) \tag{2A.1}$$

where $z_t^{'}$ is a $(1 \times q)$ row vector and where $w^{(4)}$ is a $(q \times 1)$ vector of unknown constant parameters. The first step in the derivation of the recursive least square estimator $b_t^{(4)}$ of the weight $w^{(4)}$ is to specify the OLS estimators of $w^{(4)}$ using information up to t and $t-1$

$$b_t^{(4)} = (Z_t^{'} Z_t)^{-1} Z_t^{'} Y_t$$

$$b_{t-1}^{(4)} = (Z_{t-1}^{'} Z_{t-1})^{-1} Z_{t-1}^{'} Y_{t-1}$$

where $Z_t = (Z_{t-1}^{'}, z_t)^{'}$ with Z_{t-1} a $((t-1) \times q)$ matrix and $z_t^{'}$ a $(1 \times q)$ row vector, and where $Y_t = (Y_{t-1}^{'}, y_t)^{'}$ with Y_{t-1} a $((t-1) \times 1)$ vector and y_t a scalar. To compute $b_t^{(4)}$ and $b_{t-1}^{(4)}$ we require that $t \geq q$, and $(Z_{t-1}^{'} Z_{t-1})$ and $(Z_t^{'} Z_t)$ are non singular matrices. As next step to derive $b_t^{(4)}$ as function of $b_{t-1}^{(4)}$ and data information available at time $(t-1)$, we express $(Z_t^{'} Z_t)$ in terms of $(Z_{t-1}^{'} Z_{t-1})$, and $(Z_t^{'} Y_t)$ in terms of $(Z_{t-1}^{'} Y_{t-1})$. Consider first

$$Z_t^{'} Y_t = Z_{t-1}^{'} Y_{t-1} + z_t y_t$$

$$= (Z_{t-1}^{'} Z_{t-1}) b_{t-1}^{(4)} + z_t y_t + z_t z_t^{'} b_{t-1}^{(4)} - z_t z_t^{'} b_{t-1}^{(4)}$$

$$= (Z_{t-1}^{'} Z_{t-1} + z_t z_t^{'}) b_{t-1}^{(4)} + z_t (y_t - z_t^{'} b_{t-1}^{(4)})$$

$$= Z_t^{'} Z_t w_{t-1}^{(4)} + z_t (y_t - z_t^{'} b_{t-1}^{(4)})$$

Then,

$$b_t^{(4)} = (Z_t^{'} Z_t)^{-1} (Z_t^{'} Z_t b_{t-1}^{(4)} + z_t (y_t - z_t^{'} b_{t-1}^{(4)}))$$

$$= b_{t-1}^{(4)} + (Z_t^{'} Z_t)^{-1} z_t (y_t - z_t^{'} b_{t-1}^{(4)})$$

To express $(Z_t^{'} Z_t)$ in terms of $(Z_{t-1}^{'} Z_{t-1})$, make use of the matrix inverse lemma $(A + BC^{-1}D)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$, where A, B, C, D are matrices, and

A and C are non-singular, see e.g. Harvey (1993, section 4). Then one can obtain

$$(Z_t'Z_t)^{-1} = (Z_{t-1}'Z_{t-1} + z_t z_t')^{-1}$$

$$= (Z_{t-1}'Z_{t-1})^{-1} - (Z_{t-1}'Z_{t-1})^{-1} z_t (1 + z_t'(Z_{t-1}'Z_{t-1})^{-1} z_t)^{-1} z_t'(Z_{t-1}'Z_{t-1})^{-1}$$

$$= (Z_{t-1}'Z_{t-1})^{-1}(1 + z_t'(Z_{t-1}'Z_{t-1})^{-1} z_t)^{-1}$$

The recursive least squares estimator of the weight $w^{(4)}$ is

$$b_t^{(4)} = b_{t-1}^{(4)} + (Z_{t-1}'Z_{t-1})^{-1} z_t (z_t'(Z_{t-1}'Z_{t-1})^{-1} z_t + 1)^{-1}(y_t - z_t' b_{t-1}^{(4)}) \tag{2A.2}$$

$b_t^{(4)}$ is defined recursively as equal to its previous value plus a weighted value of the prediction error $(y_t - z_t' b_{t-1}^{(4)})$ times the observed value of z_t. A minimum of k observations are needed to compute a starting value for the estimator.

Time varying model weights The model weights of the time varying averaging scheme 5 are defined as

$$y_t = z_t' w_t^{(5)} + u_t; \quad u_t \sim N(0, s^2) \tag{2A.3}$$

$$w_t^{(5)} = w_{t-1}^{(5)} + \xi_t; \quad \xi_t \sim N(0, \Sigma) \tag{2A.4}$$

where $w_t^{(5)}$ is a $(q \times 1)$ vector of random variables, and u_t and ξ_t are independently and identical distributed for $t = 1, ..., T$, and uncorrelated for all lags, $E(\xi_t, u_\tau) = 0$ for all t and τ, $t \neq \tau$, and where Σ is a diagonal matrix.

We make use of the Kalman Filter technique to compute estimators for the model weights $w_t^{(5)}$. Rewrite (2A.3) as

$$y_t = z_t' b_{t|t-1}^{(5)} + z_t'(w_t^{(5)} - b_{t|t-1}^{(5)}) + u_t \tag{2A.5}$$

and use that

$$w_t^{(5)} = b_{t|t-1}^{(5)} + w_t^{(5)} - b_{t|t-1}^{(5)} \tag{2A.6}$$

where $b_{t|t-1}^{(5)}$ is defined as an unbiased predictor of the vector of latent factors $w_t^{(5)}$. Conditional on information at time $t-1$ and using (2A.5) and (2A.6), one has

$$E(y_t) = z_t' b_{t|t-1}^{(5)}$$

$$V(y_t) = z_t' P_{t|t-1} z_t + s^2$$

$$E(w_t^{(5)}) = b_{t|t-1}^{(5)}$$

$$V(w_t^{(5)}) = P_{t|t-1}$$

and

$$E(w_t^{(5)}, y_t) = P_{t|t-1} z_t$$

where $E()$ is the expected value operator and $V()$ is the variance operator, where (following Harvey, 1993) $b_{t|t-1}^{(5)} = b_{t-1}^{(5)}$ is the best predictor in MSPE sense for $w_t^{(5)}$ and $b_{t-1}^{(5)}$ is the unbiased estimator for $w_{t-1}^{(5)}$ known at time t, and where $P_{t|t-1} = (P_{t-1} + \Sigma)$ is the best predictor for the variance of $w_t^{(5)}$ and P_{t-1} is the variance of the estimator of $w_{t-1}^{(5)}$ known at time t.

Therefore, given the assumptions on the distribution of u_t and on $b_{t|t-1}^{(5)}$ the vector $[w_t^{(5)'}, y_t]'$ is conditional normally distributed given information up to time $(t-1)$ with mean and covariance matrix given by

$$\begin{bmatrix} b_{t|t-1}^{(5)} \\ \\ z_t' b_{t|t-1}^{(5)} \end{bmatrix} \quad \begin{bmatrix} P_{t|t-1} & P_{t|t-1} z_t \\ \\ z_t' P_{t|t-1} & z_t' P_{t|t-1} z_t + s^2 \end{bmatrix} \tag{2A.7}$$

It now follows from properties of the multivariate distribution, see Appendix 2.C, that the distribution of $w_t^{(5)}$ conditional on y_t is multivariate normal with mean

$$b_t^{(5)} = b_{t|t-1}^{(5)} + P_{t|t-1} z_t (z_t' P_{t|t-1} z_t + s^2)^{-1} (y_t - z_t' b_{t|t-1}^{(5)}) \tag{2A.8}$$

and covariance matrix

$$P_t = P_{t|t-1} - P_{t|t-1} z_t (z_t' P_{t|t-1} z_t + s^2)^{-1} z_t' P_{t|t-1} \tag{2A.9}$$

Thus $b_t^{(5)}$ is the vector of estimated model weights in (2A.3) defined equal to the predictor of the latent factor $w_t^{(5)}$ plus a term that is the weighted product of the prediction error $(y_t - z_t' b_{t|t-1}^{(5)})$, the observed value of z_t, and the prediction for the variance of the latent factor estimator $P_{t|t-1}$.

Comparison The estimated weights of the recursive OLS averaging scheme 4 is a special case of the estimated weights of the time varying averaging scheme 5 computed using the Kalman Filter technique. This can be seen directly: if Σ is a matrix of zeros, $s^2 = 1$, and the updating equation of the Kalman Filter in (2A.9) is initialized as $P_k = (Z_k' Z_k)^{-1}$, the weight estimates in (2A.8) and (2A.2) are identical.
Let fix $P_k = (Z_k' Z_k)^{-1}$. Following (2A.8), the weight estimates at time $(k+1)$ given by the Kalman Filter have mean

$$b_{k+1}^{(5)} = b_k^{(5)} + ((Z_k' Z_k)^{-1} + \Sigma) z_{k+1} (z_{k+1}' ((Z_k' Z_k)^{-1} + \Sigma) z_{k+1} + s^2)^{-1} (y_{k+1} - z_{k+1}' b_k^{(5)}) \tag{2A.10}$$

where $b^{(5)}_{k+1|k} = b^{(5)}_k$, and where $(P_{k+1|k} = (Z'_k Z_k)^{-1} + \Sigma)$. s^2 is a scaling parameter. From (2A.3) $V(y_{k+1}) = (z'_{k+1}((Z'_k Z_k)^{-1} + \Sigma)z_{k+1} + s^2)$, and by scaling it with s^2, one has

$$V(y^*) = \left(z'_{k+1}\left(\frac{(Z'_k Z_k)^{-1}}{s^2} + \frac{\Sigma}{s^2}\right) z_{k+1} + 1\right)$$

$$E(w^{(5)}_{k+1}, y^*) = \frac{((Z'_k Z_k)^{-1}+\Sigma)z_k}{s}$$

where $y^* = y_{k+1}/s$. Therefore, $b^{(5)}_{k+1}$ can be written as:

$$b^{(5)}_{k+1} = b^{(5)}_k + \left(\frac{(Z'_k Z_k)^{-1}}{s^2} + \frac{\Sigma}{s^2}\right) z_{k+1}\left(z'_{k+1}\left(\frac{(Z'_k Z_k)^{-1}}{s^2} + \frac{\Sigma}{s^2}\right) z_{k+1} + 1\right)^{-1} (y_{k+1} - z'_{k+1} b^{(5)}_k) \tag{2A.11}$$

The recursive least square estimator of $w^{(4)}$ at time $k+1$ is given in (2A.2) and repeated for convenience as

$$b^{(4)}_{k+1} = b^{(4)}_k + (Z'_k Z_k)^{-1} z_{k+1}(z'_{k+1}(Z'_k Z_k)^{-1} z_{k+1} + 1)^{-1}(y_{k+1} - z'_{k+1} b^{(4)}_k) \tag{2A.12}$$

Note that s^2 is bounded from (2A.3) as $0 < s^2 < Var(y)$. Since we assume $P_k = (Z'_k Z_k)^{-1}$, and if k is sufficient large, the elements of the matrix $(Z'_k Z_k)^{-1}$ are relative small. Then by dividing for the scalar s^2 they change marginally. What really matters in such situation for comparing the two estimators in (2A.11) and (2A.12) is the signal to noise ratio (SNR), that is Σ/s^2.

- If the SNR is large, meaning that one or more diagonal elements of Σ are very large comparing to s^2, the weight estimates of the two schemes will differ substantially.
- If the SNR is on contrary small, meaning that s^2 is large compared to the diagonal elements of Σ, the weight estimates in the two schemes will be almost identical.

In our simulation exercise, a large SRN corresponds to large instability in the DGP weights. Thus, our conclusion is that in cases where the data are subject to structural instability, the time varying weight scheme is preferable to the Recursive OLS scheme.

2B Graphical examples

We develop few simulation exercises to explain graphically results in Appendix 2A. Let assume that a series is generated from the following DGP:

$$y_t = 1 + z_t w_{t,1} + u_t; \quad u_t \sim N(0, s^2) \tag{2B.1}$$

$$w_t = w_{t-1} + \xi_t; \quad \xi_t \sim N(0, \sigma^2) \tag{2B.2}$$

Table 2.5: Simulation design in exercises 2BI-2BIII

EXERCISES	I	II	III
μ_z	0.00	0.00	0.00
μ_u	0.00	0.00	0.00
μ_ξ	0.00	0.00	0.00
s^2	1.00	1.00	1.00
σ_z^2	1.00	1.00	1.00
σ_ξ^2	0.00	0.04	1.00

where $t = 1, .., T$, where $z = \{z_t\}_{t=1}^T$ is a $(T \times 1)$ normally distributed vector with mean μ_z and variance σ_z in Table 2.4.
We apply the Recursive Least Squares and the Kalman Filter algorithms to estimate $w = \{w_t\}_{t=k+1}^T$, defined as $b^{(4)}$ and $b^{(5)}$ respectively, where k are the initial observations to initialize the estimation algorithms. Precisely, we use the OLS estimate of w on the initial k observation and $P_k = (Z_k' Z_k)^{-1}$ to initialize the algorithms.

Exercise 2B.I: Zero SNR We fix $T = 240$, $k = 120$, $s^2 = 1$, $\sigma^2 = 0$, and $\beta_0 = 1$. Results are in Figure 2.10. The vector w is constant and the two estimators provide the same results.

Exercise 2B.II: Medium SNR In this exercise we fix $s^2 = 1$, $\sigma^2 = 0.04$, and $\beta_0 = 1$. Results are in Figure 2.11. The vector w has a time varying pattern. $b^{(4)}$ and $b^{(5)}$ initialize with the same value, then $b^{(4)}$ is very persistent around the value 1, $b^{(5)}$ on contrary approximates very precisely the pattern of w.

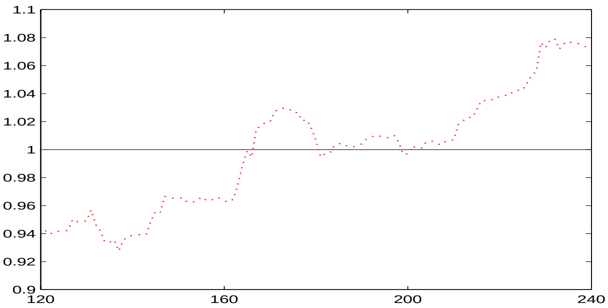
Exercise 2B.III: High SNR In this exercise we fix $s^2 = 1$, $\sigma^2 = 1$, and $\beta_0 = 1$. Results are in Figure 2.12. The vector w follows a very high volatile pattern, $b^{(5)}$ accurately estimates it, $b^{(4)}$ is on contrary a poor estimator.

2C Properties of Multivariate Normal Distribution

Let the pair of vectors x and y be jointly multivariate normal such that $(x', y')'$ has mean and covariance matrix given by

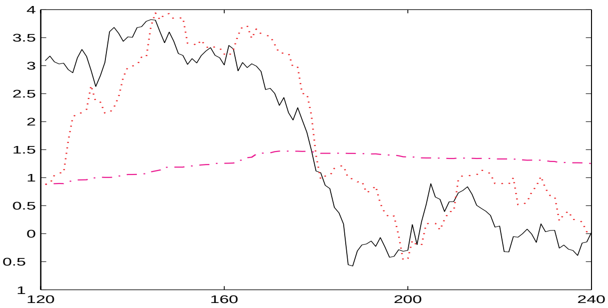
$$\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \tag{2C.1}$$

Figure 2.10: Exercise 2B.I



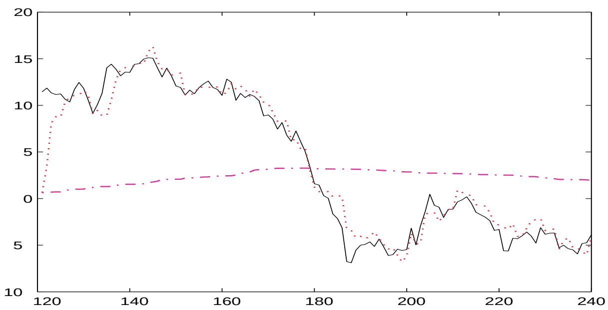
Note: The figure presents the patterns of parameter β (in - line), and estimates $\widehat{\beta}^{(4)}$ (in -. line) and $\widehat{\beta}^{(5)}$ (in .. line) in exercises 2B.I.

Figure 2.11: Exercise 2B.II



Note: The figure presents the patterns of parameter β (in - line), and estimates $\widehat{\beta}^{(4)}$ (in -. line) and $\widehat{\beta}^{(5)}$ (in .. line) in exercises 2B.II.

Figure 2.12: Exercise 2B.III



Note: The figure presents the patterns of parameter β (in - line), and estimates $\widehat{\beta}^{(4)}$ (in -. line) and $\widehat{\beta}^{(5)}$ (in .. line) in exercises 2B.III.

respectively. Then the distribution of x conditional on y is also multivariate normal with mean

$$\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y) \tag{2C.2}$$

and covariance matrix

$$\Sigma_{xx|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \tag{2C.3}$$

Note that Σ and Σ_{yy} are assumed to be non-singular, although in fact Σ_{yy}^{-1} can be replaced by a pseudo inverse.

2D Predictive densities and marginal likelihood for linear models

For Bayesian inference of models (2.1) and (2.2), we use a Normal-Inverted Gamma-2 prior densities for the vector of parameters $(\alpha_i, \beta_i, \sigma_i^2)$

$$\alpha_i, \beta_i, \sigma_i^2 \sim NIG_2(\underline{\alpha}_i, \underline{\beta}_i, \underline{V}_i, \underline{s}_i^2, \underline{\nu}_i)$$

The posteriors take the form, see for example Koop (2003, p. 37),

$$\alpha_i, \beta_i, \sigma_i^2 | F_T \sim NIG_2(\overline{\alpha}_i, \overline{\beta}_i, \overline{V}_i, \overline{s}_i^2, \overline{\nu}_i),$$

where F_T are data up to time T, where

$$\overline{V}_i = (\underline{V}_i^{-1} + X_i'X_i)^{-1},$$

$$\left[\overline{\alpha}_i, \overline{\beta}_i\right]' = \overline{V}_i(\underline{V}_i^{-1}[\underline{\alpha}_i, \underline{\beta}_i]' + X_i'X_i[\widehat{\alpha}_i, \widehat{\beta}_i]'),$$

$$\overline{\nu}_i = \underline{\nu}_i + T,$$

$$\overline{\nu}_i\overline{s}_i^2 = \underline{\nu}_i\underline{s}_i^2 + \widehat{s}_i^2 + ([\widehat{\alpha}_i, \widehat{\beta}_i] - [\widehat{\alpha}_i, \widehat{\beta}_i])(\underline{V}_i + (X_i'X_i)^{-1})^{-1}([\widehat{\alpha}_i, \widehat{\beta}_i]' - [\widehat{\alpha}_i, \widehat{\beta}_i]'),$$

where $X_i = [\iota_s, \{x_{t,i}\}_{t=1}^T]$, where $[\widehat{\alpha}_i, \widehat{\beta}_i]'$ are the OLS estimates of $[\alpha_i, \beta_i]'$ in the model m_i, and where $\widehat{s}_i^2 = ((y - \iota_s\widehat{\alpha}_i + x_i\widehat{\beta}_i)'(y - \iota_s\widehat{\alpha}_i + x_i\widehat{\beta}_i))$.
The predictive density of y_{T+1} conditional on F_T and $x_{T+1,i}$ is

$$y_{T+1}|F_T, x_{T+1,i} = t(\overline{\alpha}_i + x_{T+1,i}\overline{\beta}_i, \overline{s}_i^2(I_T + X_i\overline{V}_iX_i), \overline{\nu}_i) \tag{2D.1}$$

where $t(\cdot)$ indicates the Student t distribution.
The marginal likelihood becomes:

$$p(y|m_i) = \gamma_i \left(\frac{|\overline{V}_i|}{|\underline{V}_i|}\right)^{1/2} (\overline{\nu}_i\overline{s}_i^2)^{-\overline{\nu}_i/2} \tag{2D.2}$$

where

$$\gamma_i = \frac{\Gamma(\overline{\nu}_i/2)(\underline{\nu}_i \underline{s}_i^2)^{\underline{\nu}_i/2}}{\Gamma(\underline{\nu}_i/2)\pi^{s/2}},$$

and where $\Gamma(\cdot)$ is the Gamma function.
We note that normalizing the marginal likelihood in (2D.2) may be very difficult. The term γ_i has often numerator and denominator very huge. However, if the same priors are chosen for the parameters σ_i^2, γ_i has the same value for both the models and then it drops out when calculating model probabilities in equation (2.17). But when T is high, $\overline{\nu}_i$ is high too, making complicate the computation of (2D.2) or too sensitive to $\overline{\nu}_i$. It is often convenient to compute the log marginal likelihoods of the models m_i, to re-scale them, and finally to take the exponentials. Results of (2D.2) do not change, but numbers can be managed. However, there is not a given rule for this procedure, and the problem cannot be solved if, for example, the logarithms are high negative numbers.

2E Estimation of the Bayesian time varying model weight combinations

The model weights of the time varying weights in scheme 8 are defined as in (2A.3) and (2A.4) (z_t may assume different values). The parameters in (2A.3) and (2A.4) are the variances of the residuals in the observation equation, s^2, and the variances of the residuals in the latent equation $q_0^2, \ldots, q_i^2$, where $q_0^2, \ldots, q_i^2$ are the diagonal elements of Σ. The model parameters are collected in the $((1+i) \times 1)$ vector $\theta = (s^2, q_0^2, \ldots, q_i^2)'$. To facilitate the posterior simulation we make use of independent conjugate priors. For the variance parameters we take the inverted Gamma-2 prior

$$q_j^2 \sim \text{IG-2}(\nu_j, \delta_j) \qquad \text{for } j = 0, \ldots, i \tag{2E.1}$$

and

$$s^2 \sim \text{IG-2}(\nu_s, \delta_s), \tag{2E.2}$$

where ν_j, δ_j, $j = 0, \ldots, i$, ν_s, and δ_s are parameters which can be chosen to reflect the prior beliefs about the variances.
Posterior results are obtained using the Gibbs sampler of Geman and Geman (1984) combined with the technique of data augmentation of Tanner and Wong (1987). The latent variables $w = \{w_t\}_{t=1}^T$ are simulated alongside the model parameters θ. The complete data likelihood function is given by

$$p(y, w|z, \theta) = \prod_{t=1}^{T} p(y_t|z_t, w_t, s^2) p(w_t|w_{t-1}, q_0^2, \ldots, q_i^2) \tag{2E.3}$$

where $y = (y_1, \dots, y_T)'$ and $z = (z_1', \dots, z_T')'$. The terms $p(y_t|z_t, w_t, s^2)$ and $p(w_t|w_{t-1}, q_0^2, \dots, q_i^2)$ are normal density functions, which follows directly from (2A.3) and (2A.4) respectively. If we combine (2E.3) together with the prior density $p(\theta)$, which follows from (2E.1)-(2E.2), we obtain the posterior density

$$p(\theta, w|y, z) \propto p(\theta)p(y, w|z, \theta) \tag{2E.4}$$

The sampling scheme can be summarized as follows:

1. Draw w conditional on θ.
2. Draw θ conditional on w.

The full conditional posterior density for the latent regression parameters w in step 1 is computed using the simulation smoother as in Carter and Kohn (1994). The Kalman smoother is applied to derive the conditional mean and variance of the latent factors; for the initial value w_0 a multivariate normal prior with mean 0 is chosen as for scheme 5. To sample the parameters θ in step 2 we can use standard results in Bayesian inference. Hence, the variance parameters s^2 and q_j^2 are sampled from inverted Gamma-2 distributions.

The one-step ahead predictive density of y_{T+1} at time T conditional on y, z and z_{T+1} is given by

$$p(y_{T+1}|y, z, z_{T+1}) = \iint p(y_{T+1}|z_{T+1}, w_{T+1}, s^2)p(w_{T+1}|w_T, q_0^2, \dots, q_i^2) \\ p(\theta, w|y, z)p(z_{T+1}|z_T)dwd\theta \tag{2E.5}$$

Simulating y_{T+1} from the one-step ahead distribution (2E.5) is in fact rather straightforward. In each step of the Gibbs sampler, we use the simulated values of w_T and $(q_0, \dots, q_i^2)$, and equation (2A.4) to simulate w_{T+1}. Equation (2A.3) in combination with the simulated value of w_{T+1}, the current Gibbs draws of s^2, and the simulated value of z_{T+1} then provide a simulated value for y_{T+1}.

Chapter 3

Bayesian Model Averaging in the Presence of Structural Breaks

3.1 Introduction

A growing body of empirical evidence suggests the presence of a certain (albeit modest) level of predictability in aggregate stock returns, see Cochrane (2006) and Campbell and Thompson (in press) for recent accounts. Several financial and macro-economic variables have been reported as being useful predictors, including interest rates and different interest rate spreads such as the yield spread, term spread, and credit spread, as well as valuation ratios such as the dividend yield and the price-earnings ratio. There is, however, little consensus about which variables really are the relevant predictor variables that should enter a successful return forecasting model. Put differently, an investor who intends to use a predictive regression to forecast future stock returns faces *model uncertainty*, see Avramov (2002) and Cremers (2002).

At the same time, recent studies demonstrate that the relationship between stock returns and predictor variables is not stable over time, see Pesaran and Timmermann (2002) and Paye and Timmermann (2005), among others. Important political and economic events, such as changes in monetary policy, oil crises and recessions fundamentally change the economic environment including financial markets. In terms of predictive regressions for stock returns, an investor should take into account the possibility that parameters exhibit occasional *structural breaks*.[1]

A third related issue that investors have to cope with is the fact that parameters in return forecasting model are estimated using historical data, implying the presence

[1]An alternative approach to incorporate instability in the relationship between excess returns and predictor variables is to allow for the presence of recurrent regimes, possibly related to bull and bear states of the market. We refer the interested reader to Ang and Bekeart (2002) and Guidolin and Timmermann (2005a,b) for an analysis of the consequences of such regime-switching behavior for asset allocation decisions.

of *parameter (estimation) uncertainty*, see Kandel and Stambaugh (1996) and Barberis (2000), among others.

To date, the effects of model uncertainty and structural breaks on return predictability and asset allocation decisions have only been considered in isolation.[2] We are not aware of any attempts to incorporate both features in predictive regression models for asset returns jointly. The only exception is Pettenuzzo and Timmermann (2005), but in their framework the investor is quite limited in terms of both the number and complexity of the models that can be combined.

In this Chapter we develop a return forecasting methodology that allows for instability in the relationship between stock returns and predictor variables, for model uncertainty, and for parameter estimation uncertainty simultaneously. On the one hand, the predictive regression specification that we put forward allows for occasional structural breaks of random magnitude in the regression parameters. On the other hand, we allow for uncertainty about the inclusion of the forecasting variables in the model and about the parameter values by employing Bayesian model averaging.

We consider an empirical application to predicting monthly US excess stock returns using a set of 11 financial and macro-economic predictor variables. Our main results can be summarized as follows. We find that over the period 1966-2005, several structural breaks occurred in the relationship between the excess stock return and predictor variables such as the dividend yield and interest rates. These changes appear to be caused by important events such as the oil crisis, changes in monetary policy, the October 1987 stock market crash, and the internet bubble at the end of the 1990s. Although incorporating the different sources of uncertainty does not lead to large improvements in the statistical accuracy of excess return forecasts, their economic value in asset allocation decisions is considerable. We find that a typical investor would be willing to pay up to several hundreds of basis points annually to switch from a passive buy-and-hold strategy to an active strategy based on a return forecasting model that allows for model and parameter uncertainty as well as structural breaks in the regression parameters. The active strategy that incorporates all three sources of uncertainty performs considerably better than strategies based on more restricted return forecasting models.

The Chapter proceeds as follows. In Section 3.2 we develop the return forecasting methodology that allows for instability in the relationship between stock returns and predictor variables, for model uncertainty, and for parameter estimation uncertainty simultaneously. Given that the Bayesian analysis of our model is non-standard, we provide a detailed description of the prior specification and the simulation of the posterior distri-

[2] As both model uncertainty and the presence of structural breaks are typically handled by adopting a Bayesian framework, these features are relatively easy to combine with parameter uncertainty.

butions. In Sections 3.3 and 3.4 we report results from an empirical application of the approach developed in Section 3.2 to predicting US stock returns using a set of 11 financial and macro-economic predictor variables over the period 1966-2005. In Section 3.3 we describe the data set and discuss the choices made for prior specification. In addition, we present full-sample estimation results, which can be considered as an ex-post analysis of the occurrence of structural breaks and the relevance of the different forecasting variables. In Section 3.4 we assess the economic value of incorporating the different sources of uncertainty in investment decisions in real-time by means of an ex-ante recursive forecasting experiment. We conclude in Section 3.5.

3.2 Methodology

3.2.1 The Model

Let r_t denote the stock return in excess of the risk-free rate during period t, and let $x_t = (x_{1t}, x_{2t}, \ldots, x_{kt})'$ denote a vector of k predictor variables (which are observed at the beginning of period t) for $t = 1, \ldots, T$. The benchmark model in the literature for predicting stock returns is the standard linear regression model

$$r_t = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jt} + \varepsilon_t, \tag{3.1}$$

where $\varepsilon_t \sim N(0, \sigma^2)$. Two crucial assumptions underlying the linear regression model are, first, that the set of relevant predictor variables x_t is given and fixed, and second that the regression parameters $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ are constant over time. Both assumptions are questionable in empirical practice, and extensions of the model that drop either of the two assumptions have been developed in recent years. These are briefly discussed first, before we introduce our general model that allows for both uncertainty about the relevant predictor variables and for possible structural breaks in the regression parameters.

First, the fact that the set of predictor variables x_t in (3.1) is given and fixed *a priori* is unrealistic, in the sense that the investor rarely knows with certainty which particular forecasting variables are the relevant ones to include. Avramov (2002) and Cremers (2002) have analyzed this issue of model uncertainty, advocating the use of Bayesian model averaging where all possible models are considered and averaged according to their posterior model probabilities.

A possible way to represent model uncertainty in the linear regression is by means of a latent binary random variable $s_j = 0, 1$ determining the inclusion of x_{jt} in the model for $j = 1, \ldots, k$. The return forecasting model with uncertainty about the relevant predictor variables (but with constant parameters) then is given by

$$r_t = \beta_0 + \sum_{j=1}^{k} s_j \beta_j x_{jt} + \varepsilon_t. \tag{3.2}$$

The k s_j variables can be summarized in a k-dimensional vector $S = (s_1, \ldots, s_k)'$. The vector S can take 2^k different values, each resulting in a different regression model. Model selection is therefore defined in terms of variable selection, see George and McCulloch (1993) and Kuo and Mallick (1998). We denote each model by the index $i = (s_1, \ldots, s_k)_2$. Note that the intercept parameter β_0 is always included in the model, as typically assumed.

Second, as discussed in the introduction, there is abundant empirical evidence showing that the relationship between stock returns and typical predictor variables such as the dividend yield is not stable over time, implying that the assumption of constant regression parameters β_j as in (3.1) is invalid.

There are several ways to extend the linear regression model in order to capture parameter instability. An attractive flexible specification that allows for occasional structural breaks in the regression parameters is as follows:

$$r_t = \beta_{0t} + \sum_{j=1}^{k} \beta_{jt} x_{jt} + \varepsilon_t, \tag{3.3}$$

where $\beta_t = (\beta_{0t}, \beta_{1t}, \ldots, \beta_{kt})$ is a vector of time-dependent regression parameters, which evolve over time according to

$$\beta_{jt} = \beta_{j,t-1} + \kappa_{jt}\eta_{jt}, \qquad j = 0, \ldots, k, \tag{3.4}$$

where $\eta_{jt} \sim N(0, q_j^2)$ for $j = 0, \ldots, k$, and κ_{jt} is an unobserved uncorrelated 0/1 process with $\Pr[\kappa_{jt} = 1] = \pi_j$ for $j = 0, \ldots, k$. Hence, the jth regression parameter β_{jt} remains the same as its previous value $\beta_{j,t-1}$ unless $\kappa_{jt} = 1$ in which case it changes with η_{jt}, see, for example, Koop and Pooter (2004) and Giordani *et al.* (2007) for a similar approach. We note that the predictor variables x_t should be demeaned to exclude that any break in one of the β_{jt} entails a break in the intercept coefficient β_{0t}. Then, β_{0t} represents the unconditional equity premium.

The specification in (3.4) implies that the regression parameters β_{jt}, $j = 0, \ldots, k$, are allowed to change every time period, but they need not change at any point in time. The presence of a change is described by the latent binary random variable κ_{jt}, while the magnitude of the change is determined by η_{jt}, which is assumed to be normally distributed with mean zero. Note that the changes in the separate regression parameters are not restricted to coincide as in Pesaran and Timmermann (2002) but rather are allowed to occur at different points in time, see also Giordani *et al.* (2007).

While model uncertainty and structural breaks in the context of return prediction models have been studied in isolation, attempts to consider both features simultaneously are very rare, but see Pettenuzzo and Timmermann (2005). Using the representation of model uncertainty as given in (3.2), it actually turns out to be fairly straightforward to incorporate structural breaks as well, for example by adding the time-varying parameter specification as given in (3.4). Hence, we propose the following linear regression model for the excess stock return r_t:

$$r_t = \beta_{0t} + \sum_{j=1}^{k} s_j \beta_{jt} x_{jt} + \varepsilon_t, \tag{3.5}$$

where $\varepsilon_t \sim N(0, \sigma^2)$ and $\beta_t = (\beta_{0t}, \beta_{1t}, \ldots, \beta_{kt})'$ evolves over time according to (3.4) as before.

For inference in our model (3.5) with (3.4) we opt for a Bayesian approach. This will provide the posterior distribution of the latent κ_{jt} processes for $j = 0, \ldots, k$ and $t = 1, \ldots, T$. Bayesian inference on S leads to posterior probabilities of the 2^k possible models that can in turn be used for Bayesian model selection and Bayesian model averaging. Notice that κ_{jt} does not depend on S. At the same time the estimate of κ_{jt} can be different across different values of S and hence breaks can occur in different parameters and at different time periods across models. Below we first discuss prior specification, followed by a description of the posterior simulation algorithm.

3.2.2 Prior Specification and Posterior Simulation

The parameters in the model (3.5) with (3.4) are the inclusion variable $S = (s_1, \ldots, s_k)'$, the structural break probabilities $\pi_0, \ldots, \pi_k$ and the magnitude of the breaks in the regression parameters, $q_0^2, \ldots, q_k^2$, in addition to the variances of the residual returns, σ^2. The model parameters are collected in the $(3(1+k)\times 1)$ vector $\theta = (s_1, \ldots, s_k, \pi_0, \ldots, \pi_k, q_0^2, \ldots, q_k^2, \sigma^2)'$. To facilitate the posterior simulation we make use of independent conjugate priors. For the variable inclusion parameters we take the following prior distribution

$$\Pr[s_j = 1] = \lambda_j \qquad \text{for } j = 1, \ldots, k. \tag{3.6}$$

Hence, the parameter λ_j reflect our prior belief about the inclusion of the jth explanatory variable, see George and McCulloch (1993) and Kuo and Mallick (1998). For the structural break probability parameters we take Beta distributions

$$\pi_j \sim Beta(a_j, b_j) \qquad \text{for } j = 0, \ldots, k. \tag{3.7}$$

The parameters a_j and b_j can be set according to our prior belief about the occurrence of structural breaks. Finally, for the variance parameters we take the inverted Gamma-2 prior

$$q_j^2 \sim \text{IG-2}(\nu_j, \delta_j) \qquad \text{for } j = 0, \ldots, k \tag{3.8}$$

and

$$\sigma^2 \sim \text{IG-2}(\nu_s, \delta_s), \tag{3.9}$$

where ν_j, δ_j, $j = 0, \ldots, k$, ν_s, and δ_s are parameters which can be chosen to reflect the prior beliefs about the variances. Realistic values of the parameters in the different prior distributions depend on the problem at hand. In Section 3.3 we discuss the prior settings for our empirical application to US stock returns.

Posterior results are obtained using the Gibbs sampler of Geman and Geman (1984) combined with the technique of data augmentation of Tanner and Wong (1987). The latent variables $B = \{\beta_t\}_{t=1}^T$ and $K = \{\kappa_t\}_{t=1}^T$ with $\kappa_t = (\kappa_{0t}, \kappa_{1t}, \ldots, \kappa_{kt})'$ are simulated alongside the model parameters θ.

The complete data likelihood function is given by

$$p(r, B, K|x, \theta) = \prod_{t=1}^{T} p(r_t|S, x_t, \beta_t, \sigma^2) p(\beta_t|\beta_{t-1}, \kappa_t, q_0^2, \ldots, q_k^2) \prod_{j=0}^{k} \pi_j^{\kappa_{jt}} (1-\pi_j)^{1-\kappa_{jt}}, \tag{3.10}$$

where $r = (r_1, \ldots, r_T)$ and $x = (x_1', \ldots, x_T')'$. The terms $p(r_t|S, x_t, \beta_t, \sigma^2)$ and $p(\beta_t|\beta_{t-1}, \kappa_t, q_0^2, \ldots, q_k^2)$ are normal density functions, which follow directly from (3.5) and (3.4), respectively. If we combine (3.10) together with the prior density $p(\theta)$, which follows from (3.9)–(3.7), we obtain the posterior density

$$p(\theta, B, K|r, x) \propto p(\theta) p(r, B, K|x, \theta). \tag{3.11}$$

To derive the Gibbs sampler we combine the Kuo and Mallick (1998) algorithm for variable selection and the efficient sampling algorithm of Gerlach *et al.* (2000) to handle the (occasional) structural breaks. If we define $\theta = (S, \bar{\theta})$ with $\bar{\theta} = (\pi_0, \ldots, \pi_k, q_0^2, \ldots, q_k^2, \sigma^2)'$, the sampling scheme can be summarized as follows:

1. Draw S conditional on B, K, $\bar{\theta}$, r and x.

2. Draw K conditional on S, $\bar{\theta}$, r and x.

3. Draw B conditional on S, K, $\bar{\theta}$, r and x.

4. Draw $\bar{\theta}$ conditional S, B, K, r and x.

Step 1 is done similarly to and Kuo and Mallick (1998), which is a simplified version of the George and McCulloch (1993) algorithm. Starting from the previous iteration, the variable S is drawn from its full conditional posterior distribution. The complete data likelihood function (3.10) is computed for $s_j = 0$ and $s_j = 1$ resulting in $p_{j,0}$ and $p_{j,1}$. The full conditional posterior is then given by

$$\Pr[s_j = 1|r, x, \bar{\theta}, B, K, S_{-j}] = \frac{p_{j,1}}{p_{j,0} + p_{j,1}}, \tag{3.12}$$

for $j = 1, \ldots, k$, where $S_{-j} = (s_1, \ldots, s_{j-1}, s_{j+1}, \ldots, s_k)'$.

The (occasional) structural breaks, measured by the latent variable κ_{jt}, are drawn in step 2 using the algorithm of Gerlach *et al.* (2000), which derives its efficiency from generating κ_{jt} without conditioning on the states β_{jt}. The conditional posterior density for κ_t, $t = 1, \ldots, T$ unconditional on B is

$$\begin{aligned} p(\kappa_t|K_{-t}, S, \bar{\theta}, r, x) &\propto p(r|K, S, \bar{\theta}, x)p(\kappa_t|K_{-t}, S, \bar{\theta}, x) \\ &\propto p(r_{t+1}, \ldots, r_T|r_1, \ldots, r_t, K, S, \bar{\theta}, x) \\ &\qquad p(r_t|r_1, \ldots, r_{t-1}, \kappa_1, \ldots, \kappa_t, S, \bar{\theta}, x)p(\kappa_t|K_{-t}, S, \bar{\theta}, x), \end{aligned} \tag{3.13}$$

where $K_{-t} = \{\kappa_s\}_{s=1, s\neq t}^{T}$. Note that the term $p(\kappa_t|K_{-t}, S, \bar{\theta}, x)$ is simply given by $\prod_{j=0}^{k} \pi_j^{\kappa_{jt}}(1-\pi_j)^{1-\kappa_{jt}}$ given that κ_{jt} does not depend on s_j. The two remaining densities $p(r_{t+1}, \ldots, r_T|r_1, \ldots, r_t, K, S, \bar{\theta}, x)$ and $p(r_t|r_1, \ldots, r_{t-1}, \kappa_1, \ldots, \kappa_t, S, \bar{\theta}, x)$ can easily be evaluated as shown in Gerlach *et al.* (2000). Because κ_t can take a finite number of values, the integrating constant can easily be computed by normalization.

The full conditional posterior density for the latent regression parameters B in step 3 is computed using the simulation smoother as in Carter and Kohn (1994). The Kalman smoother is applied to derive the conditional mean and variance of the latent factors; for the initial value β_0 a multivariate normal prior with mean 0 is chosen.

Note that in case the variable x_j is not selected, the full conditional distributions of κ_{jt} and β_{jt} for $t = 1, \ldots, T$ do not depend on the data r and x. Hence, in this case we sample unconditionally from the process in (3.4) and the binary random process for κ_{jt}.

To sample the parameters $\bar{\theta}$ in step 4 we can use standard results in Bayesian inference. Hence, the probabilities π_j are sampled from Beta distributions and the variance parameters σ^2 and q_j^2 are sampled from inverted Gamma-2 distributions.

3.2.3 Using the Posterior Results

The output of the Gibbs sampler can be used to compute several quantities of interest. First, the marginal posterior distribution of the individual s_j parameters $p(s_j|r, x)$ represents the posterior probability that variable x_j is included in the model. This can used to assess the (relative) importance of the different predictor variables for forecasting stock returns. The interaction of different predictor variables can also be examined. For example, following Doppelhofer and Weeks (2005) the degree of dependence or jointness among two explanatory variables x_j and x_l can be formally computed by the following measure of jointness:

$$J_{j,l} = \log\left(\frac{p(s_j = 1 \cap s_l = 1|r,x)}{p(s_j = 1|r,x)p(s_l = 1|r,x)}\right), \tag{3.14}$$

where the numerator is the posterior joint probability of inclusion of the couple of variables x_j and x_l, and the denominator is the product of the marginal posterior probabilities of the inclusion of the jth and lth variables. We consider two variables to be significant substitutes if $J_{j,l} < -1$, and significant complements if $J_{j,l} > 1$. In addition, posterior model probabilities are easily obtained from the joint posterior distribution $p(S|r,x)$ of the inclusion variable S.

Second, we can use the simulated draws of K to do inference on the occurrence of structural breaks in the regression parameters during the sample period. Obviously, we might consider the marginal posterior distribution of a single κ_{jt}, $p(\kappa_{jt}|r,x)$, but the presence of contemporaneous breaks in different parameters can also be evaluated. Similarly, we can examine whether posterior evidence for breaks differs across models by conditioning on the inclusion/exclusion of certain variables in the model, for example, the posterior probability of a break in the regression parameter of variable x_j given that variables x_l and x_m are included in the model is given by $p(\kappa_{jt}|s_l = s_m = 1, r, x)$.

Third, the model in (3.5) with (3.4) can be used to predict future returns r_{T+h} for $h \geq 1$. As our inference is Bayesian, we can explicitly take into account parameter uncertainty, uncertainty in variable selection, and uncertainty in the occurrence of structural breaks. In the empirical application in the next section, we focus on one-step ahead forecasting. For that reason the discussion below is limited to the case $h = 1$, but it can be generalized to $h > 1$ straightforwardly.

The one-step ahead predictive density of r_{T+1} at time T conditional on r, x and x_{T+1} is given by

$$\begin{aligned} p(r_{T+1}|r,x,x_{T+1}) = &\iint \sum_S \sum_K \sum_{\kappa_{T+1}} p(r_{T+1}|S,x_{T+1},\beta_{T+1},\sigma^2) \\ &p(\beta_{T+1}|\beta_T,\kappa_{T+1},q_0^2,\ldots,q_k^2)\prod_{j=0}^{k}\pi_j^{\kappa_{j,T+1}}(1-\pi_j)^{1-\kappa_{j,T+1}}p(B,K,S,\bar{\theta}|r,x)dBd\bar{\theta}, \end{aligned} \tag{3.15}$$

where $p(r_{T+1}|S,x_{T+1},\beta_{T+1},\sigma^2)$ and $p(\beta_{T+1}|\beta_T,\kappa_{T+1},q_0^2,\ldots,q_k^2)$ follow directly from (3.5) and (3.4) and where $p(B,K,S,\bar{\theta}|r,x)$ is the posterior density.

As we average over the posterior distribution of S we implicitly take a weighted average over all possible model specifications, where the weights are the posterior model probabilities. The posterior distribution also reflects our posterior beliefs about the in-sample structural breaks K. Finally, note that we also average with respect the unknown κ_{T+1} variable to account for the possibility that a break may occur in the out-of-sample period $T+1$, where the weights are given by $\prod_{j=0}^{k} \pi_j^{\kappa_{j,T+1}}(1-\pi_j)^{1-\kappa_{j,T+1}}$.

Simulating r_{T+1} from the one-step ahead distribution (3.15) is in fact rather straightforward. In each step of the Gibbs sampler, we use the simulated values of π_j to draw the out-of-sample values of $\kappa_{j,T+1}$ for $j = 0, \ldots, k$. Given the simulated values of $\kappa_{j,T+1}$ and given the Gibbs draws of q_j^2 and β_T we can simulate β_{T+1} using (3.4). Equation (3.5) in combination with the simulated value of β_{T+1} and the current Gibbs draws of S and σ^2 then provide a simulated value for r_{T+1}.

Of course, often forecasting returns in itself is not the ultimate goal, but rather a means for determining the optimal asset allocation, for example. We postpone a detailed discussion of this issue in the context of our empirical application to Section 3.4.

3.3 Model uncertainty and structural breaks in return forecasting models for the S&P 500

3.3.1 Data

The dependent variable is the continuously compounded monthly return on the S&P 500 index in excess of the 1-month T-Bill rate, from January 1966 to December 2005, for a total of 480 observations. The set of predictors consists of $k = 11$ financial and macro-economic variables that have often been considered as potentially relevant factors for forecasting stock returns. Specifically, we include the S&P 500 index price-earnings ratio (PE), the S&P 500 index dividend yield (DY) defined as the ratio of dividends over the previous twelve months and the current stock price, the 3-month T-Bill rate ($I3$), the monthly change in the 3-month T-bill rate ($DI3$), the term spread (TS) defined as the difference between the 10-year T-bond rate and the 3-month T-bill rate, the credit spread (CS) defined as the difference between Moody's Baa and Aaa yields, the yield spread (YS) defined as the difference between the Federal funds rate and the 3-month T-bill rate, the annual inflation rate based on the producer price index (PPI) for finished goods (INF), the annual growth rate of industrial production (IP), the annual growth rate of the monetary base (MB), and the log monthly realized volatility of the S&P 500 index ($LVol$). The monthly realized volatility is computed using daily returns, where we follow French *et al.* (1987) and Marquering and Verbeek (2004) by assuming that daily returns are appropriately described by a first-order autoregressive process. In particular, we use the following realized volatility estimator,

$$\hat{\sigma}_t^2 = \sum_{t=1}^{N_t} (r_{i,t} - \bar{r}_t)^2 \left[1 + \frac{2}{N_t} \sum_{j=1}^{N_t - 1} (N_t - j) \hat{\phi}_t^j \right],$$

where $r_{i,t}$ is the return on day i in month t which has N_t trading days, $\bar{r}_t$ is the average daily return in month t, and $\hat{\phi}_t$ denotes the first-order autocorrelation estimated using daily returns within month t.

A final remark about the data concerns the fact that we take into account the typical publication lag of macroeconomic variables in order to avoid look-ahead bias. We therefore include inflation and the growth rates of industrial production and the monetary base with a two-month lag. As the financial variables are promptly available, these are included with a one-month lag.

3.3.2 Prior specification

We set the prior probability of inclusion of the variable x_{jt} equal to $\lambda_j = 0.5$ for all $j = 1, \ldots, k$. As in our framework the λ_j's are independent across j, a 'diffuse' prior for λ_j implies that all individual models have equal prior probability, as discussed in Fernández *et al.* (2001). For the hyperparameters a_j and b_j in the Beta distribution for the prior probability of breaks in the regression parameters π_j, we impose $a_j = 0.7$ and $b_j = 35$ for all j. This implies that the prior mean duration between breaks in a particular regression parameter is equal to 51 months. For the q_j^2 parameters, $j = 0, \ldots, k$, we take a very peaked prior with mode near zero to limit the number of potential breaks. Finally, the Gamma-2 prior parameters for σ^2 are 0.002 with 35 degrees of freedom.

Unreported results show that the posterior results are not very sensitive to moderate changes of the prior parameters of λ_j and σ^2. However, the posterior results are sensitive to the prior settings of the π_j and the q_j^2 parameters. For example, an increase in the prior mean of the duration between breaks leads to less breaks, see Giordani *et al.* (2007) for similar results.

3.3.3 Full-sample estimation results

We estimate the linear regression model with variable selection and occasional structural breaks in the parameters (3.5) with (3.4) using the complete sample period from January 1966 until December 2005. This enables us to provide an ex-post analysis of the relevance of the different predictor variables and possible breaks in their regression parameters[3].

[3] As noted before, the predictor variables are demeaned to exclude that possible breaks in the relation between the excess returns and some predictors automatically entail a break in the intercept β_{0t}. This

Table 3.1 provides the posterior mean for the probability of inclusion parameter λ_j. The lagged 3-month T-Bill rate is included in the return forecasting model most often, with the posterior inclusion probability being equal to 0.9. This perhaps is not surprising given that the dependent variable is the stock return in excess of the closely related 1-month T-bill rate. Other variables for which the posterior probability of inclusion is higher than the prior probability of 0.5 are the dividend yield (0.728), the yield spread (0.628), the change in the 3-month T-bill rate (0.606), and the annual growth rate of the monetary base (0.601). The dividend yield is the classical example of a financial ratio that has been scrutinized time and again for its predictive ability for stock returns, with varying degrees of success, starting with Keim and Stambaugh (1986), Campbell and Shiller (1988) and Fama and French (1988). It also has often been used in studies that consider the implications of return predictability for asset allocation, see Kandel and Stambaugh (1996), Barberis (2000), and Pettenuzzo and Timmermann (2005), among many others. The effects of monetary policy on the stock market have been examined extensively, see the survey by Sellin (2001). Patelis (1997) documents that various monetary policy indicators have predictive ability for future stock returns, primarily by affecting expected excess returns.[4] Here we find a relatively high posterior inclusion for the monetary base growth rate. Ample empirical evidence has been found for pronounced effects of monetary policy announcements on the stock market. In particular, stock prices have been shown to respond to (unexpected) changes in the Federal funds target rate, see Bernanke and Kuttner (2005) for a recent assessment. This explains the high posterior inclusion probability for the yield spread of 0.621. The posterior inclusion probability of the change in the 3-month T-bill rate of 0.606 suggests that changes in the level of interest rates bear some additional information to the interest rate level itself. The low posterior inclusion probability of the term spread of 0.220 is surprising, as the slope of the yield curve has frequently been found to be an important predictive variable for stock returns, see Fama and French (1989) and Aït-Sahalia and Brandt (2001), for example. This may be attributed to the short forecast horizon of one month considered here, for which variables related to short-term interest rates such as the level and change of the 3-month T-bill rate and the yield spread may have more predictive power. Finally, inflation and stock return volatility have particular low posterior inclusion probabilities, indicating that these variables have not been useful as predictors of stock returns over the sample period considered.

---

also implies that any breaks in the mean of the forecasting variables are captured by changes in β_{0t}. See Lettau and van Nieuwerburgh (in press) for an analysis of the implications of breaks in mean of forecasting variables for return predictability.

[4]It is also interesting to note that Jensen *et al.* (1996) report that the predictive ability of variables such as the dividend yield and the default spread depends on the monetary policy regime. In particular, predictable variation in expected stock returns is higher during periods of expansive monetary policy than during more restrictive periods. We do not explicitly consider such nonlinear effects here.

Table 3.1: Posterior probability of predictor variable selection

Variable	Marginal posterior inclusion probability
PE_{t-1}	0.336
DY_{t-1}	0.728
$I3_{t-1}$	0.900
$DI3_{t-1}$	0.606
TS_{t-1}	0.220
YS_{t-1}	0.621
CS_{t-1}	0.283
INF_{t-2}	0.122
IP_{t-2}	0.388
MB_{t-2}	0.601
$LVOL_{t-1}$	0.140

Note: The table presents the marginal posterior probabilities of variables to be selected in the predictive regression model (3.5) for monthly S&P 500 excess returns, estimated over the period January 1966-December 2005. PE = price-earnings ratio; DY = dividend yield; $I3$ = 3-month T-bill rate; $DI3$ = monthly change in the 3-month T-bill rate; TS = term spread, defined as the difference between 10-year T-bond rate and the 3-month T-bill rate; CS = credit spread, defined as the difference between Moody's Baa and Aaa yields; YS = yield spread, defined as the difference between the Federal funds rate and the 3-month T-bill rate; INF = annual inflation rate based on the PPI; IP = annual growth rate of industrial production; MB = annual growth rate of the monetary base; $LVOL$ = log monthly stock return volatility.

Table 3.2: Posterior probability of joint selection

	PE	DY	$I3$	$DI3$	TS	YS	CS	INF	IP	MB	$LVOL$
PE		0.157	0.323	0.158	0.093	0.228	0.142	0.046	0.056	0.174	0.088
DY	−0.088		0.707	0.414	0.150	0.404	0.193	0.084	0.321	0.509	0.085
$I3$	0.021	0.052		0.507	0.184	0.535	0.268	0.122	0.335	0.595	0.133
$DI3$	−0.046	−0.027	−0.038		0.130	0.285	0.147	0.054	0.249	0.336	0.068
TS	0.019	−0.010	−0.014	−0.003		0.154	0.045	0.038	0.052	0.053	0.014
YS	0.019	−0.048	−0.024	−0.091	0.017		0.172	0.087	0.258	0.376	0.125
CS	0.047	−0.013	0.013	−0.024	−0.017	−0.004		0.046	0.088	0.188	0.057
INF	0.005	−0.005	0.012	−0.020	0.011	0.011	0.011		0.041	0.069	0.003
IP	−0.074	0.039	−0.014	0.014	−0.033	0.017	−0.022	−0.006		0.234	0.033
MB	−0.028	0.071	0.054	−0.028	−0.079	0.003	0.018	−0.004	0.001		0.092
$LVOL$	0.041	−0.017	0.007	−0.017	−0.017	0.038	0.017	−0.014	−0.021	0.008	

Note: The table presents posterior probabilities of pairs of variables to be selected in the predictive regression model (3.5) for monthly S&P 500 excess returns, estimated over the period January 1966-December 2005. See Table 3.1 for a description of the predictor variables.

Table 3.3: Bivariate jointness

	DY	$I3$	$DI3$	TS	YS	CS	INF	IP	MB	$LVOL$
PE	−0.446	0.066	−0.252	0.235	0.087	0.398	0.115	−0.850	−0.151	0.625
DY		0.076	−0.064	−0.064	−0.114	−0.068	−0.056	0.128	0.152	−0.175
$I3$			−0.073	−0.072	−0.044	0.051	0.105	−0.042	0.096	0.059
$DI3$				−0.027	−0.280	−0.156	−0.321	0.059	−0.081	−0.224
TS					0.119	−0.317	0.358	−0.487	−0.906	−0.785
YS						−0.021	0.142	0.067	0.008	0.365
CS							0.293	−0.225	0.101	0.371
INF								−0.151	−0.061	−1.855
IP									0.003	−0.495
MB										0.088

Note: The table presents the degree of dependence or jointness measure defined in (3.14) among pairs of variables in the predictive regression model (3.5) for monthly S&P 500 excess returns, estimated over the period January 1966-December 2005. See Table 3.1 for a description of the predictor variables.

Additional insight into the variable selection results can be obtained from the joint selection of different variables. For that purpose, the posterior joint probabilities of inclusion for all possible pairs of variables are presented in the above diagonal part of Table 3.2. The entries below the diagonal represent the difference between the empirical posterior joint inclusion probability and the joint inclusion probability under the assumption of independence, which equals the product of the marginal probabilities of the two variables involved as given in Table 3.1. Also note that the prior probability of joint inclusion for any two variables is equal to 0.25, given that all individual prior probabilities are equal to 0.5 and independent across variables. Table 3.3 shows the values of the measure of jointness, as defined in (3.14). Obviously, part of the results in Table 3.2 follow directly from the variable-specific selection probabilities in Table 3.1. For example, given that the 3-month T-bill rate and the dividend yield have such high individual probabilities of inclusion, their combination has high posterior probability to be selected as well. Hence, the measure of jointness for $(DY, I3)$ does not have a particularly large value. The couple (PE, DY) has a posterior probability of being selected together of 0.157, which is considerably lower than the joint inclusion probability under independence of 0.245 and also fairly close to the lower bound (0.064) that is possible given their individual inclusion probabilities. Their coefficient of jointness therefore is substantially negative, indicating that these variables are close substitutes. This is not surprising given that both the price-earnings ratio and the dividend yield are well-accepted valuation measures having similar predictive content for the development of the stock market. The posterior probability of joint inclusion of the monetary base growth rate with both the dividend-yield and the 3-month T-bill rate is higher than its prior value, while the corresponding jointness measures in Table 3.3 confirm that money growth is complementary to these financial variables. Many of the other combinations of variables have a posterior inclusion probability lower than the prior value of 0.25, while it is lower than the probability under independence for an even larger number of pairs.

Finally, Table 3.4 lists the ten models which have the highest posterior model probabilities. The conclusions from this table agree with the findings from Tables 3.1-3.3 as discussed above. First, the dividend yield and both the level and change of the 3-month T-bill rate are almost always included in the most likely models, in line with their relatively high individual inclusion probabilities. The same applies to the yield spread and the monetary base growth rate, although they appear somewhat less frequently. Second, the finding that industrial production is included in five of the top ten models may seem surprising at first given its modest individual posterior inclusion probability of 0.388, but this can be explained by the positive value of its jointness measure with the dividend yield, the change in the 3-month T-bill rate, and the yield spread. Hence, it seems that

Table 3.4: Posterior model probabilities

Model	Posterior probability
C, DY, $I3$, $DI3$, IP, MB	0.0627
C, DY, $I3$, YS, MB	0.0457
C, DY, $I3$, $DI3$, MB	0.0400
C, DY, $I3$, $DI3$, YS, IP, MB	0.0313
C, DY, $I3$, $DI3$, IP	0.0307
C, DY, $I3$, $DI3$, CS, IP, MB	0.0257
C, PE, DY, $I3$, TS	0.0223
C, $DI3$, YS, IP	0.0213
C, DY, $I3$, $DI3$	0.0197
C, DY, $I3$, $DI3$, TS, YS, MB	0.0197

Note: The table lists the ten models for monthly S&P 500 excess returns, estimated over the period January 1966-December 2005, with highest posterior probabilities and their respective probabilities. See Table 3.1 for a description of the predictor variables.

macro-economic information as represented by industrial production growth is a useful complement to the information obtained from financial variables. Finally, it is worth nothing that the sum of the posterior probabilities for these ten models is 0.30, suggesting that the data is reasonably informative as to which variables are most useful for predicting stock returns. At the same time, this should not be mistaken for saying that the data is telling so much as to which forecasting models actually provide a good fit in an absolute sense. The posterior probabilities in Table 3.4 only measure the fit of these predictive regressions relative to the other possible models.

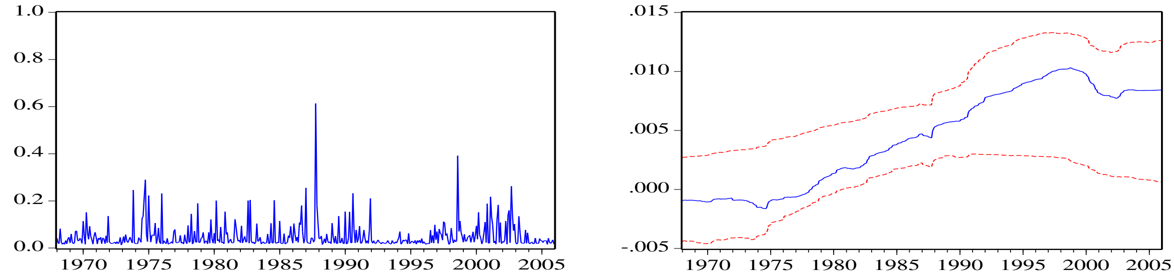
We next turn to the regression parameters and possible structural breaks therein. Figure 3.1 shows the posterior mean for the latent binary variable κ_{jt} governing the occurrence of changes in the regression parameters, together with the associated posterior mean for β_{jt}, conditional on inclusion of the variable j, that is $s_j = 1$. For the coefficients, the 25th and 75th percentiles of the posterior distributions are also shown. Several conclusions emerge from these graphs. First, the posterior means of κ_{jt} show quite erratic, 'spiky' behavior, suggesting that the probabilities of structural breaks in the parameters vary considerably from one period to the next.[5] This occurs for two reasons. On the one hand, κ_{jt} can be different across different values of S, such that breaks can occur at different times across models. On the other hand, in case a break is estimated to have occurred in a certain month, the probability of a break in the next month will be much lower.

[5]Recall that the posterior mean of κ_{jt} is identical to the posterior probability of a break occurring in the regression parameter for the jth variable x_{jt} at time t.

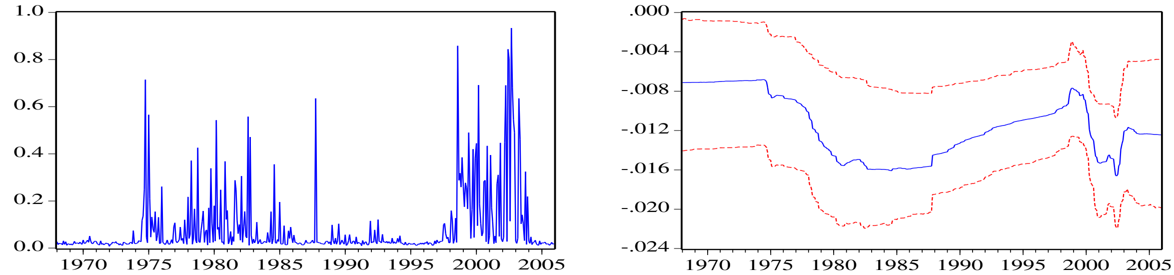
Second, despite the volatile behavior of the break probabilities, three periods with considerable probability mass for structural change can be identified: during the years 1974-1975, the years 1979-1982, and around 2001. Political reasons and the oil price shocks provide possible explanations for the first break period. The second break period coincides with "monetarist experiment" of the Federal Reserve under Chairman Volcker. Note that this often is identified as the start of a marked structural change in the Fed's monetary policy, see Clarida *et al.* (2000), among others. The third break period may obviously be related to the burst of the internet bubble. Also note that the stock market crash in October 1987 gives rise to an isolated jump in the break probability for most variables, with notable exceptions being the level of the 3-month T-bill rate, the yield spread, and inflation.

Third, the posterior means of the regression parameters also reveal several interesting findings. The pattern of the intercept β_{0t} suggests a gradual increase in the unconditional equity premium during the 1980s and 1990s from around zero to 10 percent per year, followed by a decline just before the turn of the millennium. As expected, for the price-earnings ratio and the dividend yield we find a negative and positive coefficient, respectively. The most substantial changes in these parameters occur during the period 1999-2002 (in addition to the decline in the PE coefficient during the second half of the 1970s and an increase in the DY coefficient during the same period), reflecting the large decline in the dividend yield and the corresponding large increase in the price-earnings ratio due to the dramatic boom of stock prices during that period. The coefficients for both the level and change in the 3-month T-bill rate are negative, so that higher interest rates lead to lower stock return forecasts. The most substantial breaks in these coefficients appear to have occurred around 1982, at the time the Federal Reserve switched from targeting M1 to targeting the Federal funds rate and abandoned its nonborrowed reserves operating procedure, see Thornton (2006). Temporary instabilities in the yield spread and credit spread coefficients are observed at approximately the same time. The coefficients related to inflation and industrial production growth display the largest changes around 1975, due to the oil price shocks and the higher level of inflation and slowdown in economic growth that followed. Finally, the coefficients of the monetary base and volatility experienced very large breaks around October 1987. Especially the pattern in the coefficient of volatility is interesting, showing a gradual decline up to the moment of the crash, and a gradual increase thereafter. Also note that the volatility coefficient changes sign, being negative between the end of the 1970s and early 1990s and positive before and after this period. We refer to Ghysels *et al.* (2005) and Guo and Whitelaw (2006) for recent discussions on the risk-return trade-off and empirical evidence therefor.

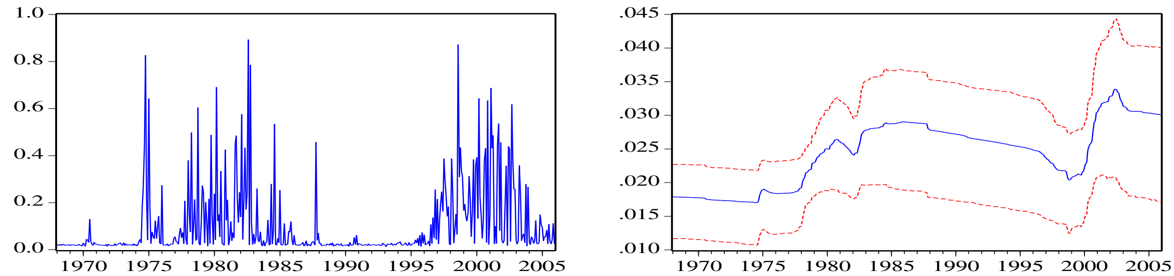
Figure 3.1: Posterior densities of the breaks and β parameters conditional on inclusion



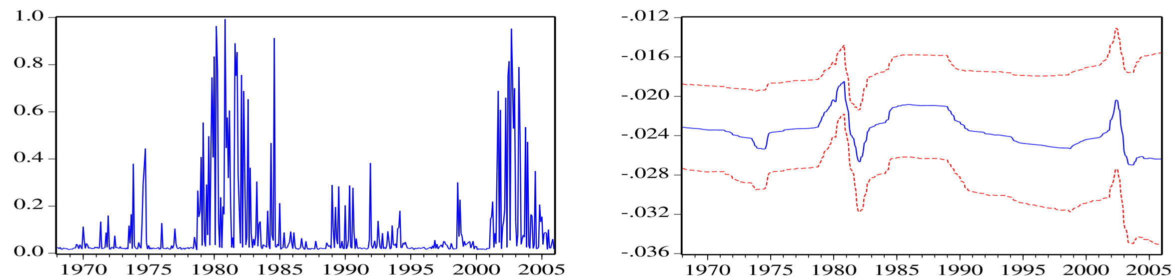
(a) Constant



(b) Price-earnings ratio (PE)

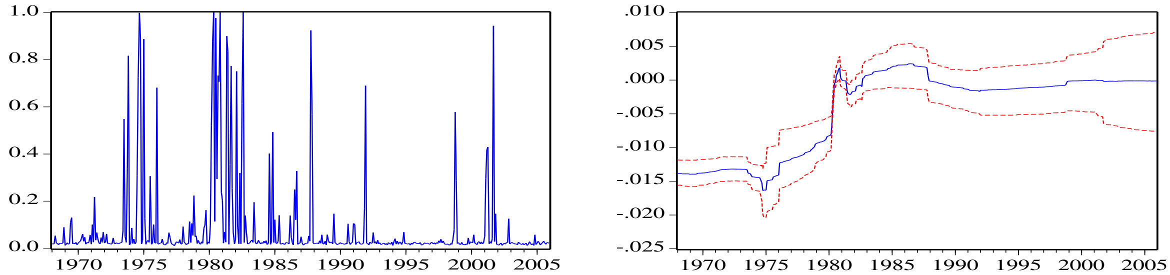


(c) Dividend yield (DY)

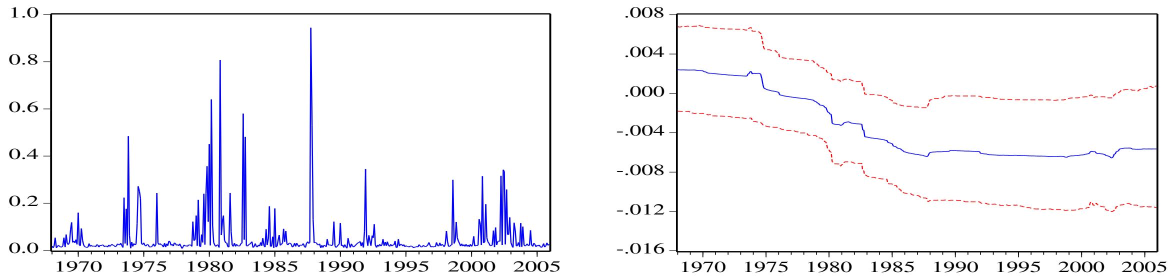


(d) 3-month T-bill rate ($I3$)

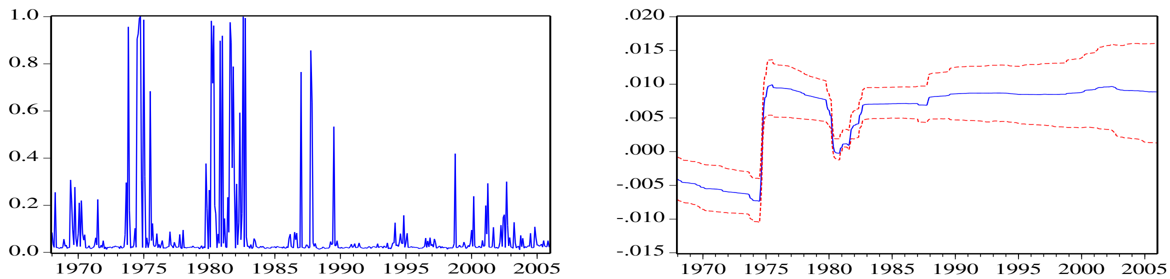
(continued on next page)



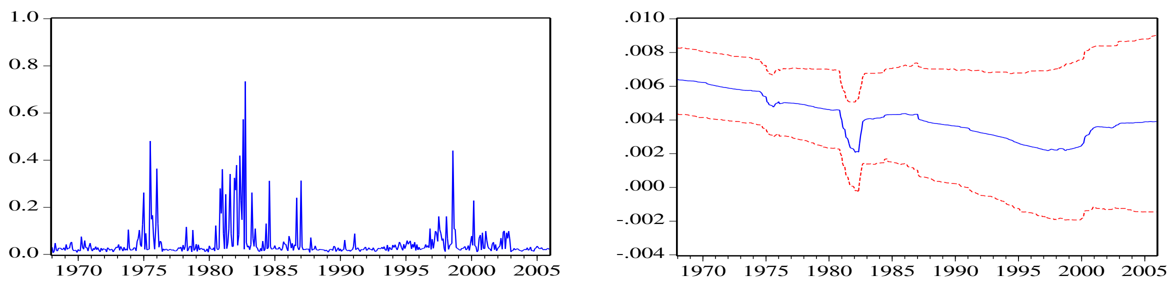
(e) Change in 3-month T-bill rate ($DI3$)



(f) Term spread (TS)

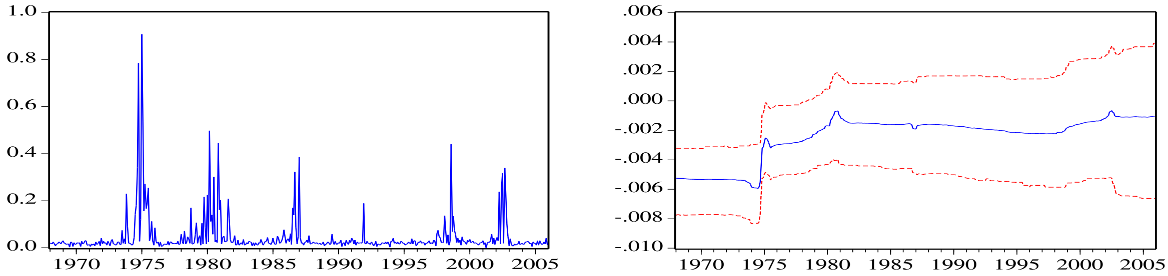


(g) Yield spread (YS)

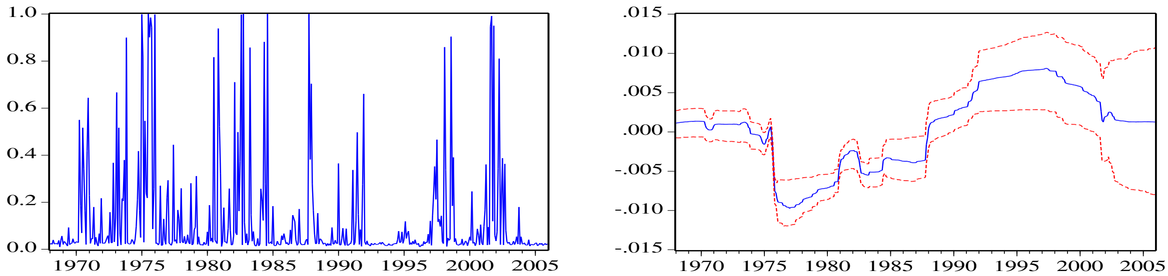


(h) Credit spread (CS)

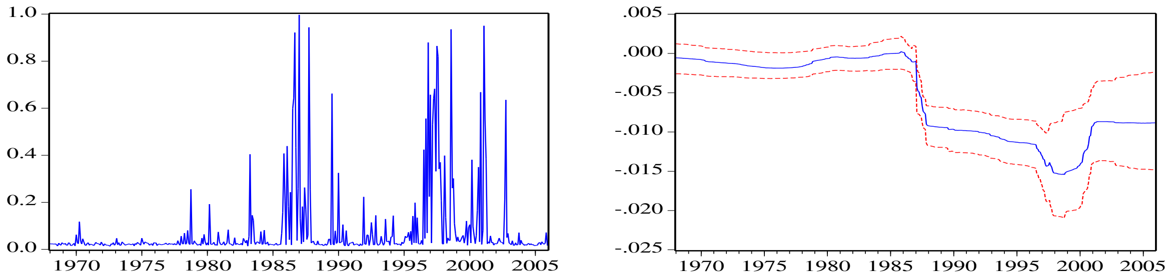
(continued on next page)



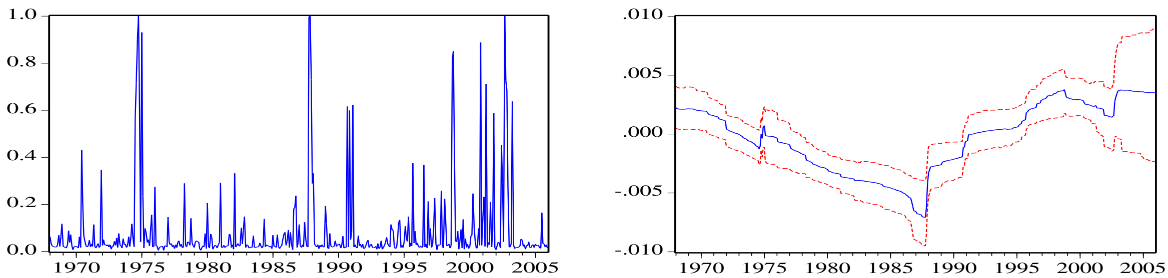
(i) Inflation (INF)



(j) Industrial production (IP)



(k) Monetary base (MB)



(l) Volatility ($LVOL$)

Note: The graphs in this figure show the posterior means (solid line) of κ_{jt} on the left side and β_{jt} on the right side, conditional upon inclusion of the jth variable ($s_j = 1$), in the predictive regression model (3.5) for monthly S&P 500 excess returns, estimated over the period January 1966-December 2005. The dashed lines in the graphs for the coefficients are the 25th and 75th percentiles of the posterior densities.

3.4 Active investment strategies allowing for model uncertainty and structural breaks

The full-sample results presented in the previous section provide a useful *ex post* characterization of the (relative) importance of financial and macroeconomic variables as predictors in return forecasting models and of possible breaks in the regression parameters. For an investor, however, both issues of variable selection and model instability are most interesting from an *ex ante* perspective. That is, the relevant questions are whether we can identify the appropriate predictor variables and detect structural breaks in regression parameters in real time, and how these may affect investment decisions. Answering these questions is the purpose of this section.

3.4.1 A utility-based performance measure

Several papers consider the effects of either parameter uncertainty, model uncertainty or model instability on optimal asset allocation decisions, see Kandel and Stambaugh (1996), Barberis (2000), Avramov (2002) and Pettenuzzo and Timmermann (2005). Most of these analyses focus on horizon effects, that is the issue how uncertainty about the relevant predictor variables or the possibility of structural breaks changes the decisions of investors with different horizons, typically ranging from a single month up to ten years. Here we only consider an active short-term investor, with an investment horizon of one month.[6] The investor's portfolio consists of stocks and riskfree bonds only. At the start of each month $T+1$, the investor decides upon the fraction of her portfolio to be invested in stocks w_{T+1}, based upon a forecast of the excess stock return r_{T+1}. The investor is assumed to maximize a power utility function with coefficient of relative risk aversion γ:

$$u(W_{T+1}) = \frac{W_{T+1}^{1-\gamma}}{1-\gamma}, \qquad \gamma > 0, \tag{3.16}$$

where W_{T+1} is the wealth at the end of period $T+1$, which is equal to

$$W_{T+1} = W_T((1-w_{T+1})\exp(r_{f,T+1}) + w_{T+1}\exp(r_{f,T+1}+r_{T+1})), \tag{3.17}$$

where W_T denotes initial wealth, and where $r_{f,T+1}$ is the riskfree rate.

[6] It is relatively straightforward to extend our analysis to the case of a long-horizon buy-and-hold investor, who solves the asset allocation problem only once at the start of the investment period. Things become more involved in case of dynamic asset allocation, that is when the long-run investor is allowed to rebalance her portfolio during the investment period, adjusting the portfolio weights to reflect new information that arrives. Solving the resulting dynamic programming problem is complicated due to the large number of state variables that enter the problem in a highly nonlinear way, see Barberis (2000) and Guidolin and Timmermann (2005b).

Without loss of generality we set initial wealth equal to one, $W_T = 1$, such that the investor's optimization problem is given by

$$\max_{w_{T+1}} E_T(u(W_{T+1})) = \max_{w_{T+1}} E_T\left(\frac{((1-w_{T+1})\exp(r_{f,T+1}) + w_{T+1}\exp(r_{f,T+1}+r_{T+1}))^{1-\gamma}}{1-\gamma}\right), \tag{3.18}$$

where E_T is the conditional expectation given information at time T. How this expectation is computed depends on the treatment of model uncertainty and model instability by the investor. Consider the most general case, both allowing for uncertainty concerning which predictor variables to include and allowing for the possibility of structural breaks in the regressions parameters, as given by model (3.5) with (3.4). The marginal predictive density for future excess stock returns $p(r_{T+1}|r, x, x_{T+1})$ in (3.15) should then be used to derive the proportion of the portfolio allocated to stocks according to (3.18). That is, the investor solves the following problem:

$$\max_{w_{T+1}} \int u(W_{T+1}) p(r_{T+1}|r, x, x_{T+1}) dr_{T+1}. \tag{3.19}$$

The integral in (3.19) is approximated by generating G independent draws $\{r^g_{T+1}\}^G_{g=1}$ from the predictive density $p(r_{T+1}|r, x, x_{T+1})$, and then using a numerical optimization method to maximize the quantity:

$$\frac{1}{G}\sum_{g=1}^{G}\left(\frac{((1-w_{T+1})\exp(r_{f,T+1}) + w_{T+1}\exp(r_{f,T+1}+r^g_{T+1}))^{1-\gamma}}{1-\gamma}\right) \tag{3.20}$$

Three further cases are included in the empirical analysis below. First, we consider an investor who incorporates model uncertainty but ignores the possibility of structural breaks in the regression parameters. This investor obtains a forecast of the excess stock return r_{T+1} from model (3.2). Second, we consider the reverse case of an investor who allows for the possibility of structural breaks in the regression parameters but ignores model uncertainty, thus using excess return forecasts from model (3.3). Third, we consider an investor who is ignorant about both model uncertainty and structural breaks, and simply includes all available predictor variables in the model assuming constant coefficients, effectively using the benchmark model (3.1) for return forecasting.[7]

As explained by Barberis (2000), the weight w_{T+1} in (3.17) cannot be left unconstrained in the optimization problem (3.18) as expected utility would be equal to $-\infty$ in that case. We consider the following two restrictions on w_{T+1}. First, we restrict

[7]We also consider a return forecasting model with the dividend yield as the only predictor variable, as in Barberis (2000). This renders a portfolio with worse performance than the model with all variables included and therefore detailed results are not reported here.

$w_{T+1} \in [-1, 2]$, allowing short-sales and leveraging of the portfolio to some extent. Second, we do not allow for short-sales or leveraging at all, by constraining w_{T+1} to be in the [0,1] interval.

In sum, in total we consider eight active investment strategies. The strategies that are based on excess return forecasts from models (3.1), (3.2), (3.3), and (3.5) are denoted as Linear, BMA (Bayesian Model Averaging), SB (Structural Break), and BMASB (Bayesian Model Averaging with Structural Breaks), respectively. The strategies without the possibility of short-selling and leveraging are indicated by means of the addition (0,1). For comparison, we include three static benchmark strategies: I) holding stocks only, II) holding a portfolio consisting of 50% stocks and 50% bonds, and III) holding bonds only.

We evaluate the different investment strategies by computing the *ex post* utility levels substituting the realized return of the portfolios at time $T+1$ in (3.16). Total utility is then obtained as the sum of $u(W_{T+1})$ across all T^* investment periods $T = T_0+1, \ldots, T_0+T^*$, where the first investment decision is made at the end of period T_0. In order to compare two alternative strategies we compute the return that equates their average utilities. For example, suppose we compare the strategy based on excess return forecasts from the benchmark model (3.1) with a fixed set of predictor variables and constant regression parameters to the strategy based on the general model (3.5) with (3.4) that incorporates model uncertainty and structural breaks. The wealth provided at time $T+1$ by the two resulting portfolios is denoted as $W_{A,T+1}$ and $W_{B,T+1}$, respectively. We then determine the value of Δ such that

$$\sum_{T=T_0}^{T_0+T^*-1} u(W_{A,T+1}) = \sum_{T=T_0}^{T_0+T^*-1} u(W_{B,T+1}/\exp(\Delta)). \tag{3.21}$$

Following Fleming *et al.* (2001), we interpret Δ as the maximum performance fee the investor would be willing to pay to switch from strategy A to strategy B. In that sense, Δ represents the economic value of model uncertainty and model instability in the example above. For comparison of multiple investment strategies, it is useful to note that the performance fee an investor is willing to pay to switch from strategy A to strategy B can also be computed as the difference between the performance fees of these strategies with respect to a third strategy C.[8] We use this property below to infer the added value of the different components of our model, that is model uncertainty, break uncertainty, and parameter uncertainty.

Finally, the portfolio weights in the active investment strategies change every month,

[8]This follows from the fact that combining (3.21) for the comparisons of strategies A and B with C, $\sum_T u(W_{C,T+1}) = \sum_T u(W_{A,T+1}/\exp(\Delta_A))$ and $\sum_T u(W_{C,T+1}) = \sum_T u(W_{B,T+1}/\exp(\Delta_B))$, gives $\sum_T u(W_{A,T+1}/\exp(\Delta_A)) = \sum_T u(W_{B,T+1}/\exp(\Delta_B))$. Using the power utility specification in (3.16), this can be rewritten as $\sum_T u(W_{A,T+1}) = \sum_T u(W_{B,T+1}/\exp(\Delta_B - \Delta_A))$.

and the portfolio must be rebalanced accordingly. Hence, transaction costs play a non-trivial role and should be taken into account when evaluating the relative performance of different strategies. Rebalancing the portfolio at the start of month $T+1$ means that the weight invested in stocks is changed from w_T to w_{T+1}. We assume that transaction costs amount to a fixed percentage c on each traded dollar. Setting the initial wealth W_T equal to 1 for simplicity, transaction costs at time $T+1$ are equal to

$$c_{T+1} = 2c|w_{T+1} - w_T| \tag{3.22}$$

where the multiplication by 2 follows from the fact that the investor rebalances her investments in both stocks and bonds. The net portfolio return is then given by $r_{T+1} - c_{T+1}$. We apply two scenarios with transaction costs of 0.1% and 0.3%. Note that for a passive strategy the inclusion of transaction costs matters only in setting up the portfolio at time T_0.

3.4.2 Empirical Results

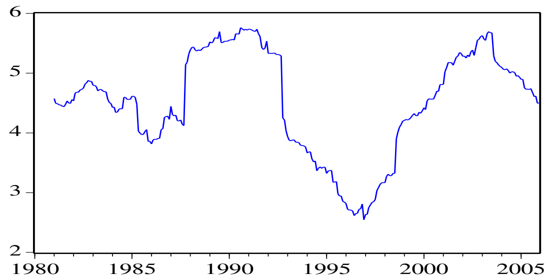
The analysis for the active investment strategies is implemented for the period from January 1976 until December 2005, involving $T^* = 360$ one month ahead excess stock return forecasts. The models are estimated recursively using an expanding window of observations, starting with the first $T_0 = 120$ months to estimate the initial models that are used to obtain the first return prediction. The investment strategies are implemented for two levels of relative risk aversion, $\gamma = 5$ and 10. Before we analyze the performance of the different portfolios, we summarize the statistical accuracy of the excess return forecasts.

Statistical accuracy of excess return forecasts

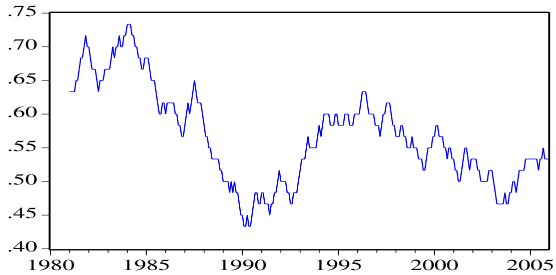
The forecasts obtained from the model allowing for uncertainty concerning which predictor variables to include and allowing for the possibility of structural breaks in the regressions parameters (3.5) with (3.4) have mean error (ME) of 0.50% and a root mean square prediction error (RMSPE) of 4.36%. This is slightly more accurate than the linear, BMA and SB forecasting models, all of which have RMSPEs equal to 4.41%. We note that the predictive regression models considerably improve upon a random walk forecast, which leads to an RSMPE of 6.15%. Hence, there is some predictability in the excess stock returns, but the improvements in statistical forecast accuracy due to more elaborate model specifications is relatively minor.

Figure 3.2 shows five-year moving averages of the excess returns' RMSPE and the hit ratio, defined as the proportion of correctly predicted signs of the excess stock return.

Figure 3.2: Statistical Accuracy of Excess Return Forecasts



(a) RMSPE



(b) Hit ratio

Note: The figure presents five-year moving averages of the RMSPE and of the hit ratio, for excess stock return forecasts obtained from the predictive regression model (3.5) allowing for model uncertainty and structural breaks in the regression parameters.

Both graphs show that the model performs quite well until October 1987, with RMSPE varying between 4% and 5% and the hit ratio between 0.6 and 0.75. The stock market crash causes a sizable upward jump in the RMSPE, and marks the beginning of a period with less accurate forecasts and a sharp decline in the hit ratio to 0.43 for the period 1986-1990. Forecast accuracy improves again considerably during the period 1991-1996 with the RMSPE reaching a low of 2.5% and the hit ratio peaking at 0.63 for the period January 1992 - December 1996. Predictability of the excess stock returns then deteriorates again due to the crises in Asia and Russia, and the internet bubble and its burst, with the hit ratio dropping below 0.5 again in 2003. During the final two years of the out-of-sample period the forecast performance of the model appears to improve again. In sum, the predictive accuracy varies considerably over time, even when a flexible forecast approach allowing for structural breaks and model uncertainty is employed.

Asset allocation and portfolio performance

Table 3.5 provides performance measures for the 11 different investment strategies considered, ignoring transaction costs for the moment. We report the ex-post utility levels and performance fees Δ relative to the three buy-and-hold portfolios, denoted as Δ_s, Δ_m and Δ_b for the stock, mixed, and bond portfolios, respectively. In addition, we consider traditional performance measures including the annualized mean and standard deviation of portfolio returns, and the Sharpe ratio (computed as the ratio of the mean monthly excess return on the portfolio and the monthly standard deviation of the portfolio return).

For the benchmark passive strategies, we observe that over the complete investment period from January 1976 until December 2005, the average annualized return on the stock portfolio is 12% with an estimated unconditional standard deviation of 15%, while the bond portfolio provides a mean return of 5.9% with a standard deviation of 0.9%. The Sharpe ratio of the stock portfolio is 0.117, while for the bond portfolio it is zero by construction. In terms of utility levels, the mixed buy-and-hold portfolio consisting of 50% stocks and 50% bonds renders the best results.

Next, consider the active investment strategy based on excess return forecasts that account for model uncertainty and structural breaks, which allows for limited short-selling and leveraging (Strategy IV: BMASB). Compared to the buy-and-hold stock portfolio, the average return increases by 1.4% for the investor with low risk aversion ($\gamma = 5$) while it decreases by 0.8% for the investor with high risk aversion ($\gamma = 10$). At the same time, portfolio risk is reduced considerably as well for the high risk averse investor, while it does not increase dramatically for the low risk averse investor. Consequently, both types of investors achieve a higher Sharpe ratio of around 0.14. The benefits of the active investment strategy also are shown convincingly by the estimates of the performance fee Δ. The investor with low risk aversion would be willing to pay 74 basis points to switch from the passive stock portfolio to the active strategy, while not surprisingly she is even more eager to switch from the passive bond portfolio, with a performance fee of 211 basis points. Due to her high risk aversion, the investor with $\gamma = 10$ is close to being indifferent between the buy-and-hold bond portfolio and the active strategy with Δ_b equal to 22 basis points. If this investor is forced to hold stocks though, she would rather have her portfolio managed actively, offering a performance fee of no less than 583 basis points. Although the active strategies do render a higher Sharpe ratio than the mixed buy-and-hold portfolio, both investors prefer this passive strategy in terms of utility levels, thus leading to negative performance fees of -10 and -52 basis points.

Eliminating the possibility of short-sales and leverage by restricting the portfolio weight w_{T+1} to lie between 0 and 1 (Strategy VIII: BMASB (0,1)) further improves the performance. Not surprisingly, the average returns decline compared to the unrestricted portfolios, by approximately 2% per year, but this is more than compensated for by the reduction in volatility. The restricted portfolios render return standard deviations of 8.3% and 6.6% for $\gamma = 5$ and 10, respectively, compared to 15.2% and 11% for their unrestricted

Table 3.5: Active portfolio performance - No transaction costs

Strategy		Mean	St dev	SR (×100)	Utility	Δ_s	Δ_m	Δ_b
Panel A: $\gamma = 5$								
I:	100% stocks	11.98	14.99	11.72	−87.85			
II:	50% stocks	8.94	7.51	11.70	−87.61			
III:	0% stocks	5.89	0.89	0.00	−88.25			
IV:	BMASB	13.38	15.18	14.24	−87.63	74.0	−9.9	211.1
V:	BMA	6.50	1.69	10.36	−88.06	−70.7	−154.6	66.4
VI:	SB	10.77	14.70	9.57	−88.33	−162.8	−246.7	−25.7
VII:	Linear	7.29	5.37	7.51	−87.99	−48.5	−132.4	88.6
VIII:	BMASB (0,1)	11.05	8.31	17.92	−87.12	249.5	165.7	386.7
IX:	BMA (0,1)	6.50	1.69	10.36	−88.06	−70.7	−154.6	66.4
X:	SB (0,1)	9.45	8.06	12.74	−87.57	97.7	13.9	234.9
XI:	Linear (0,1)	7.35	4.57	9.16	−87.91	−19.5	−103.3	117.7
Panel A: $\gamma = 10$								
I:	100% stocks	11.98	14.99	11.72	−39.92			
II:	50% stocks	8.94	7.51	11.70	−38.07			
III:	0% stocks	5.89	0.89	0.00	−38.28			
IV:	BMASB	11.18	11.00	13.89	−38.22	582.5	−51.7	22.0
V:	BMA	6.81	8.89	2.96	−39.12	272.1	−362.2	−288.5
VI:	SB	8.45	6.26	11.79	−38.01	653.8	19.5	93.2
VII:	Linear	6.60	2.85	7.11	−38.16	603.9	−30.4	43.3
VIII:	BMASB (0,1)	9.27	6.56	14.85	−37.82	723.3	89.0	162.8
IX:	BMA (0,1)	6.20	1.13	7.69	−38.19	593.5	−40.7	33.0
X:	SB (0,1)	9.43	10.28	9.92	−38.70	414.3	−219.9	−146.2
XI:	Linear (0,1)	6.68	2.51	9.03	−38.10	623.2	−11.0	62.7

Note: The table presents performance measures for active investment strategies based on one-month excess return forecasts of the S&P 500 index over the period January 1976 - December 2005 for investors with power utility function with risk aversion γ. BMASB, BMA, SB, and Linear denote strategies based on excess return forecasts from models (3.5), (3.2), (3.3), and (3.1), respectively. The addition (0,1) indicates that the portfolio weights obtained from (3.20) are restricted between 0 and 1, such that short-selling and leveraging are not allowed. Otherwise, the portfolio weights are restricted to lie between −1 and 2. Strategies I-III are benchmark passive strategies: I) holding stocks only, II) holding a mixed portfolio consisting of 50% stocks and 50% bonds, and III) holding bonds only. The table reports the average portfolio return and standard deviation (both in annualized percentage points), the Sharpe ratio (SR), and utility (computed using (3.16)). The three rightmost columns present the annualized return in basis points that an investor is willing to give up to switch from the passive stock (s), mixed (m), or bond (b) strategy to the active strategy.

counterparts. The unrestricted strategies sometimes use extreme portfolio weights that makes them relatively risky, see also Jagannathan and Ma (2003) and Marquering and Verbeek (2004). Hence, despite the sacrifice in terms of average return, the Sharpe ratios of the restricted portfolios are considerably higher at 0.179 and 0.149. The performance fees relative to the buy-and-hold stock and bond portfolios increase accordingly. Furthermore, the passive mixed portfolio is outperformed as well by the restricted active strategies, with estimates of Δ_m equal to 165 and 89 basis points for the investor with low and high risk aversion, respectively.

The added value of incorporating the different sources of uncertainty in the return forecasting model can be gauged by examining the performance of the other active investment strategies based on excess return forecasts from more restricted models. Here we focus on Strategies VIII-XI with portfolio weights restricted to be in the [0,1] interval. The comparison is facilitated by recalling that the performance fee an investor is willing to pay to switch from a certain strategy A to another strategy B is equal to the difference between the performance fees of these strategies with respect to a third strategy C. For example, given that the investor with low risk aversion is willing to pay 74 and 211 basis points to switch from the passive stock and bond portfolios to Strategy IV, it follows that she would offer a performance fee of 137 ($= 211 - 74$) basis points to switch from the bond portfolio to the stock portfolio.

From Table 3.5 we observe that allowing for parameter uncertainty only is not sufficient to convince the low risk averse investor to switch from the static stock or mixed portfolio to a dynamic portfolio, given the negative performance fees for the active strategy based on return forecasts from the linear model (3.1) that includes all predictor variables. Incorporating model uncertainty (Strategy IX) makes the active strategy even less attractive, as the estimates of Δ_s and Δ_m are lower than for Strategy XI. Allowing for structural breaks does help, leading to a higher Sharpe ratio and positive performance fees of $\Delta_s = 98$ and $\Delta_m = 14$ basis points for Strategy X. However, these are still more than 150 basis points lower than the corresponding performance fees for the strategy based on return forecasts from the general model that allows for both structural breaks and model uncertainty. Hence, once structural breaks in the predictive regression model are accounted for, it is worthwhile to allow for model uncertainty as well. Comparing the return and risk characteristics of Strategies VIII and X, we find that the main difference occurs in the mean portfolio return, which is 1.6% higher for Strategy VIII, while their standard deviations are similar at just over 8%.

The complementarity of structural breaks and model averaging is perhaps even more apparent from the results for the high risk averse investor. This investor does prefer an active strategy over passive stock or bond portfolios, even if only parameter uncertainty is incorporated in the excess return forecasts, as Strategy XI renders positive performance fees of $\Delta_s = 623$ and $\Delta_b = 63$ basis points. At the same time, this investor does not value model uncertainty or uncertainty about structural breaks in the regression coefficients positively when these features are considered in isolation. In fact, she would be willing to pay 30 basis points to switch from Strategy IX to Strategy XI in order to avoid model averaging, while the performance fee for eliminating structural breaks from consideration by switching from Strategy X to XI is no less than 210 basis points. The investor's negative opinion about these features of the return forecasting model is completely reversed though when offered the possibility to incorporate both. The performance fees of Strategy VIII based on excess return forecasts from the general model (3.5) with (3.4) are equal to $\Delta_s = 723$ and $\Delta_b = 163$, exactly 100 basis points higher than the corresponding estimates for Strategy XI.

Sub-period analysis

Figure 3.2 suggests that the accuracy of the excess return forecasts varies considerably over time. How this affects the performance of the active strategies can be seen from Table 3.6, which shows performance statistics for three sub-periods each covering a decade for the investor with high relative risk aversion ($\gamma = 10$).[9] We limit our discussion here to the restricted active portfolios (Strategies VIII-XI), given their superior performance relative to the unrestricted portfolios over the complete sample period, as discussed before. The performance of the portfolio that results from forecasts of the general model allowing for model uncertainty and structural breaks (Strategy VIII) is quite impressive during the first decade of the investment period, from January 1976 until December 1985, with a Sharpe ratio of 0.239, which is more than double the Sharpe ratios of the passive portfolios held in Strategies I and II. This positive result is due to the fact that the mean return of the active strategy during this period is actually higher than the mean return of the buy-and-hold portfolio (15.5% compared to 13.4%), while volatility is reduced by about 40%. The corresponding performance fees are positive and large. Also note that accounting for structural breaks during this sub-period seems crucial, in the sense that the portfolio based on returns forecast from the structural break model (3.3) (Strategy IX) renders even higher values of all performance measures, including the performance fee Δ_s. Model averaging (Strategy X) adds much less value, as the performance fee against the passive stock portfolio is only half as large as that of the active portfolios that account for structural breaks (415 basis points compared to 804 and 856 basis points for Strategies VIII and IX, respectively).

[9]Sub-sample results for the investor with low relative risk aversion ($\gamma = 5$) are qualitatively similar and therefore not shown to save space. Detailed results are available upon request.

Table 3.6: Active portfolio performance - Subperiod results, $\gamma = 10$

Strategy		Mean	St dev	SR (×100)	Utility	Δ_s	Δ_m	Δ_b
Panel A: 1976:1 - 1985:12								
I:	100% stocks	13.41	13.98	9.76	−12.86			
II:	50% stocks	11.04	6.96	9.81	−12.45			
III:	0% stocks	8.68	0.86	0.00	−12.50			
IV:	BMASB	22.48	15.06	26.47	−12.02	898.8	464.2	514.3
V:	BMA	4.41	10.93	−11.27	−13.61	−751.1	−1185.7	−1135.6
VI:	SB	13.91	7.34	20.57	−12.20	699.3	264.8	314.9
VII:	Linear	11.22	3.53	20.75	−12.30	590.7	156.1	206.2
VIII:	BMASB (0,1)	15.49	8.23	23.89	−12.11	803.7	369.1	419.2
IX:	BMA (0,1)	8.95	1.00	7.94	−12.47	414.9	−19.7	30.5
X:	SB (0,1)	20.20	12.91	25.77	−12.06	856.0	421.5	471.6
XI:	Linear (0,1)	10.72	2.87	20.58	−12.33	564.7	130.2	180.3
Panel B: 1986:1 - 1995:12								
I:	100% stocks	13.85	15.31	15.90	−13.38			
II:	50% stocks	9.64	7.66	15.89	−12.64			
III:	0% stocks	5.42	0.50	0.00	−12.80			
IV:	BMASB	7.61	8.26	7.63	−12.94	444.9	−313.9	−144.2
V:	BMA	11.11	10.17	16.16	−12.62	786.2	27.3	197.0
VI:	SB	6.98	4.99	9.02	−12.77	625.8	−133.1	36.6
VII:	Linear	5.22	1.97	−2.98	−12.84	556.0	−202.9	−33.2
VIII:	BMASB (0,1)	7.98	5.19	14.22	−12.68	718.8	−40.1	129.7
IX:	BMA (0,1)	5.85	0.88	14.08	−12.76	634.9	−124.0	45.7
X:	SB (0,1)	5.03	9.40	−1.21	−13.37	7.7	−751.2	−581.5
XI:	Linear (0,1)	5.77	1.78	5.59	−12.78	618.0	−140.9	28.9
Panel C: 1996:1 - 2005:12								
I:	100% stocks	8.68	15.72	9.37	−13.68			
II:	50% stocks	6.13	7.88	9.35	−12.98			
III:	0% stocks	3.58	0.51	0.00	−12.98			
IV:	BMASB	3.46	7.28	−0.48	−13.25	424.9	−277.3	−275.0
V:	BMA	4.89	3.61	10.50	−12.90	788.5	86.2	88.6
VI:	SB	4.47	5.96	4.29	−13.04	638.4	−63.8	−61.5
VII:	Linear	3.36	2.32	−2.80	−13.02	663.2	−39.0	−36.7
VIII:	BMASB (0,1)	4.35	5.47	4.04	−13.03	652.5	−49.7	−47.4
IX:	BMA (0,1)	3.78	0.97	5.94	−12.96	722.8	20.5	22.9
X:	SB (0,1)	3.05	6.99	−2.22	−13.27	409.8	−292.5	−290.2
XI:	Linear (0,1)	3.55	2.30	−0.44	−13.00	683.6	−18.7	−16.4

Note: The table presents performance measures for active investment strategies based on one-month excess return forecasts of the S&P 500 index over three 10-year sub-periods for investors with power utility function with risk aversion $\gamma = 10$. See Table 3.5 for a description of the investment strategies and evaluation criteria.

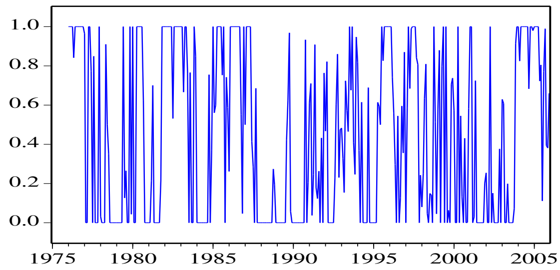
The importance of structural breaks and model uncertainty seems to be completely reversed during the second decade of the investment period, from January 1986 until December 1995. For this sub-period we find an estimate of $\Delta_s = 634$ basis points for the active BMA-based Strategy X, compared to only 8 basis points for the SB-based Strategy IX. Accounting for both features still pays off though, in the sense that the BMASB Strategy VIII renders a higher performance fee of 719 basis points against the passive stock portfolio. While the passive portfolios consisting purely of stock or bonds are still outperformed by the active strategies, this no longer holds for the mixed buy-and-hold portfolio. The return of this passive strategy is considerably higher at 9.64%, which is sufficient to render higher Sharpe ratios (despite higher volatility) and a higher level utility resulting in negative performance fees Δ_m.

The active strategies' performance further declines during the third and final decade of the investment period, from January 1996 until December 2005. Although the reduction in portfolio returns' volatility is of the same magnitude as before (or even larger), the loss in terms of average return is considerably larger. This results in a Sharpe ratio of 0.041 for the BMASB-based Strategy VIII, compared to 0.094 for the passive stock and mixed portfolios. It is quite remarkable then that the active strategy still achieves higher utility than the buy-and-hold stock portfolio (-13.03 compared to -13.68). The mixed passive portfolio in turn renders higher utility than the active strategy, resulting again in a negative performance fee Δ_m. Note that the same now holds for the passive bond portfolio. In sum, it seems that the performance of the active strategies has gradually worsened over time.

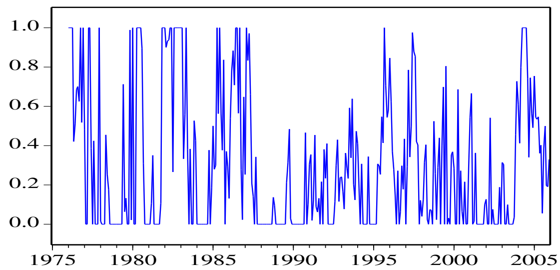
Transaction costs

Our analysis of the active investment strategies so far has ignored transaction costs. Obviously, their effects on the strategies' performance crucially depends on the average absolute change in portfolio weights, see (3.22). Figure 3.3 shows the portfolio weight for stocks in the restricted portfolios based on excess return forecasts from the general model, allowing for model uncertainty and structural breaks in the regression parameters (Strategy VIII). First of all, Figure 3.3 clearly demonstrates the effects of risk aversion on the asset allocation, in the sense that the weight for stocks in the portfolio of the investor with high risk aversion is systematically lower than for the investor with low risk aversion. The average stock weight is equal to 0.44 and 0.30 for $\gamma = 5$ and $\gamma = 10$, respectively. Second, although there are extended periods of time when the investment in stocks is at high or low levels, month-to-month variation in the portfolio composition is quite substantial. The standard deviation of the stock weight is equal to 0.43 and 0.35, while the average absolute change $|w_{T+1} - w_T|$ is equal to 0.27 and 0.20 for $\gamma = 5$

Figure 3.3: Stock portfolio weights in restricted portfolios (Strategy VII)



(a) $\gamma = 5$



(b) $\gamma = 10$

Note: The figure presents the portfolio weight for stocks in the restricted portfolios based on excess stock return forecasts from the predictive regression model (3.5) allowing for model uncertainty and structural breaks in the regression parameters (Strategy VIII).

and $\gamma = 10$, respectively. Hence, a proper analysis of the effects of transaction costs is warranted.

Tables 3.7 and 3.8 present results for the complete 30-year investment period for low (0.1%) and moderate (0.3%) levels of transaction costs, respectively. The presence of transaction costs obviously hurts the active strategies' performance. However, for both levels of transaction costs, the restricted portfolio based on return forecasts from the general model continues to outperform the buy-and-hold portfolios, although the performance fees Δ become somewhat lower. It should also be noted that for moderate levels of transaction costs, the mixed buy-and-hold portfolio renders a higher Sharpe ratio than the active portfolios, although utility is still lower.

Table 3.7: Active portfolio performance - 0.1% transaction costs

Strategy		Mean	St dev	SR (×100)	Utility	Δ_s	Δ_m	Δ_b
Panel A: $\gamma = 5$								
I:	100% stocks	11.98	14.99	11.71	−87.85			
II:	50% stocks	8.93	7.51	11.69	−87.61			
III:	0% stocks	5.89	0.89	0.00	−88.25			
IV:	BMASB	11.92	15.15	11.48	−87.85	2.3	−81.7	139.5
V:	BMA	4.49	1.69	−24.04	−88.07	−72.8	−156.8	64.4
VI:	SB	9.31	14.67	6.72	−88.54	−234.7	−318.7	−97.5
VII:	Linear	5.38	5.37	−2.78	−88.05	−66.1	−150.0	71.2
VIII:	BMASB (0,1)	10.42	8.30	15.72	−87.22	217.9	134.0	355.2
IX:	BMA (0,1)	6.45	1.69	9.53	−88.07	−72.8	−156.8	64.4
X:	SB (0,1)	8.82	8.06	10.49	−87.66	67.8	−16.1	205.1
XI:	Linear (0,1)	7.12	4.56	7.75	−87.94	−30.5	−114.4	106.8
Panel A: $\gamma = 10$								
I:	100% stocks	11.98	14.99	11.71	−39.92			
II:	50% stocks	8.93	7.51	11.69	−38.07			
III:	0% stocks	5.89	0.89	0.00	−38.28			
IV:	BMASB	10.22	10.96	11.38	−38.35	535.7	−98.6	−24.6
V:	BMA	5.82	8.85	−0.23	−39.26	224.3	−410.0	−336.1
VI:	SB	7.95	6.24	9.50	−38.08	629.0	−5.3	68.7
VII:	Linear	6.42	2.84	5.30	−38.18	595.1	−39.2	34.8
VIII:	BMASB (0,1)	8.78	6.54	12.74	−37.88	699.3	65.0	139.0
IX:	BMA (0,1)	6.17	1.13	7.03	−38.11	592.5	−41.8	32.2
X:	SB (0,1)	8.43	10.24	7.14	−38.84	365.7	−268.6	−194.6
XI:	Linear (0,1)	6.56	2.49	7.70	−38.12	617.6	−16.7	57.2

Note: The table presents performance measures for active investment strategies based on one-month excess return forecasts of the S&P 500 index over the period January 1976 - December 2005 for investors with power utility function with risk aversion γ. Transaction costs are set equal to 0.1%. See Table 3.5 for a description of the investment strategies. The table reports the average portfolio return and standard deviation (both in annualized percentage points), the Sharpe ratio (SR), and utility (computed using (3.16)). The three rightmost columns present the annualized return in basis points that an investor is willing to give up to switch from the passive stock (s), mixed (m), or bond (b) strategy to the active strategy.

Table 3.8: Active portfolio performance - 0.3% transaction costs

Strategy		Mean	St dev	SR (×100)	Utility	Δ_s	Δ_m	Δ_b
Panel A: $\gamma = 5$								
I:	100% stocks	11.97	14.99	11.70	−87.85			
II:	50% stocks	8.93	7.50	11.67	−87.61			
III:	0% stocks	5.88	0.89	0.00	−88.26			
IV:	BMASB	9.00	15.11	5.93	−88.27	−141.7	−225.7	−4.2
V:	BMA	0.46	1.92	−81.73	−88.08	−77.0	−161.1	60.5
VI:	SB	6.40	14.66	1.00	−88.97	−379.1	−463.1	−241.6
VII:	Linear	1.55	5.44	−23.05	−88.15	−101.3	−185.4	36.2
VIII:	BMASB (0,1)	9.14	8.29	11.30	−87.40	154.7	70.6	292.2
IX:	BMA (0,1)	6.35	1.69	7.87	−88.08	−77.0	−161.1	60.5
X:	SB (0,1)	7.56	8.07	5.97	−87.84	7.9	−76.2	145.4
XI:	Linear (0,1)	6.67	4.54	4.91	−88.01	−52.6	−136.7	84.9
Panel A: $\gamma = 10$								
I:	100% stocks	11.97	14.99	11.70	−39.93			
II:	50% stocks	8.93	7.50	11.67	−38.07			
III:	0% stocks	5.88	0.89	0.00	−38.28			
IV:	BMASB	8.28	10.90	6.32	−38.62	441.7	−192.6	−118.2
V:	BMA	3.86	8.83	−6.66	−39.55	127.3	−507.0	−432.6
VI:	SB	6.94	6.22	4.86	−38.23	579.2	−55.1	19.4
VII:	Linear	6.05	2.80	1.61	−38.23	577.6	−56.7	17.7
VIII:	BMASB (0,1)	7.80	6.51	8.44	−38.02	651.2	16.9	91.3
IX:	BMA (0,1)	6.12	1.13	5.71	−38.13	590.5	−43.8	30.6
X:	SB (0,1)	6.43	10.18	1.51	−39.13	267.9	−366.4	−292.0
XI:	Linear (0,1)	6.32	2.47	4.99	−38.15	606.2	−28.1	46.3

Note: The table presents performance measures for active investment strategies based on one-month excess return forecasts of the S&P 500 index over the period January 1976 - December 2005 for investors with power utility function with risk aversion γ. Transaction costs are set equal to 0.3%. See Table 3.5 for a description of the investment strategies. The table reports the average portfolio return and standard deviation (both in annualized percentage points), the Sharpe ratio (SR), and utility (computed using (3.16)). The three rightmost columns present the annualized return in basis points that an investor is willing to give up to switch from the passive stock (s), mixed (m), or bond (b) strategy to the active strategy.

3.5 Conclusion

Optimal portfolio decisions force investors to make a number of important decisions concerning the return forecasting model used. These decisions involve in particular the treatment of different sources of uncertainty, about the relevant predictor variables (model uncertainty), the values of the regression parameters (parameter uncertainty), and their stability (structural breaks). In this Chapter we have developed a framework to incorporate all three sources of uncertainty simultaneously. This extends previous research allowing for either parameter uncertainty and model uncertainty (Avramov (2002); Cremers (2002)), or parameter uncertainty and parameter instability (Pesaran *et al.* (2006)).

Our empirical results suggest, first, that over the period 1966-2005 several structural breaks occurred in the relationship between US stock returns and predictor variables such as the dividend yield and interest rates. These changes appear to be caused by important events such as the oil crisis, changes in monetary policy, and the October 1987 stock market crash. Second, we find that allowing for model uncertainty and structural breaks jointly has considerable economic value. A typical investor would be willing to pay up to several hundreds of basis points annually to switch from a passive buy-and-hold strategy to an active strategy based on a return forecasting model that allows for model and parameter uncertainty as well as structural breaks in the regression parameters. The active strategy that incorporates all three sources of uncertainty performs considerably better than strategies based on more restricted return forecasting models.

Chapter 4

Predicting the Term Structure of Interest Rates

Incorporating parameter uncertainty, model uncertainty and macroeconomic information

4.1 Introduction

Modelling and forecasting the term structure of interest rates is by no means an easy endeavor. As long yields are risk-adjusted averages of expected future short rates, yields of different maturities are intimately related and therefore tend to move together, in the cross-section as well as over time. Long and short maturities are known to react quite differently, however, to shocks hitting the economy. Furthermore, monetary policy authorities such as the Federal Reserve are actively targeting the short end of the term structure to help achieve their macroeconomic goals. Many forces are at work at moving interest rates. Identifying these forces and understanding their impact is of crucial importance.

During the last decades significant progress has been made in modelling the term structure, which has come about mainly through the development of no-arbitrage factor models. The literature on these so-called affine models was kick-started by seminal papers of Vasicek (1977) and Cox *et al.* (1985), characterized by Duffie and Kan (1996) and classified by Dai and Singleton (2000)[1]. Affine models explain yields by a small number of latent factors that can be extracted from the panel of yields for different maturities and impose cross-equation restrictions which rule out arbitrage opportunities. Affine models, provided they are properly specified, have been shown to accurately fit the term structure, see for example Dai and Singleton (2000). The models are silent, however, about the links between the latent factors and macroeconomic forces.

The current term structure literature is actively progressing to resolve this missing link. Recent studies have yielded interesting approaches for studying the joint behavior of

[1]An excellent survey of issues involving the specification and estimation of affine models set in continuous time is Piazzesi (2003), whereas discrete models are discussed in Backus *et al.* (1998).

interest rates and macroeconomic variables. One approach that has been undertaken is to extend existing term structure models by adding in observed macroeconomic variables and to study their interactions with the latent factors. A key contribution to this strand of the literature is Ang and Piazzesi (2003) who were the first to extend a standard three-factor affine model with macroeconomic variables. Studies such as Bikbov and Chernov (2005), Kim and Wright (2005), Ang *et al.* (2006a), Dai and Philippon (2006) and DeWachter and Lyrio (2006) also include various macroeconomic variables and study their explanatory power for yield movements. Studies that take a more structural approach are, amongst others, Rudebusch and Wu (2003), Wu (2005) and Hordahl *et al.* (2006) who all combine a model for the macro economy with an arbitrage-free specification for the term structure. Moving away from the realm of no-arbitrage interest rate models to that of more ad-hoc models, in particular the Nelson and Siegel (1987) model, studies such as Diebold *et al.* (2006) and Mönch (2006b) also show that adding information that reflects the state of the economy is beneficial[2].

Whereas modelling interest rate movements over time is already a strenuous task, accurately *forecasting* future rates is an equally difficult challenge. Yields of all maturities are close to being non-stationary, which makes it hard for any model to outperform the simple random walk-based no-change forecast. Several studies have documented that beating the random walk is indeed difficult, in particular for unrestricted yields-only based vector autoregressive (VAR) and standard affine models, see Duffee (2002) and Ang and Piazzesi (2003). However, all does not seem lost as recently more favorable evidence for predictability of yields has been reported. Whereas Duffee (2002) shows that more flexible affine specifications[3] can beat the random walk, Krippner (2005) and Diebold and Li (2006) show that a dynamic Nelson-Siegel factor model forecasts particularly well. Results are even more promising with models that incorporate macroeconomic information. Ang and Piazzesi (2003) and Mönch (2006a) report improved forecasts for U.S. zero-coupon yields at various horizons using affine models augmented with principal component-extracted macro factors. Hordahl *et al.* (2006) report similar results for German zero-coupon yields.

In spite of the powerful advances in term structure modelling and forecasting, a number of issues regarding estimation and forecasting have sofar been left nearly unaddressed. This Chapter tries to fill in some of these gaps by investigating the relevance of parameter

[2]Macro variables mainly seem to help in capturing the dynamics of short rates. Modelling long-term bonds remains difficult, however. Dai and Philippon (2006) show that fiscal policy can account for some of the unexplained long rate dynamics whereas DeWachter and Lyrio (2006) show that long-run inflation expectations are important for modelling long-term bond yields.

[3]Duffee (2002) denotes his preferred class of models "essentially affine" by allowing risk premia to depend on the entire state vector instead of being a multiple of volatility which is the assumption in standard affine models. Ang and Piazzesi (2003) remark that the essentially affine risk premia are not linear in the state vector and that using linear risk premia results in better forecasts.

uncertainty and, in particular, model uncertainty. Especially for VAR and affine models, which are highly parameterized if we attempt to model the complete term structure, parameter uncertainty is likely to be substantial and should be accounted for. Regarding model uncertainty, when looking at the historical time series of (U.S.) interest rates we can easily identify subperiods across which yield dynamics are quite different. Likely reasons are for example the reigns of different Fed Chairmen, most notably that of Paul Volcker, or the strong decline in interest rate levels accompanied by a pronounced widening of spreads in the early 1990's and after the burst of the Internet bubble. It will be unlikely that any individual model is capable of consistently producing accurate forecasts in each of these subperiods. As we demonstrate below, the forecasting performance of various popular term structure models does indeed vary substantially over time. In these situations, combining forecasts yields diversification gains and can therefore be an attractive alternative to relying on forecasts from a single model.

In addition to these two focal points, we also further examine the use of macroeconomic diffusion indices in term structure models. Mönch (2006a,b) documents that using factors, extracted from a large *panel* of macro series instead of *individual* series works well, in both affine models and the Nelson-Siegel model. We extend the picture by examining the use of diffusion indices also in simpler AR and VAR models. To summarize, the aim of this Chapter is threefold and consists of examining (i) parameter uncertainty, (ii) model uncertainty and (iii) the use of macro diffusion indices.

We analyze these objectives in the following manner. Using a relatively long time-series of U.S. zero-coupon bond yields, we examine the forecasting performance of a range of models that have been used in the literature. We estimate each model and generate forecasts by applying frequentist maximum likelihood techniques as well as Bayesian techniques to gauge the effects of explicitly taking into account parameter uncertainty. Furthermore, we analyze each model both with and without macro factors to assess the benefits of adding macroeconomic information. Finally, after showing the instability of the forecasting performance of the different models through subsample analysis, we consider several forecast combination approaches.

Our results can be summarized as follows. For the out-of-sample period covering 1994-2003 we show that the predictive ability of individual models varies considerably over time, irrespective of using frequentist or Bayesian estimation methods. A prime example is the Nelson and Siegel (1987) model, which predicts interest rates accurately in the 1990s but rather poorly in the early 2000s. We find that models that incorporate macroeconomic variables seem more accurate in subperiods during which the future path of interest rates is more uncertain. This is especially the case for the early 2000s with the pronounced drop in interest rates and the widening of spreads. Models without macro information do particularly well in subperiods where interest rate dynamics are more stable. An example is the early 1990s, where these models outperform the random walk RMSPE by sometimes well over 30%.

That different models forecast well in different subperiods confirms ex-post that alternative model specifications play a complementary role in approximating the interest rate data generating process. This provides a strong claim for the use of forecast combination techniques as opposed to believing in a single model. Our forecast combination results confirm this conjecture. We show that combined forecasts are consistently more accurate than the random walk benchmark across maturities and subperiods. We find that combining individual models that incorporate macro factors using Bayesian estimation techniques works extremely well, especially when using a weighting scheme that takes into account relative historical performance using a long window of forecasts. We obtain the largest gains in forecast performance for long maturities where the forecast combinations outperform the random walk by sometimes as much as 20% and the best individual model by more than 10%.

The remainder of the Chapter is organized as follows. In Section 2 we discuss the set of U.S. Treasury yields we analyze, and we provide details about the panel of macro series that we employ to obtain our macro factors. We devote Section 3 to present the different models we use to construct forecasts. In Section 4 we discuss results of the individual models whereas in Section 5 we outline and discuss results of several forecasting combination schemes. Finally, in Section 6 we conclude. The Appendix provides details on the frequentist and Bayesian techniques that we use for estimating model parameters and for constructing forecasts.

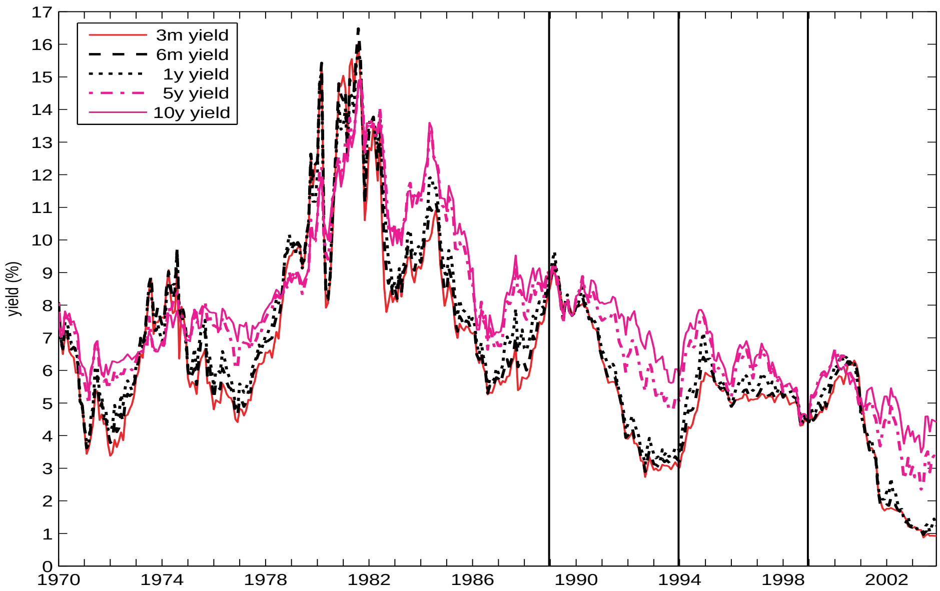
4.2 Data

4.2.1 Yield Data

The term structure data we use consists of end-of-month continuously compounded yields on U.S. zero-coupon bonds. These yields have been constructed from average bid-ask price quotes on U.S. Treasuries from the CRSP government bond files. CRSP filters the available quotes by taking out illiquid bonds and bonds with option features. The remaining quotes are used to construct forward rates using the Fama and Bliss (1987) bootstrap method as outlined in Bliss (1997). The forward rates are averaged to construct constant maturity spot rates[4]. Similar to Diebold and Li (2006) and Mönch (2006b), our dataset consists of unsmoothed Fama-Bliss yields. These unsmoothed yields exactly price

[4]We kindly thank Robert Bliss for providing us with the unsmoothed Fama-Bliss forward rates and the programs to construct the spot rates.

Figure 4.1: U.S. zero-coupon yields



Note: The figure shows time series plots for end-of-month U.S. zero coupon yields for a subset of maturities. The yields have been constructed using the Fama and Bliss (1987) bootstrap method. The sample period is January 1970 - December 2003 (408 observations). The vertical lines bound the three forecasting subsamples (1989:1 - 1993:12, 1994:1 - 1998:12 and 1999:1 - 2003:12).

Table 4.1: Summary statistics

maturity	mean	stdev	skew	kurt	min	max	JB	ρ_1	ρ_{12}	ρ_{24}
1-month	6.049	2.797	0.913	4.336	0.794	16.162	85.671	0.968	0.690	0.402
3-month	6.334	2.896	0.871	4.237	0.876	16.020	76.380	0.974	0.708	0.415
6-month	6.543	2.927	0.788	4.016	0.958	16.481	58.796	0.976	0.723	0.444
1-year	6.755	2.860	0.661	3.763	1.040	15.822	38.907	0.975	0.733	0.474
2-year	7.032	2.724	0.644	3.672	1.299	15.650	35.240	0.978	0.748	0.526
3-year	7.233	2.594	0.685	3.663	1.618	15.765	38.796	0.979	0.763	0.560
4-year	7.392	2.510	0.728	3.607	1.999	15.821	41.640	0.980	0.771	0.582
5-year	7.483	2.449	0.759	3.478	2.351	15.005	42.454	0.982	0.786	0.607
6-year	7.611	2.406	0.791	3.437	2.663	14.979	45.236	0.983	0.797	0.626
7-year	7.659	2.344	0.841	3.488	3.003	14.975	51.562	0.983	0.787	0.623
8-year	7.728	2.320	0.841	3.365	3.221	14.936	49.798	0.984	0.809	0.651
9-year	7.767	2.317	0.877	3.427	3.389	15.018	54.765	0.985	0.813	0.656
10-year	7.745	2.266	0.888	3.496	3.483	14.925	57.117	0.985	0.796	0.647

Note: The table shows summary statistics for our sample of end-of-month continuously compounded U.S. zero-coupon yields. Reported are the mean, standard deviation, skewness, kurtosis, minimum, maximum, the Jarque-Bera test statistic for normality and the 1^{st}, 12^{th} and 24^{th} sample autocorrelation. The results shown are for annualized yields (in %). The sample period is January 1970 - December 2003 (408 monthly observations).

the included U.S. Treasury securities. Smoothed yields on the other hand, which can be obtained by fitting a Nelson-Siegel curve on the unsmoothed yields (see Bliss, 1997 for details), do not have this property, and, moreover, using these may give the Nelson-Siegel model an unfair advantage over the other models in terms of fitting and forecasting the term structure.

Throughout our analysis we use yields with $N = 13$ different maturities of $\tau = 1$, 3 and 6 months and 1, 2,..., 10 years. We denote yields by $y^{(\tau_i)}$ for $i = 1, \ldots, N$. To estimate the Nelson-Siegel models we follow Diebold and Li (2006) and Diebold *et al.* (2006) by including additional maturities of 9, 15, 18, 21 and 30 months in order to increase the number of observations at the short end of the curve.

Our sample period covers January 1970 till December 2003 for a total of 408 monthly observations. Similar to Duffee (2002) and Ang and Piazzesi (2003) we include data from well before the Volcker period, despite the reservations expressed in Rudebusch and Wu (2003) that it is likely that the pricing of interest rate risk and the relationship between yields and macroeconomic variables have changed during such a long time span. We do so for two main reasons: (i) to have enough observations to sufficiently accurately identify the parameters of the models we consider, some of which are highly parameterized, and (ii) to assess forecasting performance over (sub-)periods with strikingly different characteristics.

Figure 4.1 shows time-series plots for a subsample of the 13 maturities whereas Table 4.1 reports summary statistics. The stylized facts common to yield curve data are clearly visible: the sample average curve is upward sloping and concave, volatility is decreasing with maturity, autocorrelations are very high and increasing with maturity and the null of normality is rejected due to positive skewness and excess kurtosis. Correlations between yields of different maturities are high, especially for close-together maturities. Even the maturities which are furthest apart (1 month and 10 years) still have a correlation of 86%.

4.2.2 Macroeconomic Data

Our macroeconomic dataset originates from Stock and Watson (2005) and consists of 116 series[5]. The macro variables are classified in 15 categories: (1) output and income, (2) employment and hours, (3) retail, (4) manufacturing and trade sales, (5) consumption, (6) housing starts and sales, (7) inventories, (8) orders, (9) stock prices, (10) exchange rates, (11) federal funds rate, (12) money and credit quantity aggregates, (13) price indexes, (14) average hourly earnings and (15) miscellaneous. Table 4.2 lists the series included in

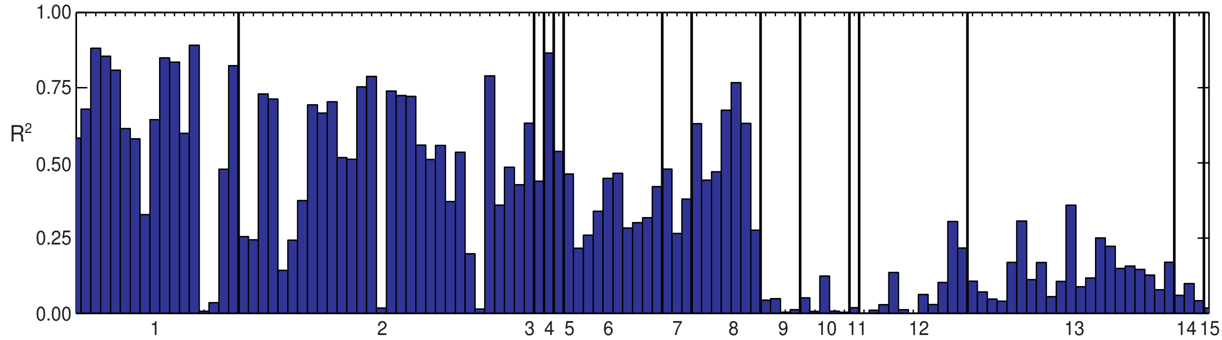
[5] We exclude all interest and spread related series from the original 132 series in the panel dataset (we discarded 16 series in total). We do include the federal funds rate because it closely follows the federal funds target rate. The latter is the key monetary policy instrument of the Federal Reserve. The federal funds rate will therefore be important for capturing the movements of (especially) the short end of the term structure.

Table 4.2: Macro-economic series

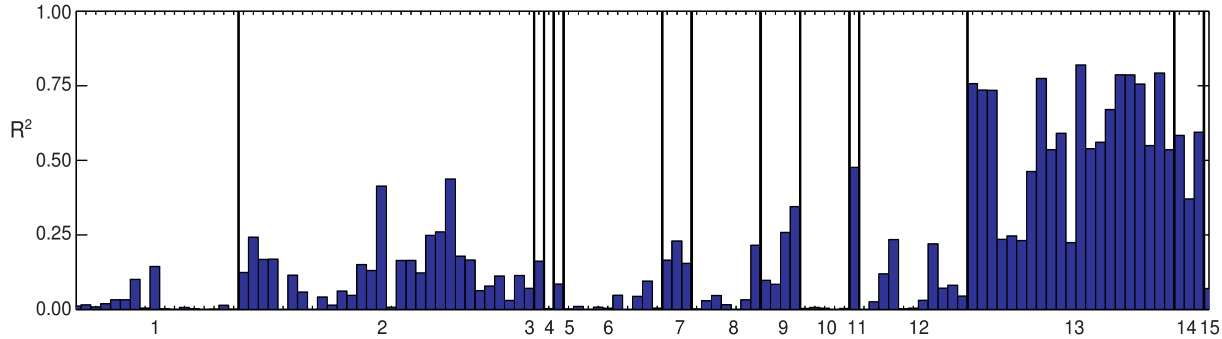
group	code	description	group	code	description
1	7	Personal Income (AR, Bil. Chain 2000 $) (TCB)	6	4	Houses Authorized By Build. Permits:South(Thou.U.)S.A.
1	7	Personal Income Less Transfer Payments (AR, Bil. Chain 2000 $) (TCB)	6	4	Houses Authorized By Build. Permits:West(Thou.U.)S.A.
1	7	Industrial Production Index - Total Index	7	1	Napm Inventories Index (Percent)
1	7	Industrial Production Index - Products, Total	7	7	Manufacturing And Trade Inventories (Bil. Chain 2000 $) (TCB)
1	7	Industrial Production Index - Final Products	7	8	Ratio, Mfg. And Trade Inventories To Sales (Based On Chain 2000 $) (TCB)
1	7	Industrial Production Index - Consumer Goods	8	1	Purchasing Managers' Index (Sa)
1	7	Industrial Production Index - Durable Consumer Goods	8	1	Napm New Orders Index (Percent)
1	7	Industrial Production Index - Nondurable Consumer Goods	8	1	Napm Vendor Deliveries Index (Percent)
1	7	Industrial Production Index - Business Equipment	8	7	Mfrs' New Orders, Consumer Goods And Materials (Bil. Chain 1982 $) (TCB)
1	7	Industrial Production Index - Materials	8	7	Mfrs' New Orders, Durable Goods Industries (Bil. Chain 2000 $) (TCB)
1	7	Industrial Production Index - Durable Goods Materials	8	7	Mfrs' New Orders, Nondefense Capital Goods (Mil. Chain 1982 $) (TCB)
1	7	Industrial Production Index - Nondurable Goods Materials	8	7	Mfrs' Unfilled Orders, Durable Goods Indus. (Bil. Chain 2000 $) (TCB)
1	7	Industrial Production Index - Manufacturing (Sic)	9	7	S&P's Common Stock Price Index: Composite (1941-43=10)
1	7	Industrial Production Index - Residential Utilities	9	7	S&P's Common Stock Price Index: Industrials (1941-43=10)
1	7	Industrial Production Index - Fuels	9	8	S&P's Composite Common Stock: Dividend Yield (% Per Annum)
1	1	Napm Production Index (Percent)	9	7	S&P's Composite Common Stock: Price-Earnings Ratio (%,Nsa)
1	8	Capacity Utilization (Mfg) (TCB)	10	7	United States;Effective Exchange Rate(Merm)(Index No.)
2	1	Index Of Help-Wanted Advertising In Newspapers (1967=100;Sa)	10	7	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)
2	1	Employment: Ratio; Help-Wanted Ads:No. Unemployed Clf	10	7	Foreign Exchange Rate: Japan (Yen Per U.S.$)
2	7	Civilian Labor Force: Employed, Total (Thous.,Sa)	10	7	Foreign Exchange Rate: United Kingdom (Cents Per Pound)
2	7	Civilian Labor Force: Employed, Nonagric.Industries (Thous.,Sa)	10	7	Foreign Exchange Rate: Canada (Canadian $ Per U.S.$)
2	1	Unemployment Rate: All Workers, 16 Years & Over (%,Sa)	11	1	Interest Rate: Federal Funds (Effective) (% Per Annum,Nsa)
2	8	Unemploy.By Duration: Average(Mean)Duration In Weeks (Sa)	12	7	Money Stock: M1(Curr,Trav.Cks,Dem Dep,Other Ck'able Dep)(Bil$,Sa)
2	7	Unemploy.By Duration: Persons Unempl.Less Than 5 Wks (Thous.,Sa)	12	7	Money Stock:M2(M1+O'nite Rps,Euro$,G/P&B/D Mmmfs&Sav&Sm Time Dep(Bil$,Sa)
2	7	Unemploy.By Duration: Persons Unempl.5 To 14 Wks (Thous.,Sa)	12	7	Money Stock: M3(M2+Lg Time Dep,Term Rp's&Inst Only Mmmfs)(Bil$,Sa)
2	7	Unemploy.By Duration: Persons Unempl.15 Wks + (Thous.,Sa)	12	7	Money Supply - M2 In 1996 Dollars (Bci)
2	7	Unemploy.By Duration: Persons Unempl.15 To 26 Wks (Thous.,Sa)	12	7	Monetary Base, Adj For Reserve Requirement Changes(Mil$,Sa)
2	7	Unemploy.By Duration: Persons Unempl.27 Wks + (Thous,Sa)	12	7	Depository Inst Reserves:Total, Adj For Reserve Req Chgs(Mil$,Sa)
2	7	Average Weekly Initial Claims, Unemploy. Insurance (Thous.) (TCB)	12	7	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs(Mil$,Sa)
2	7	Employees On Nonfarm Payrolls: Total Private	12	7	Commercial & Industrial Loans Oustanding In 1996 Dollars (Bci)
2	7	Employees On Nonfarm Payrolls - Goods-Producing	12	1	Wkly Rp Lg Com'l Banks:Net Change Com'l & Indus Loans(Bil$,Saar)
2	7	Employees On Nonfarm Payrolls - Mining	12	7	Consumer Credit Outstanding - Nonrevolving(G19)
2	7	Employees On Nonfarm Payrolls - Construction	12	8	Ratio, Consumer Installment Credit To Personal Income (Pct.) (TCB)
2	7	Employees On Nonfarm Payrolls - Manufacturing	13	7	Producer Price Index: Finished Goods (82=100,Sa)
2	7	Employees On Nonfarm Payrolls - Durable Goods	13	7	Producer Price Index: Finished Consumer Goods (82=100,Sa)
2	7	Employees On Nonfarm Payrolls - Nondurable Goods	13	7	Producer Price Index:I ntermed Mat.Supplies & Components(82=100,Sa)
2	7	Employees On Nonfarm Payrolls - Service-Providing	13	7	Producer Price Index: Crude Materials (82=100,Sa)
2	7	Employees On Nonfarm Payrolls - Trade, Transportation, And Utilities	13	7	Spot market price index: bls & crb: all commodities(1967=100)
2	7	Employees On Nonfarm Payrolls - Wholesale Trade	13	7	Index Of Sensitive Materials Prices (1990=100)(Bci-99a)
2	7	Employees On Nonfarm Payrolls - Retail Trade	13	1	Napm Commodity Prices Index (Percent)
2	7	Employees On Nonfarm Payrolls - Financial Activities	13	7	Cpi-U: All Items (82-84=100,Sa)
2	7	Employees On Nonfarm Payrolls - Government	13	7	Cpi-U: Apparel & Upkeep (82-84=100,Sa)
2	7	Employee Hours In Nonag. Establishments (AR, Bil. Hours) (TCB)	13	7	Cpi-U: Transportation (82-84=100,Sa)
2	1	Avg Weekly Hrs of Prod or Nonsup Workers On Priv. Nonf. Payrolls - Goods-Producing	13	7	Cpi-U: Medical Care (82-84=100,Sa)
2	8	Avg Weekly Hrs of Prod or Nonsup Workers On Priv. Nonf. Payrolls - Mfg Overtime Hours	13	7	Cpi-U: Commodities (82-84=100,Sa)
2	1	Average Weekly Hours, Mfg. (Hours) (TCB)	13	7	Cpi-U: Durables (82-84=100,Sa)
2	1	Napm Employment Index (Percent)	13	7	Cpi-U: Services (82-84=100,Sa)
3	7	Sales Of Retail Stores (Mil. Chain 2000 $) (TCB)	13	7	Cpi-U: All Items Less Food (82-84=100,Sa)
4	7	Manufacturing And Trade Sales (Mil. Chain 1996 $) (TCB)	13	7	Cpi-U: All Items Less Shelter (82-84=100,Sa)
5	7	Real Consumption (AC) a0m224/gmdc (a0m224 is from TCB)	13	7	Cpi-U: All Items Less Midical Care (82-84=100,Sa)
6	4	Housing Starts:Nonfarm(1947-58);Total Farm&Nonfarm(1959-)(Thous.,Saar)	13	7	Pce, Impl Pr Defl:Pce (1987=100)
6	4	Housing Starts:Northeast (Thous.U.)S.A.	13	7	Pce, Impl Pr Defl:Pce; Durables (1987=100)
6	4	Housing Starts:Midwest(Thous.U.)S.A.	13	7	Pce, Impl Pr Defl:Pce; Nondurables (1996=100)
6	4	Housing Starts:South (Thous.U.)S.A.	13	7	Pce, Impl Pr Defl:Pce; Services (1987=100)
6	4	Housing Starts:West (Thous.U.)S.A.	14	7	Avg Hourly Earnings of Prod or Nonsup Workers On Priv. Nonf. Payrolls - Goods-Producing
6	4	Housing Authorized: Total New Priv Housing Units (Thous.,Saar)	14	7	Avg Hourly Earnings of Prod or Nonsup Workers On Priv. Nonf. Payrolls - Construction
6	4	Houses Authorized By Build. Permits:Northeast(Thou.U.)S.A	14	7	Avg Hourly Earnings of Prod or Nonsup Workers On Priv. Nonf. Payrolls - Manufacturing
6	4	Houses Authorized By Build. Permits:Midwest(Thou.U.)S.A.	15	8	U. Of Mich. Index Of Consumer Expectations(Bcd-83)

Note: The table lists the individual macro series that we use to construct macro factors. The series are categorized in 15 groups: (1) real output and income, (2) employment and hours, (3) real retail, (4) manufacturing and trade sales, (5) consumption, (6) housing starts and sales, (7) real inventories, (8) orders, (9) stock prices, (10) exchange rates, (11) federal funds rate, (12) money and credit quantity aggregates, (13) prices indexes, (14) average hourly earnings and (15) miscellaneous. The transformations applied to original series are coded as: $1 \equiv$ no transformation (levels are used), $4 \equiv$ logarithm of the level, $7 \equiv$ annual first differences of the log levels and $8 \equiv$ annual first differences of the levels. The sample period is January 1970 - December 2003 (408 observations).

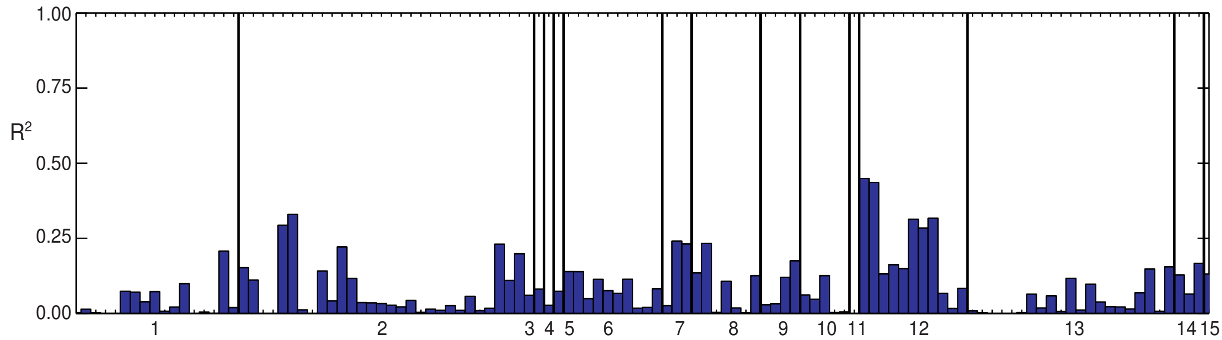
Figure 4.2: R^2 in regressions of individual macro series on PCA factors



(a) PCA factor #1



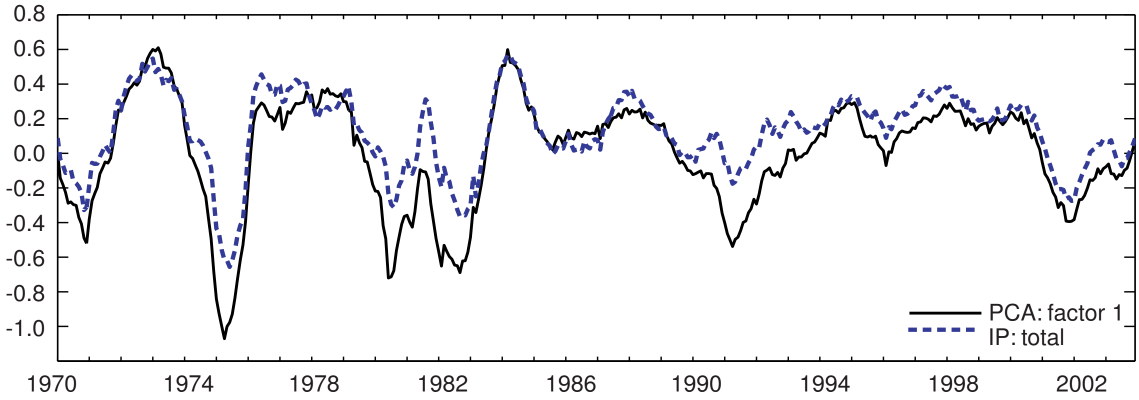
(b) PCA factor #2



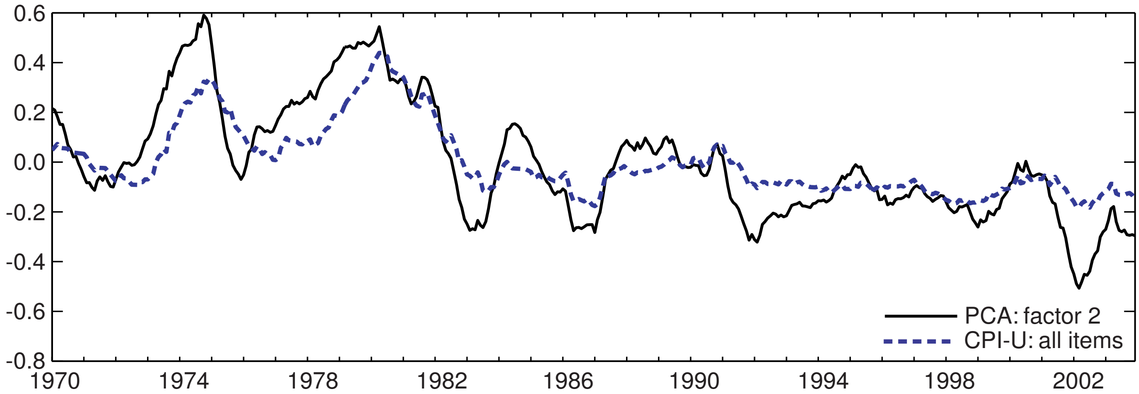
(c) PCA factor #3

Note: The figure shows R^2s when regressing the individual series in the macro panel on each of the first three macro factors. The macro dataset consists of 116 series (transformed to ensure stationarity) and the sample period is January 1970 - December 2003 (408 monthly observations). Panels (a), (b) and (c) show the results for the first, second and third macro factor respectively. In each panel the macro series are grouped according to the 15 categories as indicated on the horizontal axis. The group categories are (1) real output and income, (2) employment and hours, (3) real retail, (4) manufacturing and trade sales, (5) consumption, (6) housing starts and sales, (7) real inventories, (8) orders, (9) stock prices, (10) exchange rates, (11) federal funds rate, (12) money and credit quantity aggregates, (13) prices indexes, (14) average hourly earnings and (15) miscellaneous.

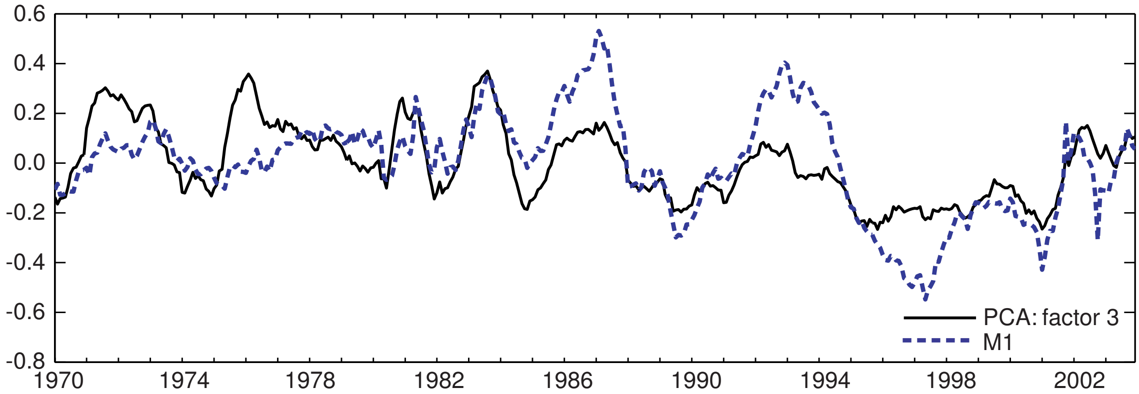
Figure 4.3: Macro factors compared to individual macro series



(a) PCA factor #1 - IP:total



(b) PCA factor #2 - CPI-U:total



(c) PCA factor #3 - M1

Note: The figure shows timeseries plots of the first three macro factors and the main individual macro series within the category to which the factor is most related. The first factor is plotted together with Industrial Production Index: Total Index (R^2 is 0.88), the second factor is plotted with the Consumer Price Index: All Items (R^2 is 0.77) and the third factor is plotted with Money Stock: M1 (R^2 is 0.44). The macro dataset consists of 116 (transformed to ensure stationarity) series and the sample period used is January 1970 - December 2003 (408 monthly observations). The group categories are (1) real output and income, (2) employment and hours, (3) real retail, (4) manufacturing and trade sales, (5) consumption, (6) housing starts and sales, (7) real inventories, (8) orders, (9) stock prices, (10) exchange rates, (11) federal funds rate, (12) money and credit quantity aggregates, (13) prices indexes, (14) average hourly earnings and (15) miscellaneous.

the macro dataset and their designated category.

We transform the monthly recorded macro series, whenever necessary, to ensure stationarity by using log levels, annual differences or annual log differences. Column 2 of Table 4.2 lists the applied transformation. We follow Ang and Piazzesi (2003), Mönch (2006b) and Diebold *et al.* (2006) in our use of annual growth rates. Monthly growth rates series are very noisy and are therefore expected to add little information when included in the various term structure models. Outliers in each individual series are replaced by the median value of the previous five observations, see Stock and Watson (2005) for details.

We need to be careful about the timing of the macro series relative to the interest rate series to prevent the use of information that has not been released yet at the time when a forecast is made. The interest rates we use are recorded at the end of the month. Although macro figures tend to be released at the beginning or in the middle of the month, they are usually released with a lag of one to sometimes several months. We accommodate for a potential look-ahead bias[6] by lagging all macro series by one month, except for S&P variables, exchange rates and the federal funds rate which are all monthly averages.

We extract a small number of common factors from our dataset, similar to Mönch (2006a) who, based on the work of Bernanke *et al.* (2005), builds a no-arbitrage Factor-Augmented VAR with four factors from a large panel of macroeconomic variables. To this end we apply static principal component analysis, see Stock and Watson (2002a,b), to the full panel of macro series which we standardize to have zero mean and unit variance. The use of common factors instead of individual macro series allows us to incorporate information beyond that contained in commonly used variables such as CPI, PPI, employment, output gap or capacity utilization, while at the same time ensuring that the number of model parameters remains manageable.

For the full sample period, the first common factor explains 35% of the variation in the macro panel. The second and third factors explain an additional 19% and 8%, whereas the first 10 factors together explain an impressive 85%. Figure 4.2 shows the R^2 when regressing each individual macro series on each of first three factors separately, which allows us to attach economic labels to these factors. The first factor closely resembles the series in the real output and employment categories (categories 1 and 2) and can therefore be labelled *business cycle* or *real activity* factor. The second factor loads mostly on inflation measures (category 13) which allows for the designation *inflation* factor. The third factor, although the correlations are much lower than for factors one and two, is mostly related to money stock and reserves (category 12) and could thus be labelled a

[6]Note that Ang and Piazzesi (2003) and Mönch (2006b) use *contemporaneous* macro information to construct their term structure forecasts. With contemporaneous information there is the risk that it may exaggerate the benefits from using macroeconomic series when forecasting yields.

monetary aggregates or *money stock* factor. Figure 4.3 corroborates these interpretations graphically through time-series plots of the three macro factors with Industrial Production (total), Consumer Price Index (all items) and Money Stock (M1) respectively.

We have chosen to include the first three factors as additional explanatory variables in the term structure model because, together, these factors explain over 60% of the variation in the macro panel[7]. Given that we want to construct interest rate forecasts we also need to forecast the macro factors. We explain in Section 3.1 in detail how we do so.

4.3 Models

We assess the individual and combined forecasting performance of a range of models that are commonly used in the literature and in practice. Since previous studies have shown that more parsimonious models often outperform sophisticated models we consider models with different levels of complexity. Our models range from unrestricted linear specifications for yield levels (AR and VAR models), models that impose a parametric structure on factor loadings (the Nelson-Siegel class of models) to models that impose cross-sectional restrictions to rule out arbitrage opportunities (affine models). In this section we present the different models. We defer to the appendix all specific details regarding the frequentist and Bayesian techniques to draw inference and to generate (multi-step ahead) forecasts.

4.3.1 Adding macro factors

The approach we use to incorporate the three macro factors is the following. Denote M_t as the (3×1) vector containing the time t values of the macro factors, which have been extracted from the full panel of macro series. We add the factors to each of the term structure models, contemporaneously[8] as well as lagged by one month to capture any delayed effects of macroeconomic news on the term structure. The exogenous explanatory macro information that we add to the models is denoted by X_t, and is thus given by $X_t = (M_t' \; M_{t-1}')'$.

Our approach implies that when we forecast yields, we also need to model and forecast the macro factors. We tackle this issue by following Ang and Piazzesi (2003) in only allowing for a *uni*directional link from macro variables to yields. Although this can be argued to be a restrictive assumption as it does not allow for a potentially rich *bi*directional

[7] We also examined using more factors but the forecasting results were very similar. With only one or two factors we obtained worse results.

[8] Contemporaneous in the sense of same-month values for stock prices, exchange rates and the federal funds rate but one-month lagged values for the remaining macro series, see Section 2.2 for further details.

feedback[9], it enables us to model the time-series behavior of the macro factors separately, which considerably facilitates estimation. In particular, information criteria suggest to model and forecast M_t using a VAR(3) model:

$$M_t = c + \Phi_1 M_{t-1} + \Phi_2 M_{t-2} + \Phi_3 M_{t-3} + H\xi_t, \qquad \varepsilon_t \sim \mathcal{N}(0, I) \tag{4.1}$$

where c is a (3×1) vector, Φ_i for $i = 1, ..., 3$ is a (3×3) matrix and H a (3×3) lower triangular Cholesky matrix. We estimate the macro VAR using both frequentist and Bayesian techniques as we also use both types of inference for the term structure models.

4.3.2 Models

Random walk

The first model that we consider is a random walk for each maturity τ_i, $i = 1, \ldots, N$,

$$y_t^{(\tau_i)} = y_{t-1}^{(\tau_i)} + \sigma^{(\tau_i)} \varepsilon_t^{(\tau_i)}, \qquad \varepsilon_t^{(\tau_i)} \sim \mathcal{N}(0, 1) \tag{4.2}$$

In this model any h-step ahead forecast $\hat{y}_{T+h}^{(\tau_i)}$ is equal to the most recent observed value $y_T^{(\tau_i)}$. It is natural to qualify this no-change model as the benchmark against which to judge the predictive power of other models. Duffee (2002), Ang and Piazzesi (2003), Mönch (2006b) and Diebold and Li (2006) all show, using different models and different forecast periods, that beating the random walk is quite an arduous task. The reported first order autocorrelation coefficients in Table 4.1 indeed confirm that yields are potentially non-stationary as these are all very close to unity. We denote the Random Walk by **RW**.

AR model

Although unreported results indicate that the null of a unit root for yield levels cannot be rejected statistically, the assumption of a random walk is difficult to interpret from an economic point of view. The random walk assumption implies that interest rates can roam around freely and do not revert back to a long-term mean, something which contradicts the Federal Reserve's monetary policy targets. The second model that we therefore consider is a first-order univariate autoregressive model which allows for mean-reversion

$$y_t^{(\tau_i)} = c^{(\tau_i)} + \phi^{(\tau_i)} y_{t-1}^{(\tau_i)} + \psi^{(\tau_i)\prime} X_t + \sigma^{(\tau_i)} \varepsilon_t^{(\tau_i)}, \qquad \varepsilon_t^{(\tau_i)} \sim \mathcal{N}(0, 1) \tag{4.3}$$

[9] In a forecasting exercise using German zero-coupon yields, Hordahl *et al.* (2006) show that term-structure information helps little in forecasting macro-economic variables (more specifically (i) inflation and (ii) the output gap) which is a justification for forecasting macro variables outside the term structure models. The authors note, however, that this might be due to the fact that their proposed macroeconomic model has an imperfect ability to describe the joint dynamics of German macroeconomic variables. Diebold *et al.* (2006) and Ang *et al.* (2006a) allow for bi-directional effects between macro and latent yield factors but both studies find that the causality from macro variables to yields is much higher than vice versa.

where $c^{(\tau_i)}$, $\phi^{(\tau_i)}$ and $\sigma^{(\tau_i)}$ are scalar parameters and $\psi^{(\tau_i)}$ is a (6×1) vector containing the coefficients on the macro factors. We construct forecasts both with and without macro factors by setting $\psi^{(\tau_i)} = 0$. We denote the yield-only model by **AR** and the model with macro factors by **AR-X**. For this and all other models we construct *iterated* forecasts[10].

VAR model

Vector autoregressive (VAR) models create the possibility to use the history of other maturities on top of any maturity's own history as additional information. We use the following first-order VAR specification[11],

$$Y_t = c + \Phi Y_{t-1} + \Psi X_t + H\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, I) \tag{4.4}$$

where Y_t contains the yields for all 13 maturities; $Y_t = [y_t^{(1m)}, ..., y_t^{(10y)}]'$, c is a (13×1) vector, Φ a (13×13) matrix, Ψ a (13×6) matrix and H is the lower triangular Cholesky decomposition of the (unrestricted) residual variance matrix $S = HH'$ containing $\frac{1}{2}N(N+1) = 91$ free parameters. As noted in the introduction, our approach is similar in spirit to the VAR models used in Evans and Marshall (1998, 2001) and Ang and Piazzesi (2003) in the sense that we impose exogeneity of macroeconomic variables with respect to yields.

A well-known drawback of using an unrestricted VAR model for yields is that forecasts can only be constructed for those maturities used in the estimation of the model. As we want to construct forecasts for 13 maturities, this results in a considerable number of parameters that need to be estimated. As an attempt to mitigate estimation error, and subsequently, to reduce the forecast error variance, we summarize the information contained in the explanatory vector Y_{t-1} by replacing it with a small number of common factors that drive yield curve dynamics. Similar to Litterman and Scheinkman (1991) and many other studies, we find that the first 3 principal components explain almost all the variation in yields (over 99%). We replace Y_{t-1} in (4.4) accordingly with the (13×3) factor matrix F_{t-1}[12]:

$$Y_t = c + \Phi F_{t-1} + \Psi X_t + H\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, I) \tag{4.5}$$

[10] Another approach is to construct *direct* forecasts by regressing $y_t^{(\tau_i)}$ directly on its h-month lagged value $y_{t-h}^{(\tau_i)}$ as in Diebold and Li (2006). For the state-space form of the Nelson-Siegel model and the affine model, such an approach is, however, infeasible. Therefore, and for matters of consistency, we choose to construct iterated forecasts for all the models. Whether iterated forecasts are more accurate than direct forecasts is a matter of ongoing debate, see the discussion in e.g. Marcellino *et al.* (2006).

[11] For both the AR and VAR models we examined the benefits of including more lags by analyzing AR(p) and VAR(p) models with $p = 2, \ldots, 12$. We found that using multiple lags resulted in nearly identical forecasts compared to the AR(1) and VAR(1) models and these results are therefore not reported nor were they included in the forecasting combination procedures in Sections 4 and 5.

[12] The time subscript '$t-1$' indicates that we extract the common factors using the history of yields up until $t-1$, thereby not using the vector of observations for time t.

where Φ is now a (13×3) matrix. The VAR model without and with macroeconomic variables is denoted by **VAR** and **VAR-X** respectively.

Nelson-Siegel model

Diebold and Li (2006) show that using the in essence static Nelson and Siegel (1987) model as a dynamic factor model generates highly accurate interest rate forecasts. The Nelson-Siegel model differs from the unrestricted VAR model in (4.5) by imposing a parametric structure on the factor loadings. The factor loadings Φ are specified as exponential functions of maturity and a single parameter λ. Following Diebold *et al.* (2006), the state-space representation of the three-factor model, with a first-order autoregressive representation for the dynamics of the state vector, is given by

$$\begin{aligned} y_t^{(\tau_i)} &= \beta_{1,t} + \beta_{2,t}\left[\frac{1-\exp(-\tau_i/\lambda)}{\tau_i/\lambda}\right] + \beta_{3,t}\left[\frac{1-\exp(-\tau_i/\lambda)}{\tau_i/\lambda} - \exp(-\tau_i/\lambda)\right] + \varepsilon_t^{(\tau_i)} && (4.6) \\ \beta_t &= a + \Gamma\beta_{t-1} + u_t && (4.7) \end{aligned}$$

The state vector $\beta_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$ contains the latent factors at time t which can be interpreted as level, slope and curvature factors (see Diebold and Li, 2006 for details). The parameter λ governs the exponential decay towards zero of the factor loadings on $\beta_{2,t}$ and $\beta_{3,t}$, a is a (3×1) vector of parameters and Γ a (3×3) matrix of parameters. We assume that the measurement equation and state equation errors in (4.6) and (4.7) are normally distributed and mutually uncorrelated,

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{18\times 1} \\ 0_{3\times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) \qquad (4.8)$$

where H is a diagonal (18×18) matrix and Q a full (3×3) matrix. We follow Diebold and Li (2006) by adding five maturities (τ = 9, 15, 18, 21 and 30 months) to the short end of the yield curve to estimate the Nelson-Siegel model in (4.6)-(4.8). We use two different estimation procedures: a *two-step* approach and a *one-step* approach. With the frequentist approach we apply both the two-step and one-step estimation procedure whereas with Bayesian analysis we consider only the one-step procedure.

The *two-step* approach is discussed in Diebold and Li (2006) and involves fixing λ and estimating the factors β_t in a first step using the cross-section of yields for each month t. Given the estimated time-series for the factors from the first step, the second step consists of modelling the factors in (4.7) by fitting either separate AR(1) models, thereby assuming that both Γ and Q are diagonal, or a single VAR(1) model. We denote these approaches by **NS2-AR** and **NS2-VAR** respectively.

The *one-step* approach follows from Diebold *et al.* (2006) and involves jointly estimating (4.6)-(4.8) as a state space model using the Kalman filter. In this approach we assume that Γ and Q are both full matrices and that λ is now estimated alongside the other parameters. We denote the one-step model by **NS1**.

Diebold *et al.* (2006) show that the Nelson-Siegel can be extended to incorporate macroeconomic variables by adding these as observable factors to the state vector and writing the model in companion form:

$$y_t^{(\tau_i)} = \beta_{1,t} + \beta_{2,t}\left[\frac{1-\exp(-\tau_i/\lambda)}{\tau_i/\lambda}\right] + \beta_{3,t}\left[\frac{1-\exp(-\tau_i/\lambda)}{\tau_i/\lambda} - \exp(-\tau_i/\lambda)\right] + \varepsilon_t^{(\tau_i)} \tag{4.9}$$

$$f_t = a + \Gamma f_{t-1} + \eta_t \tag{4.10}$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{18\times1} \\ 0_{12\times1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) \tag{4.11}$$

The state vector now also contains observable factors, $f_t = (\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, M_t, M_{t-1}, M_{t-2})$. The dimensions of a, Γ and Q are increased appropriately and η_t is given by $\eta_t = (u_t', \xi_t', 0, ..., 0)'$. The companion form enables us to incorporate the VAR(3) specification for the macro factors. We impose structure on Γ and Q to accommodate for the effects of macro factors while maintaining the unidirectional causality from macro factors to yields[13]. In particular, the lower left (9×3) block of Γ consists of zeros whereas Q is block diagonal with a non-zero (3×3) block Q_1 for the yield factors and a non-zero (3×3) block Q_2 for the macro factors. All other blocks on the diagonal contain only zeros. The Nelson-Siegel model with macro factors can again be estimated using either a two-step approach with AR or VAR dynamics for the yield factors, denoted by **NS2-AR-X** and **NS2-VAR-X**, or using the one-step approach, denoted by **NS1-X**.

Affine model

Models that impose no-arbitrage restrictions have been examined for their forecast accuracy in for example Duffee (2002), Ang and Piazzesi (2003) and Mönch (2006a). The attractive property of the class of no-arbitrage models is that sound theoretical cross-sectional restrictions are imposed on factor loadings to rule out arbitrage opportunities. In this study we consider a Gaussian-type discrete time affine no-arbitrage model using the set-up from Ang and Piazzesi (2003).

In particular, we assume that the vector of K underlying latent factors, or state variables, Z_t, which are assumed to drive movements in the yield curve, follow a Gaussian VAR(1) process

$$Z_t = \mu + \Psi Z_{t-1} + u_t \tag{4.12}$$

[13]Note that the macro factors are prevented from entering the measurement equations directly by only allowing the factor loadings of β_t to be non-zero in (4.9). Diebold *et al.* (2006) impose this restriction to maintain the assumption that three factors are sufficient for describing the dynamics of interest rates. Relaxing this restriction would result in a substantial number of additional parameters.

where $u_t \sim \mathcal{N}(0, \Sigma\Sigma')$ with Σ a lower triangular Choleski matrix, μ a $(K \times 1)$ vector and Ψ a $(K \times K)$ matrix. The short interest rate is assumed to be an affine function of the factors

$$r_t = \delta_0 + \delta_1' Z_t \tag{4.13}$$

where δ_0 is a scalar and δ_1 a $(K \times 1)$ vector. Furthermore, we adopt a standard form for the pricing kernel, which is assumed to price all assets in the economy,

$$m_{t+1} = \exp\big(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' u_{t+1}\big)$$

We specify market prices of risk to be time-varying and affine in the state variables

$$\lambda_t = \lambda_0 + \lambda_1 Z_t \tag{4.14}$$

with λ_0 a $(K \times 1)$ vector and λ_1 a $(K \times K)$ matrix[14]. Under the assumption that bond prices are an exponentially-affine function of the state variables,

$$P_t^{(\tau)} = \exp[A^{(\tau)} + B^{(\tau)'} Z_t] \tag{4.15}$$

we can recursively estimate the price of a $\tau-$period bond using

$$P_t^{(\tau)} = \mathbb{E}_t[m_{t+1} P_{t+1}^{(\tau-1)}] \tag{4.16}$$

where the expectation is taken under the risk-neutral measure. Ang and Piazzesi (2003) show that this results in the following recursive formulas for the bond pricing coefficients $A^{(\tau)}$ and $B^{(\tau)}$:

$$\begin{aligned} A^{(\tau+1)} &= A^{(\tau)} + B^{(\tau)'}[\mu - \Sigma\lambda_0] + \frac{1}{2}B^{(\tau)'}\Sigma\Sigma' B^{(\tau)} - \delta_0 && (4.17) \\ B^{(\tau+1)'} &= B^{(\tau)'}[\Psi - \Sigma\lambda_1] - \delta_1' && (4.18) \end{aligned}$$

when starting from $A^{(0)} = 0$ and $B^{(0)} = 0$. If bond *prices* are exponentially affine in the state variables then *yields* are affine in the state variables since $P_t^{(\tau)} = \exp[-y_t^{(\tau)}\tau]$. Consequently, it follows that $y_t^{(\tau)} = a^{(\tau)} + b^{(\tau)'} Z_t$ with $a^{(\tau)} = -A^{(\tau)}/\tau$ and $b^{(\tau)} = -B^{(\tau)}/\tau$. To estimate the model we deviate from the popular Chen and Scott (1993) approach and assume that every yield is contaminated with measurement error.

To summarize, we specify the following affine model

$$\begin{aligned} y_t^{(\tau_i)} &= a^{(\tau_i)} + b^{(\tau_i)} Z_t + \varepsilon_t^{(\tau_i)} && (4.19) \\ Z_t &= \mu + \Psi Z_{t-1} + u_t && (4.20) \end{aligned}$$

[14]Risk premia are constant over time if λ_1 equals zero. With λ_0 also equal to zero, risk premia are non-existent altogether.

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{13\times1} \\ 0_{3\times1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) \tag{4.21}$$

where $Q = \Sigma\Sigma'$ and $a^{(\tau_i)}$ and $b^{(\tau_i)}$ are recursive functions of the parameters that govern the dynamics of the state variables and of the risk premia parameters. We denote this model by **ATSM**.

We extend the model to include observable macroeconomic factors in a similar way as for the Nelson-Siegel model

$$y_t^{(\tau_i)} = a^{(\tau_i)} + b^{(\tau_i)} f_t + \varepsilon_t^{(\tau_i)} \tag{4.22}$$

$$f_t = \mu + \Psi f_{t-1} + \eta_t \tag{4.23}$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{13\times1} \\ 0_{12\times1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) \tag{4.24}$$

with $f_t = (Z_t, M_t, M_{t-1}, M_{t-2})$. The dimensions of $a^{(\tau_i)}$, $b^{(\tau_i)}$, μ, Ψ and Q are again increased as appropriate and the state equation (4.23) is written in companion form. As in the Nelson-Siegel model, Q is block diagonal with only two non-zero blocks, Q_1 and Q_2. We denote the affine model with macroeconomic factors by **ATSM-X**.

Adding macroeconomic variables to affine models can cause estimation problems as it further increases the number of parameters in these already highly parameterized models[15]. To speed up and to facilitate the estimation procedure, we therefore use the two-step approach of Ang *et al.* (2006b) by making the latent yield factors observable. Contrary to Ang *et al.* (2006b) who directly use the observed short rate and the term-spread as measures of the level and slope of the yield curve, we use principal component analysis to extract the first three common factors from the full set of yields and use these as our observable state variables.

4.4 Forecasting

4.4.1 Forecast procedure

We divide our dataset into an initial estimation sample which covers the period 1970:1 - 1988:12 (228 observations) and a forecasting sample which is comprised of the remaining period 1989:1 - 2003:12 (180 observations). The forecasting period is further divided in three 60-month subperiods; 1989:1 - 1993:12, 1994:1 - 1998:12 and 1999:1 - 2003:12. The

[15] Contrary to the reduced form affine model of Ang and Piazzesi (2003), Hordahl *et al.* (2006) use a structural affine model with macroeconomic variables in which the number of parameters can be kept down. They show that their model leads to better longer horizon forecasts compared to the Ang-Piazzesi model, which indicates that instead of only imposing no-arbitrage restrictions, which is the case in affine models, imposing also *structural* equations seems to mitigate overparameterization.

initial subperiod is primarily used as a training sample to start up the forecast combinations which we discuss in Section 5. Consequently, we report forecast results for the sample 1994:1 - 2003:12 (120 observations) and the last two subsamples (60 observations each). The vertical lines in Figure 4.1 serve to identify the subperiods.

We recursively estimate all models using an expanding window of all data from 1970:1 onwards. We construct point forecasts for four different horizons: $h = 1, 3, 6$ and 12 months ahead. As mentioned in the previous section, for horizons beyond $h = 1$ month we compute iterated forecasts when using frequentist techniques whereas for Bayesian inference we compute the mean of each model's h-month ahead predictive density.

4.4.2 Forecast evaluation

To evaluate the out-of-sample forecasts we compute a number of different popular error metrics per maturity and forecast horizon. We focus in particular on the Root Mean Squared Prediction Error (RMSPE)[16]. Similar to Hordahl *et al.* (2006) we also summarize the forecasting performance of each model over *all* maturities by computing the Trace Root Mean Squared Prediction Error (TRMSPE), see Christoffersen and Diebold (1998) for details.

To test the statistical accuracy of (combined) forecasts of all models relative to our random walk benchmark model, we apply, like Hordahl *et al.* (2006) and Mönch (2006b), the White (2000) "reality check" test with the stationary bootstrap approach of Politis and Romano (1994). We carry out the test using 1000 block-bootstraps of the forecast error series with an average block-length of 12 months.

4.4.3 Forecasting results: individual models

Tables 4.3-4.6 report out-of-sample results for the period 1994:1-2003:12 for the four selected forecast horizons. Panels A and B of each table contain results for the models with and without macro factors. The results with the frequentist approach are shown in the left hand side panels whereas those with Bayesian inference are given in the right hand side panels. Subsample results are reported in Tables 4.7-4.10 for the period 1994:1-1999:12 and Tables 4.11-4.14 for the period 1999:1-2003:12.

[16]Other forecast performance statistics such as the Mean Prediction Error (MPE), Mean Absolute Prediction Error (MAPE) and the R^2 when regressing observed h-month ahead yields on the corresponding forecasts are not reported but are available upon request. It would be interesting to evaluate the different forecasting models from a truly economic point of view by gauging the performance of bond portfolios but such an analysis is beyond the scope of this Chapter and is therefore left for further research. Results that can give an indication of the likely economic profitability of interest forecasts are available upon request. In particular, we have analyzed the Hit Rate which we compute as the percentage of correctly predicted signs of changes in interest rates with values about 50% indicating sign predictability.

The first row in each table shows the values of the different forecast evaluation metrics for the random walk (reported in basis point errors) whereas all other rows show values relative to the random walk. Relative values for any forecast that are below one are highlighted in bold to indicate that these forecasts are on average more accurate than those of the random walk. Stars indicate statistically significant outperformance according to White's reality check test.

Full sample results

Sample 1994:1 - 2003:12

The results for the 1-month horizon are not very encouraging. For nearly all maturities the random walk shows better statistics than any of the models based on yields only, even when parameter uncertainty is incorporated. The results are in line, however, with other studies showing that it is very difficult to outperform the RW for short horizon forecasts. Especially for short horizons the near unit root behavior of yields seems to dominate and model-based yield forecasts add little.

Incorporating macroeconomic information as an additional source of information improves forecasts for the AR and VAR models. The (T)RMSPE statistics are now very close and often marginally better than those of the RW. The largest improvements are shown for the shortest maturities, in particular the 3-month maturity where the relative RMPSE is now 0.95. Detailed inspection of the forecasts reveals that macroeconomic information helps especially to reduce the forecast bias. However, the improvements do not appear substantial enough for the AR-X model to produce significantly better forecasts, as judged by the White reality check test. The evidence for more complex model specifications is mixed but, in general, adding macroeconomic information worsens accuracy. For example, for the 6-month maturity the relative RMSPE increases from 1.10 to 1.71 for the Nelson-Siegel model when including macro factors.

The results for the 3-month forecast horizon are very similar to those for the 1-month horizon, although the RMSPE is now higher in absolute terms. The latter is expected since the yield curve is subject to more new information when the forecast horizon lengthens. It still proves very difficult for any of the models to provide forecasts that are more accurate than the random walk. The AR-X model is again the only model that shows promising results, which can again be attributed to the macro factors, as it gives a lower TRMSPE statistic than that of the random walk. The improvement is, however, not statistically significant. What is striking though is that whereas with the frequentist approach without macro factors the RMSPE goes up for $h = 3$ compared to $h = 1$, with the Bayesian approach the RMSPE actually goes down for some models, in particular the

Table 4.3: [T]RMSPE 1994:1 - 2003:12, 1-month forecast horizon

	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	101.59	30.12	21.18	21.82	25.71	29.12	30.48	29.3	27.95	101.59	30.12	21.18	21.82	25.71	29.12	30.48	29.3	27.95
Panel A: Individual models without macro factors																		
AR	1.02	1.04	1.07	1.06	1.05	1.03	1.01	1.01	1.01	1.02	1.03	1.05	1.05	1.04	1.02	1.01	1.01	1.01
VAR	1.06	**0.83**	1.03	1.23	1.14	1.13	1.04	1.05	1.11	1.04	**0.87**	1.10	1.30	1.15	1.11	1.01	1.02	1.03
NS2-AR	1.10	**0.94**	1.13	1.27	1.24	1.19	1.11	1.06	1.07	-	-	-	-	-	-	-	-	-
NS2-VAR	1.04	**0.94**	**0.96***	1.10	1.10	1.11	1.06	1.03	1.06	-	-	-	-	-	-	-	-	-
NS1	1.06	1.16	1.09	1.08	1.05	1.10	1.07	1.04	1.06	1.08	1.08	1.05	1.10	1.07	1.15	1.10	1.04	1.08
ATSM	1.07	**0.84**	**0.93**	1.15	1.23	1.18	1.04	1.08	1.07	1.08	**0.84**	**0.93**	1.15	1.23	1.19	1.05	1.09	1.07
Panel B: Individual models with macro factors																		
AR-X	**0.99**	**0.98**	**0.95**	**0.96**	**0.98**	**0.98**	**0.99**	1.00	**0.99**	**0.99**	**0.97**	**0.95**	**0.96**	**0.98**	**0.98**	**0.99**	1.00	**0.99**
VAR-X	1.02	**0.83**	**0.99**	1.03	1.01	1.12	1.02	1.02	1.03	1.02	**0.83**	1.00	1.04	1.01	1.12	1.02	1.03	1.04
NS2-AR-X	1.09	**0.90**	1.22	1.31	1.28	1.17	1.05	1.06	1.06	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.05	**0.83**	1.05	1.17	1.20	1.13	1.03	1.05	1.05	-	-	-	-	-	-	-	-	-
NS1-X	1.05	**0.98**	1.01	1.04	1.08	1.10	1.04	1.05	1.06	1.28	1.11	1.66	1.71	1.58	1.31	1.11	1.22	1.21
ATSM-X	1.13	**0.85**	1.13	1.18	1.29	1.42	1.04	**0.99**	1.06	1.12	**0.86**	1.15	1.16	1.26	1.39	1.03	1.01	1.11
Panel C: Forecast combinations without macro factors																		
FC-EW	1.02	**0.91**	**0.98**	1.08	1.08	1.09	1.02	1.02	1.03	1.01	**0.92**	**0.99**	1.07	1.06	1.07	1.01	1.02	1.02
FC-MSPE-exp	1.02	**0.89**	**0.98**	1.07	1.06	1.08	1.02	1.02	1.03	1.01	**0.91**	**0.99**	1.06	1.05	1.06	1.00	1.02	1.02
FC-MSPE-60	1.02	**0.90**	**0.98**	1.07	1.07	1.08	1.02	1.02	1.03	1.01	**0.91**	**0.99**	1.06	1.06	1.06	1.01	1.02	1.02
FC-MSPE-24	1.02	**0.89**	**0.98**	1.06	1.06	1.08	1.02	1.02	1.03	1.01	**0.91**	**0.98**	1.05	1.05	1.06	1.01	1.02	1.02
FC-MSPE-12	1.01	**0.89**	**0.97**	1.05	1.06	1.07	1.02	1.02	1.03	1.01	**0.91**	**0.98**	1.04	1.04	1.06	1.01	1.02	1.02
BMA	-	-	-	-	-	-	-	-	-	1.02	**0.95**	1.01	1.06	1.04	1.06	1.01	1.02	1.02
Panel D: Forecast combinations with macro factors																		
FC-EW-X	1.01	**0.85**	**0.97**	1.03	1.06	1.09	1.00	1.01	1.01	1.00	**0.87**	**0.98**	1.00	1.02	1.08	**0.99**	**0.99**	**0.99**
FC-MSPE-X-exp	1.00	**0.85**	**0.96**	1.01	1.04	1.07	1.00	1.01	1.01	**0.99**	**0.87**	**0.95**	**0.97**	1.00	1.05	**0.99**	1.00	**0.99**
FC-MSPE-X-60	1.00	**0.86**	**0.96**	1.01	1.04	1.07	1.00	1.01	1.01	**0.99**	**0.87**	**0.95**	**0.97**	1.00	1.05	**0.99**	1.00	**0.99**
FC-MSPE-X-24	1.00	**0.85**	**0.94**	1.01	1.04	1.07	1.00	1.01	1.01	**0.99**	**0.87**	**0.94**	**0.96**	**0.99**	1.04	**0.99**	1.00	**0.99**
FC-MSPE-X-12	1.00	**0.84**	**0.93**	1.01	1.04	1.07	1.00	1.01	1.01	**0.99**	**0.87**	**0.93**	**0.95**	**0.98**	1.04	**0.99**	1.00	**0.99**
BMA-X	-	-	-	-	-	-	-	-	-	**0.99**	**0.87**	**0.95**	**0.98**	1.00	1.04	**0.99**	1.00	**0.99**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	1.00	**0.86**	**0.94**	1.02	1.05	1.08	1.01	1.01	1.02	**0.98**	**0.86**	**0.91**	**0.95**	**0.99**	1.05	**0.99**	1.00	1.00
FC-MSPE-ALL-exp	1.00	**0.85**	**0.94**	1.01	1.04	1.07	1.01	1.01	1.02	**0.99**	**0.86**	**0.93**	**0.97**	1.00	1.04	**0.99**	1.00	1.00
FC-MSPE-ALL-60	1.00	**0.86**	**0.94**	1.01	1.04	1.07	1.01	1.01	1.02	**0.99**	**0.86**	**0.93**	**0.97**	1.00	1.04	**0.99**	1.00	1.00
FC-MSPE-ALL-24	1.00	**0.85**	**0.93**	1.01	1.04	1.07	1.01	1.01	1.02	**0.99**	**0.86**	**0.93**	**0.97**	1.00	1.04	**0.99**	1.00	1.00
FC-MSPE-ALL-12	1.00	**0.85**	**0.92**	1.00	1.03	1.07	1.01	1.01	1.02	**0.98**	**0.85**	**0.91**	**0.94**	**0.98**	1.03	**0.99**	1.00	1.00
BMA-ALL	-	-	-	-	-	-	-	-	-	**0.98**	**0.87**	**0.91**	**0.95**	**0.98**	1.03	**0.99**	1.00	1.01

Note: The table reports the [Trace] Root Mean Squared Prediction Error ([T]RMPSE) for *individual* yield models, with and without macro factors, estimated using the frequentist and Bayesian estimation approach (Panels A and B). Panels C - E show results for different forecast *combination* methods for the frequentist and Bayesian estimated models. All results are for a 1-month horizon for the out-of-sample period 1994:1 - 2003:12. The following model abbreviations are used: RW stands for the Random Walk, (V)AR for the first-order (Vector) Autoregressive Model, NS2-(V)AR for two-step Nelson-Siegel model with a (V)AR specification for the factors, NS1 for the one-step Nelson-Siegel model, ATSM for the affine model, FC-EW and FC-MSPE for the forecast combination based on equal weights and MSPE-based weights respectively, BMA for the Bayesian Model Averaging forecast. The affix 'X' indicates that macro factors have been added as additional variables. For the forecast combinations '-X' indicates that only forecasts are combined from model with macro factors whereas '-ALL' indicates that forecasts from all the models are combined. For the FC-MSPE method '-exp', '-60', '-24' and '-12' means that the model weights are determined using an expanding, or a 60, 24 or 12-month moving window respectively. The first line in the table reports the value of [T]RMSPE (expressed in basis points) for the Random Walk model (RW) while all other lines reports statistics relative to those of the RW. Bold numbers indicate outperformance relative to the RW whereas *, ** and *** indicate significant outperformance at the 90%, 95% and 99% level respectively according to the White (2000) reality check using 1000 block-bootstraps and an average block-length of 12.

Table 4.4: [T]RMSPE 1994:1 - 2003:12, 3-month forecast horizon

	FREQUENTIST INFERENCE									**BAYESIAN INFERENCE**								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	195.81	53.61	48.24	50.71	55.36	59.86	57.25	53.47	49.72	195.81	53.61	48.24	50.71	55.36	59.86	57.25	53.47	49.72
Panel A: Individual models without macro factors																		
AR	1.05	1.11	1.10	1.09	1.08	1.04	1.02	1.03	1.03	1.04	1.08	1.08	1.07	1.05	1.03	1.03	1.03	1.01
VAR	1.10	**0.90**	1.08	1.21	1.20	1.16	1.09	1.08	1.13	1.10	1.04	1.22	1.31	1.23	1.14	1.04	1.03	1.03
NS2-AR	1.13	1.02	1.16	1.24	1.26	1.23	1.13	1.07	1.06	-	-	-	-	-	-	-	-	-
NS2-VAR	1.05	**0.94**	**0.99**	1.08	1.11	1.11	1.06	1.03	1.05	-	-	-	-	-	-	-	-	-
NS1	1.06	1.09	1.09	1.11	1.10	1.10	1.06	1.02	1.03	1.01	**0.98**	**0.99**	1.03	1.02	1.04	1.01	1.00	1.02
ATSM	1.06	**0.85**	**0.96**	1.11	1.18	1.14	1.02	1.07	1.06	1.06	**0.85**	**0.96**	1.11	1.19	1.15	1.03	1.08	1.06
Panel B: Individual models with macro factors																		
AR-X	**0.98**	**0.98**	**0.95**	**0.96**	**0.98**	**0.98**	**0.99**	**0.99**	**0.99**	**0.98**	**0.94**	**0.95**	**0.95**	**0.98**	**0.98**	**0.99**	1.00	1.00
VAR-X	**0.99**	**0.87**	**0.98**	1.00	1.00	1.03	**0.99**	**0.99**	1.00	1.02	**0.84**	**0.96**	1.03	1.07	1.11	1.03	1.01	1.01
NS2-AR-X	1.13	1.03	1.24	1.27	1.28	1.20	1.08	1.07	1.04	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.07	**0.85**	1.04	1.13	1.19	1.16	1.05	1.05	1.03	-	-	-	-	-	-	-	-	-
NS1-X	1.03	**0.84**	**0.96**	1.04	1.10	1.10	1.04	1.03	1.03	1.02	**0.84**	1.01	1.06	1.09	1.04	**0.99**	1.04	1.05
ATSM-X	1.04	**0.80**	**0.94**	1.04	1.14	1.20	1.03	1.00	1.01	1.04	**0.80**	**0.95**	1.02	1.14	1.19	1.03	1.00	1.02
Panel C: Forecast combinations without macro factors																		
FC-EW	1.04	**0.94**	1.02	1.09	1.10	1.09	1.03	1.02	1.03	1.02	**0.96**	1.02	1.07	1.07	1.05	1.01	1.01	1.01
FC-MSPE-exp	1.04	**0.94**	1.02	1.09	1.10	1.09	1.04	1.03	1.04	1.02	**0.96**	1.03	1.07	1.07	1.06	1.01	1.02	1.01
FC-MSPE-60	1.04	**0.94**	1.03	1.09	1.11	1.09	1.04	1.03	1.04	1.02	**0.97**	1.03	1.07	1.07	1.06	1.01	1.02	1.01
FC-MSPE-24	1.03	**0.93**	1.01	1.09	1.10	1.08	1.03	1.03	1.03	1.02	**0.96**	1.01	1.06	1.07	1.05	1.01	1.01	1.01
FC-MSPE-12	1.03	**0.93**	1.00	1.07	1.09	1.07	1.03	1.02	1.03	1.02	**0.95**	1.00	1.05	1.06	1.05	1.01	1.01	1.01
BMA	-	-	-	-	-	-	-	-	-	1.01	**0.98**	1.01	1.03	1.03	1.03	1.00	1.00	1.01
Panel D: Forecast combinations with macro factors																		
FC-EW-X	1.00	**0.87**	**0.96**	1.01	1.05	1.06	1.00	1.00	**0.99**	**0.97**	**0.85**	**0.91**	**0.96**	1.00	1.02	**0.98**	**0.98**	**0.98**
FC-MSPE-X-exp	1.00	**0.87**	**0.95**	1.01	1.04	1.05	1.01	1.00	1.00	**0.98**	**0.85**	**0.92**	**0.97**	1.00	1.02	**0.99**	**0.99**	**0.98**
FC-MSPE-X-60	1.00	**0.87**	**0.96**	1.01	1.04	1.05	1.01	1.00	1.00	**0.98**	**0.85**	**0.93**	**0.97**	1.00	1.02	**0.99**	**0.99**	**0.98**
FC-MSPE-X-24	1.00	**0.86**	**0.94**	1.00	1.03	1.05	1.00	1.00	1.00	**0.97**	**0.84**	**0.92**	**0.96**	**0.99**	1.02	**0.98**	**0.98**	**0.98**
FC-MSPE-X-12	1.00	**0.85**	**0.92**	1.00	1.04	1.06	1.00	1.00	1.00	**0.97**	**0.84**	**0.90**	**0.95**	**0.99**	1.02	**0.98**	**0.99**	**0.98**
BMA-X	-	-	-	-	-	-	-	-	-	**0.97**	**0.87**	**0.92**	**0.96**	**0.99**	1.01	**0.98**	**0.98**	**0.98**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	1.00	**0.86**	**0.93**	1.01	1.05	1.06	1.01	1.00	1.00	**0.98**	**0.86**	**0.92**	**0.97**	1.00	1.02	**0.99**	**0.99**	**0.98**
FC-MSPE-ALL-exp	1.00	**0.86**	**0.94**	1.01	1.05	1.06	1.01	1.01	1.01	**0.98**	**0.85**	**0.93**	**0.98**	1.01	1.02	**0.99**	**0.99**	**0.99**
FC-MSPE-ALL-60	1.01	**0.86**	**0.95**	1.02	1.06	1.06	1.02	1.01	1.01	**0.98**	**0.86**	**0.93**	**0.98**	1.01	1.03	**0.99**	**0.99**	**0.99**
FC-MSPE-ALL-24	1.00	**0.86**	**0.95**	1.02	1.06	1.06	1.01	1.01	1.01	**0.98**	**0.86**	**0.93**	**0.98**	1.01	1.02	**0.99**	**0.99**	**0.99**
FC-MSPE-ALL-12	1.00	**0.85**	**0.93**	1.01	1.05	1.06	1.01	1.00	1.00	**0.98**	**0.85**	**0.91**	**0.96**	1.00	1.02	**0.98**	**0.99**	**0.99**
BMA-ALL	-	-	-	-	-	-	-	-	-	**0.97**	**0.88**	**0.92**	**0.96**	**0.99**	1.01	**0.98**	**0.99**	**0.99**

Note: The table reports forecast results for a 3-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 4.3 for further details.

Table 4.5: [T]RMSPE 1994:1 - 2003:12, 6-month forecast horizon

Models	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	300.94	83.60	82.31	85.20	89.24	92.74	86.36	79.23	72.50	300.94	83.60	82.31	85.20	89.24	92.74	86.36	79.23	72.50
Panel A: Individual models without macro factors																		
AR	1.07	1.15	1.12	1.10	1.10	1.06	1.03	1.04	1.04	1.06	1.11	1.10	1.09	1.08	1.05	1.06	1.05	1.08
VAR	1.20	1.11	1.22	1.31	1.31	1.24	1.14	1.15	1.21	1.24	1.34	1.43	1.47	1.38	1.23	1.10	1.13	1.22
NS2-AR	1.12	1.05	1.12	1.18	1.22	1.20	1.11	1.06	1.06	-	-	-	-	-	-	-	-	-
NS2-VAR	1.05	1.02	1.03	1.09	1.11	1.10	1.04	1.02	1.06	-	-	-	-	-	-	-	-	-
NS1	1.06	1.16	1.12	1.13	1.11	1.08	1.02	1.00	1.03	1.00	**0.99**	**0.99**	1.02	1.02	1.01	**0.99**	**0.99**	1.02
ATSM	1.06	**0.95**	1.02	1.12	1.17	1.12	1.01	1.07	1.07	1.07	**0.96**	1.02	1.13	1.18	1.13	1.02	1.07	1.08
Panel B: Individual models with macro factors																		
AR-X	1.00	**0.97**	**0.96**	**0.97**	**0.99**	1.00	1.01	1.00	1.01	**0.98**	**0.93**	**0.96**	**0.96**	**0.99**	**0.99**	1.00	1.00	**0.99**
VAR-X	**0.98**	**0.93**	**0.98**	1.00	1.01	1.00	**0.97**	**0.97**	**0.99**	1.03	**0.91**	**0.99**	1.06	1.11	1.11	1.01	**0.99**	1.00
NS2-AR-X	1.13	1.10	1.22	1.24	1.26	1.18	1.06	1.05	1.04	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.07	**0.94**	1.07	1.14	1.19	1.15	1.04	1.03	1.03	-	-	-	-	-	-	-	-	-
NS1-X	1.02	**0.87**	**0.96**	1.03	1.09	1.08	1.01	1.00	1.01	**0.97**	**0.87**	**0.93**	**0.97**	1.00	**0.98**	**0.96**	**0.99**	1.01
ATSM-X	1.02	**0.84**	**0.95**	1.04	1.11	1.12	**0.99**	**0.98**	1.01	1.02	**0.85**	**0.96**	1.04	1.11	1.12	**0.99**	**0.98**	1.01
Panel C: Forecast combinations without macro factors																		
FC-EW	1.05	1.02	1.05	1.10	1.11	1.08	1.02	1.02	1.04	1.04	1.04	1.07	1.10	1.09	1.05	1.01	1.02	1.05
FC-MSPE-exp	1.05	1.03	1.07	1.10	1.11	1.08	1.02	1.01	1.03	1.03	1.04	1.07	1.09	1.07	1.04	**0.99**	1.00	1.03
FC-MSPE-60	1.05	1.03	1.07	1.11	1.12	1.09	1.02	1.02	1.03	1.03	1.04	1.06	1.08	1.07	1.04	**0.99**	1.00	1.02
FC-MSPE-24	1.05	1.02	1.06	1.11	1.12	1.09	1.03	1.03	1.04	1.04	1.01	1.04	1.08	1.09	1.06	1.01	1.02	1.05
FC-MSPE-12	1.05	1.02	1.06	1.11	1.13	1.09	1.03	1.03	1.05	1.03	1.02	1.06	1.10	1.10	1.06	1.01	1.03	1.06
BMA	-	-	-	-	-	-	-	-	-	1.01	1.00	1.02	1.03	1.03	1.02	**0.99**	1.00	1.00
Panel D: Forecast combinations with macro factors																		
FC-EW-X	**0.99**	**0.90**	**0.96**	1.01	1.04	1.04	**0.99**	**0.98**	**0.99**	**0.97**	**0.88**	**0.93**	**0.97**	1.00	1.00	**0.97**	**0.96**	**0.97**
FC-MSPE-X-exp	**0.97**	**0.90**	**0.96**	**0.99**	1.01	1.01	**0.96**	**0.95**	**0.95**	**0.95**	**0.89**	**0.94**	**0.96**	**0.98**	**0.98**	**0.94**	**0.94**	**0.93**
FC-MSPE-X-60	**0.97**	**0.91**	**0.97**	1.00	1.02	1.01	**0.96**	**0.96**	**0.96**	**0.95**	**0.90**	**0.95**	**0.97**	**0.99**	**0.98**	**0.94**	**0.94**	**0.93**
FC-MSPE-X-24	1.00	**0.90**	**0.96**	1.00	1.03	1.04	1.00	**0.99**	1.00	**0.97**	**0.89**	**0.95**	**0.98**	1.00	1.00	**0.97**	**0.97**	**0.98**
FC-MSPE-X-12	1.00	**0.89**	**0.94**	1.00	1.04	1.05	1.00	1.00	1.00	**0.97**	**0.88**	**0.94**	**0.97**	1.01	1.01	**0.97**	**0.97**	**0.98**
BMA-X	-	-	-	-	-	-	-	-	-	**0.97**	**0.91**	**0.94**	**0.97**	**0.99**	**0.99**	**0.97**	**0.97**	**0.97**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	**0.99**	**0.90**	**0.95**	1.00	1.04	1.03	**0.98**	**0.98**	**0.99**	**0.98**	**0.92**	**0.95**	**0.99**	1.02	1.01	**0.97**	**0.98**	1.00
FC-MSPE-ALL-exp	**0.98**	**0.91**	**0.96**	1.01	1.04	1.02	**0.97**	**0.96**	**0.97**	**0.97**	**0.91**	**0.96**	**0.99**	1.00	**0.99**	**0.96**	**0.96**	**0.97**
FC-MSPE-ALL-60	**0.99**	**0.92**	**0.98**	1.02	1.05	1.04	**0.97**	**0.97**	**0.98**	**0.97**	**0.92**	**0.97**	1.00	1.01	1.00	**0.96**	**0.96**	**0.97**
FC-MSPE-ALL-24	1.02	**0.93**	**0.99**	1.04	1.07	1.06	1.01	1.01	1.01	1.00	**0.93**	**0.97**	1.02	1.04	1.03	**0.99**	**0.99**	1.00
FC-MSPE-ALL-12	1.03	**0.93**	**0.99**	1.05	1.08	1.07	1.02	1.02	1.03	1.01	**0.93**	**0.99**	1.04	1.05	1.03	**0.99**	1.00	1.02
BMA-ALL	-	-	-	-	-	-	-	-	-	**0.97**	**0.93**	**0.95**	**0.98**	1.00	**0.99**	**0.97**	**0.98**	**0.99**

Note: The table reports forecast results for a 6-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 4.3 for further details.

Table 4.6: [T]RMSPE 1994:1 - 2003:12, 12-month forecast horizon

	FREQUENTIST INFERENCE									**BAYESIAN INFERENCE**								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	452.51	136.94	140.61	145.03	146.89	141.77	121.21	108.58	98.96	452.51	136.94	140.61	145.03	146.89	141.77	121.21	108.58	98.96
Panel A: Individual models without macro factors																		
AR	1.10	1.15	1.11	1.09	1.10	1.09	1.07	1.09	1.10	1.07	1.07	1.10	1.08	1.08	1.06	1.06	1.06	1.10
VAR	1.43	1.36	1.41	1.44	1.42	1.40	1.41	1.46	1.55	1.64	1.74	1.77	1.76	1.68	1.58	1.52	1.57	1.65
NS2-AR	1.10	1.02	1.04	1.06	1.10	1.14	1.13	1.12	1.13	-	-	-	-	-	-	-	-	-
NS2-VAR	1.08	1.09	1.07	1.08	1.08	1.09	1.07	1.07	1.12	-	-	-	-	-	-	-	-	-
NS1	1.09	1.21	1.15	1.13	1.10	1.09	1.05	1.04	1.08	**0.99**	1.00	**0.99**	1.00	1.00	1.00	**0.97**	**0.98**	1.01
ATSM	1.10	1.06	1.07	1.11	1.14	1.12	1.04	1.12	1.13	1.10	1.06	1.08	1.12	1.14	1.12	1.05	1.13	1.13
Panel B: Individual models with macro factors																		
AR-X	1.02	**0.95**	**0.95**	**0.98**	1.00	1.03	1.06	1.05	1.06	1.01	**0.94**	**0.96**	**0.98**	1.01	1.03	1.04	1.03	1.02
VAR-X	**0.98**	**0.97**	**0.97**	**0.98**	**0.99**	**0.99**	**0.96**	**0.97**	**0.99**	1.14	1.06	1.09	1.10	1.13	1.14	1.13	1.15	1.20
NS2-AR-X	1.14	1.15	1.19	1.18	1.19	1.16	1.09	1.09	1.08	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.11	1.07	1.13	1.14	1.17	1.15	1.07	1.06	1.05	-	-	-	-	-	-	-	-	-
NS1-X	1.01	**0.91**	**0.96**	1.00	1.05	1.06	1.01	1.00	1.01	**0.96**	**0.93**	**0.94**	**0.96**	**0.97**	**0.97**	**0.95**	**0.96**	**0.98**
ATSM-X	1.02	**0.93**	**0.99**	1.04	1.07	1.08	**0.99**	1.00	1.02	1.02	**0.93**	**0.99**	1.03	1.07	1.07	**0.99**	1.00	1.04
Panel C: Forecast combinations without macro factors																		
FC-EW	1.08	1.08	1.08	1.09	1.10	1.09	1.07	1.08	1.11	1.08	1.09	1.10	1.11	1.10	1.08	1.05	1.07	1.10
FC-MSPE-exp	1.09	1.11	1.10	1.11	1.11	1.10	1.07	1.08	1.10	1.05	1.09	1.09	1.08	1.07	1.05	1.02	1.04	1.06
FC-MSPE-60	1.09	1.10	1.10	1.11	1.11	1.11	1.07	1.08	1.10	1.04	1.07	1.07	1.06	1.06	1.04	1.01	1.03	1.04
FC-MSPE-24	1.09	1.08	1.08	1.10	1.11	1.11	1.07	1.08	1.10	1.06	1.05	1.06	1.08	1.09	1.08	1.04	1.05	1.06
FC-MSPE-12	1.11	1.11	1.11	1.13	1.15	1.13	1.09	1.10	1.11	1.04	1.10	1.12	1.14	1.14	1.11	1.07	1.08	1.09
BMA	-	-	-	-	-	-	-	-	-	1.02	1.01	1.03	1.03	1.03	1.02	1.01	1.01	1.03
Panel D: Forecast combinations with macro factors																		
FC-EW-X	1.00	**0.94**	**0.97**	**0.99**	1.02	1.03	1.00	1.00	1.00	**0.96**	**0.91**	**0.93**	**0.96**	**0.98**	**0.99**	**0.96**	**0.97**	**0.98**
FC-MSPE-X-exp	**0.95**	**0.95**	**0.97**	**0.97**	**0.98**	**0.97**	**0.94**	**0.94**	**0.92**	**0.94**	**0.94**	**0.95**	**0.95**	**0.96**	**0.95**	**0.93**	**0.93**	**0.92**
FC-MSPE-X-60	**0.96**	**0.97**	1.00	**0.99**	**0.99**	**0.98**	**0.94**	**0.94**	**0.92**	**0.94**	**0.95**	**0.96**	**0.96**	**0.96**	**0.95**	**0.92**	**0.92**	**0.91**
FC-MSPE-X-24	1.01	**0.95**	**0.98**	1.00	1.02	1.04	1.02	1.02	1.02	1.00	**0.94**	**0.97**	**0.99**	1.01	1.02	1.01	1.01	1.01
FC-MSPE-X-12	1.03	**0.95**	**0.98**	1.00	1.03	1.06	1.04	1.04	1.04	1.02	**0.96**	**0.99**	1.00	1.03	1.04	1.03	1.04	1.05
BMA-X	-	-	-	-	-	-	-	-	-	**0.98**	**0.95**	**0.96**	**0.97**	**0.99**	**0.99**	**0.98**	**0.98**	**0.98**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	**0.99**	**0.93**	**0.95**	**0.97**	1.00	1.01	**0.99**	1.00	1.01	1.00	**0.97**	**0.98**	1.00	1.01	1.01	**0.99**	1.01	1.03
FC-MSPE-ALL-exp	**0.98**	**0.96**	**0.98**	**0.99**	1.00	1.00	**0.97**	**0.97**	**0.97**	**0.98**	**0.97**	**0.99**	**0.99**	**0.99**	**0.98**	**0.96**	**0.97**	**0.98**
FC-MSPE-ALL-60	**0.99**	**0.99**	1.00	1.01	1.02	1.01	**0.97**	**0.97**	**0.97**	**0.97**	**0.98**	**0.99**	**0.99**	**0.99**	**0.98**	**0.95**	**0.96**	**0.96**
FC-MSPE-ALL-24	1.04	1.00	1.03	1.05	1.07	1.07	1.04	1.03	1.04	1.03	0.99	1.02	1.04	1.06	1.06	1.03	1.03	1.04
FC-MSPE-ALL-12	1.11	1.05	1.08	1.10	1.12	1.14	1.11	1.11	1.10	1.09	1.05	1.08	1.11	1.12	1.11	1.07	1.08	1.09
BMA-ALL	-	-	-	-	-	-	-	-	-	**0.99**	**0.97**	**0.98**	1.00	1.01	1.01	**0.99**	**0.99**	1.00

Note: The table reports forecast results for a 12-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 4.3 for further details.

Nelson-Siegel model.

For a 6-month horizon more models start to outperform the random walk for more maturities, as indicated by a larger number of relative RMSPEs below 1, although the results are still by no means impressive, and the best model only improves the random walk by a few percentage points. Taking into account macroeconomic information as well as parameter uncertainty results in reasonably accurate forecasts although there is still no significant outperformance. Incorporating parameter uncertainty is very beneficial for the Nelson-Siegel model. The Bayesian estimation of the state-space form of the model substantially reduces the relative RMSPE compared to the frequentist approach. Models that keep struggling are the VAR and affine models. In both cases this is most likely due to the large number of yields (compared to for example Duffee, 2002 and Ang and Piazzesi, 2003) that we use in estimation, resulting in a large number of parameters[17]. Note that the VAR model with Bayesian inference does worse than when estimated using maximum likelihood. This can be explained by realizing that Bayesian analysis requires drawing inference on the variance parameters of each of the 13 maturities in addition to doing so for all the other parameters. With maximum likelihood this is not necessary as we only generate point forecasts.

The longest horizon that we consider is $h = 12$. Two models produce forecasts that consistently outperform the random walk across all maturities: the frequentist VAR-X model and the Bayesian NS1-X model. For both models, the TRMSPEs are smaller compared to the random walk. RMSPEs are on average 5% lower, although for the NS1-X the differences are not significant. For all other models, the benefits of adding macro factors are evident with all relative MSPE going down considerably. Compared to the frequentist results, the Bayesian VAR model still struggles.

It is interesting to compare our results with those of Mönch (2006b) as he uses an almost identical forecasting sample (1994:1 - 2003:9) but a much shorter estimation period (1983:1 - 1993:12) for the VAR, NS2-AR and NS2-VAR model. Our results for the RW are identical, as they should be, which is a convenient check on our results. The RMSPEs we find for the VAR(1) on yields and a 1-month horizon are somewhat higher for maturities below five years whereas for longer maturities they are very similar. For a 12-month horizon the differences are larger as Mönch reports RMPSEs which are roughly 20% lower than ours. The differences will partly be due to using a slightly different set of maturities and our use of yield-factors when estimating the VAR instead of using lagged yields directly. The main reason for the different sets of results will, however, be due to

[17] An obvious solution to this problem would be to estimate the affine models using a smaller set of yields. The reason we do not follow this strategy here is because we want to use a similar number of yields as in Mönch (2006a).

our much longer estimation sample. It seems that including the 1970s and beginning of 1980s leads to poorer yield forecasts compared to those obtained when starting the sample after the Volcker period. For the NS2-AR and NS2-VAR the 1-month ahead results are again very similar. However, whereas Mönch finds that NS2-AR outperforms NS2-VAR for a 6- and 12-month horizon we find that NS2-VAR is usually more accurate. Our affine model without macro variables provides similar results as for the $A_0(3)$ model that Mönch considers for $h = 1$ but less accurate results for $h = 6$ and $h = 12$. However, we forecast the 1-month maturity much more accurately which is most likely due to the fact that we estimate the short rate parameters δ_0 and δ_1 using *only* data on the 1-month yield instead of estimating these simultaneously with the other model parameters. It is interesting to note that none of the models we consider here have an out-of-sample performance which is as good as that of the FAVAR model advocated by Mönch. It would therefore be worthwhile to add this model to the model consideration set but we leave this for further research.

As an overall summary for the 1994:1-2003:12 period we can remark that our results for the individual models are not very encouraging as interest rate predictability appears to be rather low. This may be attributed to a number of possible causes with one main reason being the out-of-sample period we select. Except for Mönch (2006b) who reports very promising out-of-sample results for his FAVAR model for nearly the same period, Duffee (2002), Ang and Piazzesi (2003), Diebold and Li (2006) and Hordahl *et al.* (2006) all use an out-of-sample period that ranges from roughly the mid 1990s till 2000. As we also include the period from 2000 onwards, a possible explanation for our poor forecasting results seems to be locked up in that period. Figure 4.1 surely indicates that the interest rate behavior during that period with its pronounced widening of spreads is rather different from the stable second half of the 1990s. The subsample results reported in Mönch (2006b) for the period 2000:1-2003:9 indicate that the VAR, NS2-AR and NS2-VAR models perform poorly compared to the RW which is evidence that forecastability is indeed low during that period. Through analyzing the subsamples 1994:1-1998:12 and 1999:1-2003:12 we hope get more insight on this issue.

Subsample results

Sample 1994:1 - 1998:12

This five year subsample is the period that has been most heavily investigated in other forecasting studies, with positive results found for different models. For example, Duffee (2002) reports forecast results for affine models that hold up favorably against the random walk for the period 1995:1-1998:12. Similarly, Ang and Piazzesi (2003) show that a no-

Table 4.7: [T]RMSPE 1994:1 - 1998:12, 1-month forecast horizon

	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	95.45	27.42	17.79	19.05	25.15	28.18	28.70	27.53	26.60	95.45	27.42	17.79	19.05	25.15	28.18	28.70	27.53	26.60
Panel A: Individual models without macro factors																		
AR	1.00	**0.98**	**0.97**	**0.98**	**0.99**	1.00	1.01	1.01	1.02	1.00	**0.98**	**0.97**	**0.98**	**0.99**	1.00	1.00	1.01	1.01
VAR	1.02	**0.82**	1.02	1.31	1.05	1.03	1.02	1.05	1.15	1.02	**0.88**	1.19	1.45	1.14	1.07	**0.98**	**0.99**	1.03
NS2-AR	**0.96**	**0.86**	**0.90**	**0.98**	**0.92**	**0.97**	1.03	**0.98**	**0.98**	-	-	-	-	-	-	-	-	-
NS2-VAR	1.02	**0.99**	**0.97**	1.06	**0.98**	1.03	1.06	1.02	1.04	-	-	-	-	-	-	-	-	-
NS1	1.02	1.21	1.08	1.04	**0.95**	1.01	1.02	1.01	1.05	1.03	1.05	**0.99**	1.06	**0.97**	1.03	1.05	1.03	1.09
ATSM	1.05	**0.82**	**0.87**	1.04	1.00	1.08	1.05	1.12	1.06	1.05	**0.81**	**0.87**	1.04	1.01	1.08	1.06	1.12	1.07
Panel B: Individual models with macro factors																		
AR-X	**0.96**	**0.96**	**0.88**	**0.88**	**0.93**	**0.95***	**0.97**	**0.97**	**0.97**	**0.97**	**0.95**	**0.90**	**0.91**	**0.95**	**0.97**	**0.98**	**0.98**	**0.98**
VAR-X	**0.97**	**0.82**	**0.94**	**0.98**	**0.93**	1.05	**0.96**	**0.98**	1.03	**0.97**	**0.84**	**0.97**	1.01	**0.95**	1.05	**0.97**	**0.98**	1.03
NS2-AR-X	1.02	**0.84**	1.32	1.33	1.20	1.03	**0.94**	**0.99**	1.02	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.03	**0.81**	1.17	1.27	1.22	1.08	**0.97**	1.01	1.04	-	-	-	-	-	-	-	-	-
NS1-X	1.01	**0.88**	**0.87**	**0.99**	1.09	1.05	**0.97**	1.03	1.08	1.30	1.03	1.82	1.89	1.71	1.39	1.09	1.20	1.23
ATSM-X	1.16	**0.87**	1.38	1.37	1.31	1.52	**0.93**	**0.91**	1.05	1.16	**0.89**	1.41	1.40	1.31	1.52	**0.94**	**0.94**	1.03
Panel C: Forecast combinations without macro factors																		
FC-EW	**0.98**	**0.88**	**0.91**	1.01	**0.95**	1.01	1.01	1.02	1.02	**0.99**	**0.89**	**0.96**	1.06	1.00	1.02	1.00	1.02	1.02
FC-MSPE-exp	**0.98**	**0.87**	**0.92**	1.00	**0.96**	1.01	1.01	1.02	1.02	**0.99**	**0.88**	**0.96**	1.04	1.00	1.02	1.00	1.01	1.02
FC-MSPE-60	**0.98**	**0.87**	**0.92**	1.00	**0.96**	1.01	1.01	1.02	1.02	**0.99**	**0.88**	**0.96**	1.04	1.00	1.02	1.00	1.01	1.02
FC-MSPE-24	**0.99**	**0.87**	**0.91**	**0.99**	**0.96**	1.01	1.02	1.02	1.02	**0.99**	**0.89**	**0.94**	1.02	1.00	1.02	1.01	1.01	1.02
FC-MSPE-12	**0.98**	**0.87**	**0.91**	**0.98**	**0.95**	1.01	1.01	1.01	1.02	**0.99**	**0.89**	**0.93**	1.00	**0.99**	1.02	1.00	1.01	1.01
BMA	-	-	-	-	-	-	-	-	-	**0.99**	**0.91**	**0.97**	1.06	**0.98**	1.01	1.00	1.01	1.02
Panel D: Forecast combinations with macro factors																		
FC-EW-X	**0.98**	**0.82**	**0.98**	1.04	1.03	1.05	**0.95**	**0.96**	1.00	1.00	**0.87**	1.06	1.10	1.07	1.10	**0.96**	**0.97**	**0.99**
FC-MSPE-X-exp	**0.97**	**0.82**	**0.96**	1.01	1.00	1.03	**0.95**	**0.97**	1.00	**0.99**	**0.87**	**0.99**	1.02	1.00	1.06	**0.96**	**0.97**	**0.99**
FC-MSPE-X-60	**0.97**	**0.83**	**0.97**	1.00	**0.99**	1.03	**0.95**	**0.97**	1.00	**0.99**	**0.88**	**0.99**	1.02	1.00	1.06	**0.96**	**0.98**	**0.99**
FC-MSPE-X-24	**0.97**	**0.82**	**0.94**	1.00	**0.99**	1.02	**0.95**	**0.97**	1.00	**0.98**	**0.88**	**0.98**	1.01	1.00	1.05	**0.96**	**0.98**	**0.99**
FC-MSPE-X-12	**0.97**	**0.82**	**0.92**	**0.99**	1.00	1.02	**0.95**	**0.97**	1.00	**0.98**	**0.87**	**0.97**	1.00	**0.99**	1.05	**0.96**	**0.98**	**0.99**
BMA-X	-	-	-	-	-	-	-	-	-	**0.99**	**0.87**	1.01	1.05	1.02	1.05	**0.96**	**0.97**	**0.99**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	**0.96**	**0.82**	**0.87**	**0.94**	**0.95**	1.00	**0.97**	**0.98**	1.00	**0.97**	**0.84**	**0.91**	**0.96**	**0.97**	1.02	**0.97**	**0.98**	1.00
FC-MSPE-ALL-exp	**0.96**	**0.81**	**0.87**	**0.94**	**0.95**	1.00	**0.97**	**0.98**	1.00	**0.97**	**0.84**	**0.91**	**0.96**	**0.96**	1.02	**0.97**	**0.98**	1.00
FC-MSPE-ALL-60	**0.96**	**0.82**	**0.88**	**0.94**	**0.94**	1.00	**0.97**	**0.98**	1.00	**0.97**	**0.84**	**0.91**	**0.96**	**0.97**	1.02	**0.97**	**0.98**	1.00
FC-MSPE-ALL-24	**0.96**	**0.82**	**0.88**	**0.94**	**0.95**	1.00	**0.97**	**0.98**	1.00	**0.97**	**0.85**	**0.91**	**0.97**	**0.97**	1.02	**0.97**	**0.99**	**0.99**
FC-MSPE-ALL-12	**0.96**	**0.82**	**0.87**	**0.93**	**0.94**	**0.99**	**0.97**	**0.98**	1.00	**0.97**	**0.84**	**0.90**	**0.94**	**0.95**	1.01	**0.97**	**0.99**	**0.99**
BMA-ALL	-	-	-	-	-	-	-	-	-	**0.96**	**0.85**	**0.90**	**0.96**	**0.96**	1.00	**0.97**	**0.98**	1.01

Note: The table reports forecast results for a 1-month horizon for the out-of-sample period 1994:1 - 1998:12. See Table 4.3 for further details.

Table 4.8: [T]RMSPE 1994:1 - 1998:12, 3-month forecast horizon

	FREQUENTIST INFERENCE									**BAYESIAN INFERENCE**								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	194.55	41.39	37.86	44.46	54.74	61.30	58.86	56.01	51.56	194.55	41.39	37.86	44.46	54.74	61.30	58.86	56.01	51.56
Panel A: Individual models without macro factors																		
AR	1.00	**0.96**	**0.95**	**0.96**	**0.97**	**0.99**	1.01	1.02	1.03	1.00	**0.93**	**0.93**	**0.94**	**0.96**	**0.99**	1.01	1.04	1.02
VAR	1.02	**0.84**	1.02	1.14	1.00	**0.97**	1.02	1.05	1.17	1.05	1.11	1.33	1.37	1.16	1.05	1.00	**0.99**	1.04
NS2-AR	**0.88**	**0.74**	**0.74**	**0.82**	**0.81**	**0.86**	**0.93**	**0.91****	**0.95**	-	-	-	-	-	-	-	-	-
NS2-VAR	**0.99**	1.01	**0.97**	1.01	**0.95**	**0.98**	1.02	1.00	1.06	-	-	-	-	-	-	-	-	-
NS1	**0.98**	1.19	1.06	1.02	**0.93**	**0.95**	**0.97**	**0.97**	1.04	**0.98**	**0.87**	**0.92**	1.00	**0.97**	**0.98**	**0.99**	**0.99**	1.05
ATSM	1.01	**0.83**	**0.89**	1.01	**0.98**	1.01	1.02	1.07	1.08	1.01	**0.82**	**0.88**	1.00	**0.97**	1.00	1.02	1.07	1.08
Panel B: Individual models with macro factors																		
AR-X	**0.96**	**0.90**	**0.89**	**0.90****	**0.93*****	**0.96****	**0.98**	**0.97**	**0.98**	**0.98**	**0.89**	**0.96**	**0.95***	**0.97**	**0.98**	**0.99**	**0.99**	**0.99**
VAR-X	**0.95**	**0.82**	**0.90**	**0.96**	**0.95***	**0.97**	**0.95**	**0.95**	1.02	**0.94**	**0.82**	**0.90**	**0.97**	**0.96**	**0.99**	**0.93**	**0.93**	1.01
NS2-AR-X	1.01	1.10	1.39	1.28	1.16	1.01	**0.90**	**0.93**	**0.97**	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.04	0.92	1.24	1.24	1.19	1.08	**0.96**	**0.98**	1.02	-	-	-	-	-	-	-	-	-
NS1-X	**0.97**	**0.65**	**0.88**	1.00	1.04	1.01	**0.94**	**0.97**	1.04	1.05	**0.89**	1.21	1.23	1.18	1.07	**0.98**	1.01	1.06
ATSM-X	1.01	**0.85**	1.12	1.13	1.10	1.16	**0.92**	**0.90**	**0.98**	1.00	**0.85**	1.12	1.13	1.09	1.15	**0.91**	**0.89**	**0.93**
Panel C: Forecast combinations without macro factors																		
FC-EW	**0.96**	**0.87**	**0.90**	**0.96**	**0.93**	**0.95**	**0.98**	**0.99**	1.03	**0.99**	**0.91**	**0.97**	1.02	**0.99**	**0.99**	**0.99**	1.01	1.02
FC-MSPE-exp	**0.98**	**0.87**	**0.92**	**0.98**	**0.96**	**0.98**	1.00	1.01	1.05	1.00	**0.92**	**0.98**	1.03	1.01	1.01	1.01	1.02	1.04
FC-MSPE-60	**0.98**	**0.86**	**0.92**	**0.98**	**0.96**	**0.98**	1.00	1.01	1.05	1.01	**0.92**	**0.99**	1.04	1.01	1.01	1.01	1.02	1.04
FC-MSPE-24	**0.97**	**0.85**	**0.89**	**0.96**	**0.94**	**0.96**	**0.99**	1.00	1.03	**0.99**	**0.89**	**0.95**	1.01	**0.99**	1.00	**0.99**	1.01	1.02
FC-MSPE-12	**0.97**	**0.85**	**0.89**	**0.95**	**0.93**	**0.96**	**0.99**	**0.99**	1.03	**0.99**	**0.89**	**0.93**	**0.99**	**0.98**	**0.99**	**0.99**	1.01	1.02
BMA	-	-	-	-	-	-	-	-	-	**0.98**	**0.89**	**0.94**	1.00	**0.98**	**0.99**	**0.99**	1.00	1.03
Panel D: Forecast combinations with macro factors																		
FC-EW-X	**0.95**	**0.81**	**0.96**	1.00	**0.99**	**0.99**	**0.94**	**0.94**	**0.98**	**0.96**	**0.84**	**0.96**	1.00	1.00	1.00	**0.95**	**0.95**	**0.98**
FC-MSPE-X-exp	**0.96**	**0.82**	**0.96**	**0.99**	**0.98**	**0.99**	**0.95**	**0.95**	**0.99**	**0.97**	0.84	**0.97**	1.01	1.00	1.01	**0.96**	**0.96**	**0.99**
FC-MSPE-X-60	**0.96**	**0.83**	**0.96**	**0.99**	**0.98**	**0.99**	**0.95**	**0.95**	**0.99**	**0.97**	**0.85**	**0.97**	1.01	1.00	1.01	**0.96**	**0.96**	**0.99**
FC-MSPE-X-24	**0.95**	**0.82**	**0.94**	**0.98**	**0.98**	**0.99**	**0.94**	**0.94**	**0.98**	**0.96**	**0.84**	**0.96**	1.00	**0.99**	1.00	**0.95**	**0.95**	**0.98**
FC-MSPE-X-12	**0.95**	**0.80**	**0.92**	**0.99**	1.00	1.00	**0.94**	**0.94**	**0.98**	**0.96**	**0.84**	**0.95**	1.00	1.00	1.01	**0.94**	**0.95**	**0.98**
BMA-X	-	-	-	-	-	-	-	-	-	**0.96**	**0.86**	**0.96**	**0.99**	1.00	**0.99**	**0.95**	**0.96**	**0.98**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	**0.92**	**0.72**	**0.80**	**0.88**	**0.90**	**0.93**	**0.94**	**0.94**	**0.99**	**0.95**	**0.79**	**0.87**	**0.93**	**0.94**	**0.96**	**0.96**	**0.96**	**0.99**
FC-MSPE-ALL-exp	**0.94**	**0.74**	**0.83**	**0.91**	**0.93**	**0.95**	**0.96**	**0.96**	1.01	**0.96**	**0.79**	**0.89**	**0.96**	**0.97**	**0.98**	**0.97**	**0.98**	1.01
FC-MSPE-ALL-60	**0.94**	**0.75**	**0.84**	**0.92**	**0.93**	**0.96**	**0.96**	**0.96**	1.01	**0.97**	**0.80**	**0.91**	**0.97**	**0.97**	**0.98**	**0.97**	**0.98**	1.01
FC-MSPE-ALL-24	**0.93**	**0.74**	**0.84**	**0.92***	**0.92**	**0.95**	**0.94**	**0.95**	**0.99**	**0.95**	**0.79**	**0.89**	**0.95****	**0.96**	**0.97**	**0.95**	**0.97**	**0.99**
FC-MSPE-ALL-12	**0.93**	**0.75***	**0.84**	**0.92***	**0.94**	**0.96**	**0.94**	**0.95**	**0.98**	**0.95**	**0.79***	**0.87**	**0.94**	**0.96**	**0.98**	**0.95**	**0.96**	**0.99**
BMA-ALL	-	-	-	-	-	-	-	-	-	**0.95**	**0.81**	**0.88**	**0.95**	**0.96**	**0.97**	**0.96**	**0.97**	1.00

Note: The table reports forecast results for a 3-month horizon for the out-of-sample period 1994:1 - 1998:12. See Table 4.3 for further details.

Table 4.9: [T]RMSPE 1994:1 - 1998:12, 6-month forecast horizon

	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	293.03	62.10	62.98	70.70	82.43	91.58	88.63	83.33	78.96	293.03	62.10	62.98	70.70	82.43	91.58	88.63	83.33	78.96
Panel A: Individual models without macro factors																		
AR	**0.99**	**0.91**	**0.90**	**0.92**	**0.93**	**0.96**	1.00	1.02	1.04	1.00	**0.89**	**0.88**	**0.90**	**0.94**	**0.98**	1.02	1.04	1.07
VAR	1.03	**0.95**	1.03	1.08	**0.97**	**0.93**	1.03	1.09	1.19	1.09	1.45	1.49	1.39	1.10	**0.92**	**0.96**	1.05	1.18
NS2-AR	**0.82**	**0.65**	**0.64***	**0.70****	**0.72*****	**0.77*****	**0.87*****	**0.88*****	**0.93**	-	-	-	-	-	-	-	-	-
NS2-VAR	**0.98**	1.01	**0.95**	**0.96**	**0.92**	**0.93**	**0.99**	1.01	1.07	-	-	-	-	-	-	-	-	-
NS1	**0.95**	1.19	1.03	**0.97**	**0.89**	**0.88**	**0.93**	**0.95**	1.01	**0.97**	**0.83**	**0.93**	**0.99**	**0.99**	**0.97**	**0.97**	**0.99**	1.03
ATSM	1.00	**0.89**	**0.90**	**0.97**	**0.96**	**0.96**	1.00	1.07	1.08	1.01	**0.89**	**0.91**	**0.98**	**0.96**	**0.97**	1.00	1.07	1.08
Panel B: Individual models with macro factors																		
AR-X	**0.99**	**0.89**	**0.93**	**0.96**	**0.96**	**0.99**	1.01	1.00	1.01	1.01	**0.91**	1.02	1.01	1.02	1.01	1.02	1.00	1.00
VAR-X	**0.94**	**0.85**	**0.90****	**0.95****	**0.95****	**0.94****	**0.93***	**0.95****	1.00	**0.88**	**0.79**	**0.81**	**0.88**	**0.90**	**0.90**	**0.86**	**0.89**	**0.97**
NS2-AR-X	1.04	1.31	1.44	1.33	1.2	1.02	**0.90***	**0.92***	**0.95**	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.08	1.17	1.35	1.32	1.26	1.12	**0.97**	**0.98**	1.01	-	-	-	-	-	-	-	-	-
NS1-X	**0.94**	**0.73**	**0.92**	1.00	1.04	**0.99**	**0.91**	**0.93**	**0.98**	1.01	**0.97**	1.11	1.14	1.12	1.02	**0.96**	**0.98**	1.01
ATSM-X	1.00	**0.92**	1.11	1.14	1.09	1.09	**0.90**	**0.90**	**0.98**	**0.99**	**0.92**	1.11	1.13	1.09	1.08	**0.90**	**0.90**	**0.96**
Panel C: Forecast combinations without macro factors																		
FC-EW	**0.94**	**0.87**	**0.87**	**0.90**	**0.88**	**0.90**	**0.96**	**0.99**	1.03	**0.98**	**0.94**	**0.96**	**0.98**	**0.94**	**0.94**	**0.98**	1.01	1.05
FC-MSPE-exp	**0.95**	**0.90**	**0.90**	**0.93**	**0.90**	**0.92**	**0.96**	**0.98**	1.02	**0.96**	**0.98**	**0.98**	**0.99**	**0.94**	**0.93**	**0.95**	**0.99**	1.03
FC-MSPE-60	**0.95**	**0.90**	**0.90**	**0.93**	**0.90**	**0.92**	**0.96**	**0.98**	1.02	**0.97**	**0.99**	1.01	1.00	**0.95**	**0.94**	**0.95**	**0.99**	1.03
FC-MSPE-24	**0.96**	**0.86**	**0.88**	**0.93**	**0.93**	**0.94**	**0.98**	1.00	1.03	**0.98**	**0.90**	**0.91**	**0.97**	**0.97**	**0.97**	**0.99**	1.02	1.05
FC-MSPE-12	**0.97**	**0.89***	**0.90****	**0.96**	**0.95**	**0.95**	**0.98**	1.00	1.03	**0.99**	**0.92**	**0.94**	**0.99**	1.00	**0.98**	**0.99**	1.02	1.04
BMA	-	-	-	-	-	-	-	-	-	**0.97**	**0.90**	**0.93**	**0.96**	**0.96**	**0.96**	**0.98**	1.00	1.03
Panel D: Forecast combinations with macro factors																		
FC-EW-X	**0.96**	**0.88**	1.00	1.02	1.01	**0.98**	**0.93***	**0.93**	**0.97**	**0.95**	**0.87**	**0.96**	1.00	**0.99**	**0.98**	**0.93**	**0.94**	**0.97**
FC-MSPE-X-exp	**0.90**	**0.88**	**0.96**	**0.96**	**0.94**	**0.92**	**0.87**	**0.87***	**0.90**	**0.91**	**0.88**	**0.97**	**0.97**	**0.95**	**0.93**	**0.88**	**0.89**	**0.91**
FC-MSPE-X-60	**0.90**	**0.88**	**0.96**	**0.96**	**0.94**	**0.92**	**0.87***	**0.87***	**0.90**	**0.91**	**0.88**	**0.97**	**0.97**	**0.95**	**0.93**	**0.89**	**0.89**	**0.91**
FC-MSPE-X-24	**0.96**	**0.90**	**0.98**	1.00	**0.99**	**0.98**	**0.93***	**0.94***	**0.97**	**0.97**	**0.90**	1.00	1.02	1.01	**0.99**	**0.95**	**0.95**	**0.98**
FC-MSPE-X-12	**0.96**	**0.88**	**0.97**	1.00	1.01	**0.99**	**0.94***	**0.94***	**0.97**	**0.97**	**0.90**	**0.99**	1.02	1.02	1.00	**0.95**	**0.96**	**0.98**
BMA-X	-	-	-	-	-	-	-	-	-	**0.96**	**0.90**	**0.97**	1.00	1.00	**0.98**	**0.95**	**0.96**	**0.98**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	**0.88**	**0.69****	**0.76****	**0.83*****	**0.85*****	**0.87*****	**0.90***	**0.92**	**0.97**	**0.93**	**0.80****	**0.85****	**0.90*****	**0.91*****	**0.92*****	**0.94***	**0.96**	1.00
FC-MSPE-ALL-exp	**0.87**	**0.74****	**0.81****	**0.85*****	**0.86****	**0.87****	**0.88**	**0.90**	**0.94**	**0.91**	**0.82***	**0.89***	**0.91***	**0.91**	**0.90**	**0.90**	**0.92**	**0.96**
FC-MSPE-ALL-60	**0.88**	**0.76***	**0.83***	**0.87****	**0.87***	**0.88***	**0.88**	**0.90**	**0.94**	**0.91**	**0.83***	**0.90**	**0.92**	**0.91**	**0.91**	**0.90**	**0.93**	**0.96**
FC-MSPE-ALL-24	**0.93**	**0.79****	**0.87****	**0.92****	**0.92****	**0.93**	**0.93**	**0.94**	**0.98**	**0.96**	**0.82***	**0.88***	**0.95**	**0.97**	**0.96**	**0.96**	**0.97**	1.00
FC-MSPE-ALL-12	**0.95**	**0.82****	**0.88*****	**0.95****	**0.96**	**0.96**	**0.95***	**0.96***	**0.99**	**0.97**	**0.85**	**0.91**	**0.97**	**0.99**	**0.97**	**0.96**	**0.98**	1.00
BMA-ALL	-	-	-	-	-	-	-	-	-	**0.95**	**0.84****	**0.90****	**0.94*****	**0.95****	**0.95****	**0.95**	**0.97**	1.00

Note: The table reports forecast results for a 6-month horizon for the out-of-sample period 1994:1 - 1998:12. See Table 4.3 for further details.

Table 4.10: [T]RMSPE 1994:1 - 1998:12, 12-month forecast horizon

	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	423.6	99.7	104.32	113.44	126.05	133.67	124.01	114.05	109.81	423.6	99.7	104.32	113.44	126.05	133.67	124.01	114.05	109.81
Panel A: Individual models without macro factors																		
AR	**0.96**	**0.84**	**0.82**	**0.83**	**0.85**	**0.91**	1.00	1.04	1.07	**0.96**	**0.80**	**0.80**	**0.85**	**0.86**	**0.93**	**0.99**	1.02	1.10
VAR	1.14	1.14	1.15	1.13	1.00	**0.97**	1.13	1.24	1.31	1.32	1.72	1.69	1.58	1.30	1.13	1.19	1.29	1.36
NS2-AR	**0.82**	**0.69**	**0.66*****	**0.67*****	**0.66*****	**0.73*****	**0.89*****	**0.93****	**0.97**	-	-	-	-	-	-	-	-	-
NS2-VAR	**0.94**	**0.93**	**0.87**	**0.86**	**0.82**	**0.85**	**0.97**	1.01	1.08	-	-	-	-	-	-	-	-	-
NS1	**0.93**	1.13	**0.98**	**0.91**	**0.82**	**0.83***	**0.92**	**0.96**	1.02	**0.95**	**0.90**	**0.94****	**0.98**	**0.96*****	**0.94****	**0.95*****	**0.97*****	1.00
ATSM	**0.96**	**0.87**	**0.86**	**0.89**	**0.87**	**0.89**	**0.97**	1.07	1.08	**0.96**	**0.87**	**0.86**	**0.90**	**0.87**	**0.89**	**0.97**	1.07	1.08
Panel B: Individual models with macro factors																		
AR-X	1.01	**0.84**	**0.93*****	**0.97*****	**0.96*****	1.00	1.05	1.02	1.04	1.05	**0.98**	1.09	1.10	1.07	1.06	1.06*	1.03**	1.02*
VAR-X	**0.90**	**0.85**	**0.86*****	**0.89*****	**0.89*****	**0.89*****	**0.90*****	**0.93*****	**0.97**	**0.74**	**0.61**	**0.59*****	**0.58*****	**0.58*****	**0.62*****	**0.77*****	**0.86***	**0.95**
NS2-AR-X	1.09	1.49	1.50	1.38	1.23	1.06	**0.92****	**0.94****	**0.94****	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.15	1.44	1.49	1.42	1.31	1.16	1.00*	1.00**	1.00	-	-	-	-	-	-	-	-	-
NS1-X	**0.91**	**0.86**	**0.96**	**0.99**	**0.99**	**0.93****	**0.85*****	**0.87*****	**0.90****	**0.99**	**0.99**	1.07	1.08	1.06	1.00	**0.94*****	**0.96*****	**0.97***
ATSM-X	**0.98**	1.05	1.13	1.13	1.06	1.02	**0.87*****	**0.88*****	**0.94****	**0.98**	1.05	1.13	1.13	1.06	1.02	**0.87*****	**0.88*****	**0.96****
Panel C: Forecast combinations without macro factors																		
FC-EW	**0.91**	**0.83**	**0.80**	**0.81**	**0.79**	**0.83***	**0.94**	1.00	1.04	**0.93**	**0.85**	**0.85**	**0.86**	**0.83**	**0.86**	**0.95**	1.00	1.05
FC-MSPE-exp	**0.92**	**0.86**	**0.83**	**0.83**	**0.80**	**0.84**	**0.95**	1.02	1.06	**0.91**	**0.88**	**0.86**	**0.85**	**0.82**	**0.84**	**0.92**	**0.99**	1.04
FC-MSPE-60	**0.92**	**0.86**	**0.83**	**0.83**	**0.80**	**0.84**	**0.95**	1.02	1.06	**0.91**	**0.87**	**0.86**	**0.85**	**0.82**	**0.85**	**0.92**	1.00	1.04
FC-MSPE-24	**0.93**	**0.85**	**0.83**	**0.85**	**0.84**	**0.87**	**0.96**	1.00	1.04	**0.95**	**0.84**	**0.86**	**0.91**	**0.91**	**0.92**	**0.97**	1.00	1.04
FC-MSPE-12	**0.96**	**0.93**	**0.91**	**0.93**	**0.90**	**0.90**	**0.97**	1.01	1.05	1.01	**0.99**	1.02	1.05	1.01	**0.97**	**0.99**	1.03	1.07
BMA										**0.94**	**0.85**	**0.89**	**0.91**	**0.91***	**0.91****	**0.95**	**0.97**	1.01
Panel D: Forecast combinations with macro factors																		
FC-EW-X	**0.95**	**0.99**	1.04	1.04	1.00	**0.96****	**0.91*****	**0.92*****	**0.94****	**0.89**	**0.82*****	**0.87*****	**0.90*****	**0.89*****	**0.89*****	**0.88*****	**0.90*****	**0.93****
FC-MSPE-X-exp	**0.80**	**0.92**	**0.94**	**0.88**	**0.82**	**0.78****	**0.75*****	**0.77*****	**0.78****	**0.79**	**0.83*****	**0.85****	**0.81****	**0.77****	**0.75*****	**0.77*****	**0.80*****	**0.83***
FC-MSPE-X-60	**0.80**	**0.92**	**0.93**	**0.88**	**0.81**	**0.78****	**0.75*****	**0.77*****	**0.78****	**0.79**	**0.83*****	**0.85****	**0.81****	**0.77****	**0.75*****	**0.77*****	**0.80*****	**0.83***
FC-MSPE-X-24	**0.96**	**0.97**	1.02	1.02	**0.99**	**0.96**	**0.93*****	**0.93*****	**0.95*****	**0.96**	**0.91*****	**0.98*****	**0.99****	**0.97****	**0.96*****	**0.95*****	**0.94*****	**0.97****
FC-MSPE-X-12	**0.97**	**0.97**	1.00	1.01	**0.99**	**0.98**	**0.95****	**0.95*****	**0.97***	**0.99**	**0.96****	1.03	1.03	1.01	**0.99**	**0.98****	**0.97****	**0.99***
BMA-X										**0.94*****	**0.90*****	**0.94*****	**0.96*****	**0.96*****	**0.95*****	**0.93*****	**0.94*****	**0.96****
Panel E: Forecast combinations with all models																		
FC-EW-ALL	**0.83**	**0.68*****	**0.71*****	**0.75*****	**0.76*****	**0.79*****	**0.85*****	**0.89*****	**0.94**	**0.85**	**0.70*****	**0.72*****	**0.76*****	**0.77*****	**0.81*****	**0.88*****	**0.92*****	**0.97**
FC-MSPE-ALL-exp	**0.77**	**0.72*****	**0.73*****	**0.72*****	**0.69*****	**0.71*****	**0.78*****	**0.83*****	**0.87**	**0.81**	**0.75*****	**0.77*****	**0.76*****	**0.73*****	**0.75*****	**0.82****	**0.87***	**0.92**
FC-MSPE-ALL-60	**0.78**	**0.73*****	**0.74*****	**0.73*****	**0.70*****	**0.72*****	**0.79*****	**0.84*****	**0.88**	**0.81**	**0.76*****	**0.78*****	**0.77*****	**0.74*****	**0.76*****	**0.82****	**0.88***	**0.92**
FC-MSPE-ALL-24	**0.89**	**0.83*****	**0.87*****	**0.88*****	**0.86*****	**0.86*****	**0.89*****	**0.91*****	**0.95**	**0.94**	**0.83*****	**0.90*****	**0.94*****	**0.92*****	**0.93*****	**0.95****	**0.96****	1.00
FC-MSPE-ALL-12	**0.96**	**0.91**	**0.94**	**0.95**	**0.93**	**0.94**	**0.96**	**0.98**	1.01	1.01	**0.93**	1.01	1.04	1.01	**0.99**	1.00	1.01	1.05
BMA-ALL										**0.91**	**0.81*****	**0.87*****	**0.90*****	**0.90*****	**0.90*****	**0.92*****	**0.94*****	**0.97**

Note: The table reports forecast results for a 12-month horizon for the out-of-sample period 1994:1 - 1998:12. See Table 4.3 for further details.

arbitrage Gaussian VAR model predicts well 1-month ahead for the period 1996:1-2000:12 while Diebold and Li (2006) report outperforming forecasts for the Nelson-Siegel model for the period 1994:1-2000:12[18]. These studies suggest that there should be a high degree of predictability for this subperiod. Tables 4.7-4.10 confirm this claim. Even for a 1-month horizon it is already possible to outperform the random walk. The AR-X model in particular performs well across all maturities with results for the frequentist approach being slightly better than for the Bayesian approach. The latter is most likely due to the fact that the prior information based solely on the initial sample does not fit well with this period of smooth interest rates. The TRMSPEs are lower than for the random walk but the White test does not indicate significant improvements. The NS2-AR and VAR-X models also do well although the 2-year and 10-year maturities still seem difficult to forecast. The affine models render poor forecasts in this subsample, except for the 5- and 7-year maturities. This differs from Ang and Piazzesi (2003) who show that an affine model augmented with an inflation and a real activity factor forecasts better than the random walk for maturities up to and including five years. This difference in results could be due to the substantially larger number of yields that we use in estimation. Furthermore, Ang and Piazzesi (2003) do not forecast beyond a 1-month horizon.

For the 3-month horizon other models also start to predict well, but especially for 6- and 12-months ahead predictability is evident. The VAR-X model and the NS2-AR model in particular now produce forecasts that are significantly better than the no-change forecast with relative RMSPE being lower by sometimes as much as 30-40%. Adding macro factors seems to reduce forecast accuracy. Except for the VAR-X model, incorporating parameter uncertainty does not seem to help either. The performance of the affine models also improves. Interestingly, for shorter maturities simple affine models do better than their counterparts with macro information, but the evidence is just the opposite for longer forecast horizons. However, the affine models are never the best performing models for any maturity, which is a result also found by Diebold and Li (2006).

Comparing our results to those of Diebold and Li (2006) makes sense, since that study has the largest overlap in the set of models considered[19]. Results for $h = 1$ for the RW, AR, VAR and NS2-AR models are nearly identical in terms of RMSPE although we find slightly different MPEs (in our case the MPE is in general positive whereas Diebold and Li report mainly negative values). For $h = 6$ we find lower RMSPEs for the maturities below five years whereas for the AR and VAR models results are very similar, despite the different

[18]Hordahl *et al.* (2006) construct 1 through 12-months ahead forecasts for the period 1995:1-1998:12 but these authors apply their structural model to German zero-coupon data and their results might therefore not be directly comparable to the results for U.S. data.

[19]Although the forecast period of Diebold and Li (2006) contains 24 months more, a comparison still seems interesting to conduct.

way in which we estimate the VAR model. We find MPEs (not reported) that are positive, as opposed to negative values in Diebold and Li. A detailed analysis of the prediction errors reveals that for the sample period 1999:1-2000:12, during the yield hike, all the models are consistently producing forecasts that are too low resulting in substantially negative forecasting errors, which explains why Diebold and Li find negative MPEs. For the 12-month horizon we also find that the NS2-AR model substantially outperforms the RW, AR and VAR model. Contrary to Diebold and Li we find that the forecast performance of NS2-VAR is at least similar to that of the AR and VAR models. We do confirm the superior performance of the NS2-AR model for this subsample.

Sample 1999:1 - 2003:12

During this subperiod, interest rates initially go up until the end of 2000 after which they decline sharply by roughly 5% to a level of 1% for the short rate accompanied by a substantial widening of spreads between long and short rates. Forecasts results are shown in Table 4.11 - 4.14. Although adding macro factors again improves forecasts, the only model that seems to be able to compete with the RW is the Bayesian NS1-X model and only consistently so for the longest horizons. The frequentist AR-X model does well for shorter maturities. The VAR model shows a strikingly poor performance with very large positive MPEs indicating that the VAR model cannot cope with the downward trend in interest rates. The Bayesian ATSM-X model does better than the Bayesian VAR and predicts the short end of the curve reasonably well. This shows that imposing no-arbitrage restrictions helps but not enough to beat simple univariate models.

Rolling TRMSPE

The subsample results clearly show that different models perform well during different subsamples. An obvious example is the NS2-AR model which comfortably outperforms all other models for the first subsample but produces disappointing forecasts for the second subsample. Similar conclusions can be drawn for other models. To further illustrate how the forecasting performance of different models varies over time we compute TRMSPEs using a 60-month rolling window. Figures 4.4-4.7 show results for all forecast horizons considered and for a subset of models[20]. Each graph shows the rolling TRMSPE of the RW, AR, VAR, NS1 and ATSM models, either without (left panels) or with macro factors (right panels)

The patterns for the two five year subsamples reappear. TRMSPEs are fairly stable

[20]Note that the graphs only depict model specifications that were estimated using both frequentist and Bayesian inference. As a result, the NS2-AR and NS2-VAR are not included but these graphs are available on request.

Table 4.11: [T]RMSPE 1999:1 - 2003:12, 1-month forecast horizon

	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	107.37	32.59	24.09	24.29	26.26	30.02	32.17	30.97	29.24	107.37	32.59	24.09	24.29	26.26	30.02	32.17	30.97	29.24
Panel A: Individual models without macro factors																		
AR	1.04	1.08	1.12	1.11	1.10	1.05	1.01	1.01	1.01	1.03	1.07	1.10	1.08	1.08	1.03	1.01	1.01	1.00
VAR	1.08	**0.85**	1.04	1.18	1.22	1.22	1.05	1.04	1.07	1.05	**0.86**	1.05	1.19	1.16	1.14	1.03	1.04	1.04
NS2-AR	1.19	1.00	1.24	1.42	1.48	1.36	1.16	1.11	1.14	-	-	-	-	-	-	-	-	-
NS2-VAR	1.06	**0.89**	**0.96**	1.13	1.20	1.18	1.06	1.04	1.07	-	-	-	-	-	-	-	-	-
NS1	1.09	1.12	1.10	1.11	1.13	1.17	1.10	1.05	1.06	1.11	1.09	1.08	1.12	1.14	1.25	1.13	1.05	1.08
ATSM	1.09	**0.86**	**0.96**	1.21	1.41	1.27	1.03	1.05	1.07	1.09	**0.86**	**0.96**	1.20	1.41	1.27	1.03	1.06	1.08
Panel B: Individual models with macro factors																		
AR-X	1.01	**0.99**	**0.99**	1.00	1.02	1.00	1.01	1.01	1.01	1.00	**0.97**	**0.97**	**0.99**	1.01	1.00	1.00	1.01	1.01
VAR-X	1.06	**0.83**	1.02	1.06	1.07	1.18	1.06	1.06	1.04	1.06	**0.83**	1.01	1.05	1.06	1.17	1.06	1.07	1.04
NS2-AR-X	1.14	**0.93**	1.17	1.29	1.34	1.28	1.13	1.12	1.10	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.07	**0.85**	**0.98**	1.10	1.18	1.17	1.07	1.07	1.06	-	-	-	-	-	-	-	-	-
NS1-X	1.07	1.05	1.08	1.06	1.08	1.14	1.10	1.06	1.05	1.25	1.17	1.57	1.58	1.46	1.23	1.12	1.23	1.18
ATSM-X	1.10	**0.84**	**0.96**	1.04	1.26	1.33	1.12	1.05	1.07	1.08	**0.83**	**0.98**	**0.98**	1.22	1.27	1.10	1.05	1.18
Panel C: Forecast combinations without macro factors																		
FC-EW	1.05	**0.92**	1.02	1.12	1.18	1.16	1.03	1.03	1.04	1.03	**0.95**	1.01	1.08	1.12	1.12	1.01	1.02	1.02
FC-MSPE-exp	1.04	**0.91**	1.01	1.10	1.15	1.14	1.02	1.03	1.04	1.02	**0.93**	1.01	1.07	1.10	1.10	1.01	1.02	1.02
FC-MSPE-60	1.04	**0.92**	1.01	1.11	1.16	1.14	1.03	1.03	1.04	1.02	**0.93**	1.01	1.07	1.10	1.10	1.01	1.02	1.02
FC-MSPE-24	1.04	**0.91**	1.01	1.10	1.15	1.13	1.03	1.03	1.04	1.02	**0.93**	1.01	1.07	1.10	1.09	1.01	1.02	1.02
FC-MSPE-12	1.04	**0.90**	1.00	1.10	1.15	1.13	1.02	1.03	1.03	1.02	**0.93**	1.00	1.06	1.09	1.09	1.01	1.02	1.02
BMA-ALL	-	-	-	-	-	-	-	-	-	1.03	**0.98**	1.02	1.07	1.09	1.10	1.02	1.02	1.02
Panel D: Forecast combinations with macro factors																		
FC-EW-X	1.03	**0.88**	**0.96**	1.03	1.09	1.13	1.05	1.04	1.03	**0.99**	**0.87**	**0.93**	**0.93**	**0.99**	1.05	1.00	1.01	**0.99**
FC-MSPE-X-exp	1.03	**0.87**	**0.96**	1.02	1.07	1.11	1.04	1.04	1.02	**0.99**	**0.86**	**0.92**	**0.94**	**0.99**	1.04	1.00	1.01	**0.99**
FC-MSPE-X-60	1.03	**0.87**	**0.96**	1.01	1.07	1.11	1.04	1.04	1.03	**0.99**	**0.86**	**0.92**	**0.94**	1.00	1.04	1.01	1.01	**0.99**
FC-MSPE-X-24	1.03	**0.87**	**0.95**	1.01	1.08	1.11	1.04	1.04	1.03	**0.99**	**0.86**	**0.92**	**0.93**	**0.99**	1.04	1.01	1.02	1.00
FC-MSPE-X-12	1.02	**0.86**	**0.93**	1.01	1.07	1.11	1.04	1.04	1.02	**0.99**	**0.86**	**0.90**	**0.92**	**0.97**	1.03	1.00	1.02	1.00
BMA-X	-	-	-	-	-	-	-	-	-	**0.99**	**0.87**	**0.91**	**0.93**	**0.97**	1.04	1.01	1.02	1.00
Panel E: Forecast combinations with all models																		
FC-EW-ALL	1.04	**0.89**	**0.97**	1.06	1.14	1.15	1.04	1.04	1.03	**0.99**	**0.87**	**0.91**	**0.95**	1.01	1.07	1.01	1.01	1.00
FC-MSPE-ALL-exp	1.03	**0.88**	**0.97**	1.05	1.11	1.13	1.04	1.03	1.03	1.00	**0.87**	**0.94**	**0.97**	1.03	1.07	1.00	1.01	1.00
FC-MSPE-ALL-60	1.04	**0.88**	**0.97**	1.05	1.12	1.13	1.04	1.04	1.03	1.00	**0.87**	**0.94**	**0.97**	1.04	1.07	1.01	1.01	1.00
FC-MSPE-ALL-24	1.04	**0.88**	**0.96**	1.05	1.12	1.13	1.04	1.04	1.03	1.00	**0.87**	**0.93**	**0.97**	1.03	1.06	1.01	1.02	1.01
FC-MSPE-ALL-12	1.03	**0.87**	**0.94**	1.04	1.10	1.12	1.04	1.03	1.03	**0.99**	**0.86**	**0.91**	**0.94**	1.00	1.05	1.01	1.02	1.01
BMA-ALL	-	-	-	-	-	-	-	-	-	1.00	**0.89**	**0.91**	**0.95**	**0.99**	1.06	1.01	1.02	1.01

Note: The table reports forecast results for a 1-month horizon for the out-of-sample period 1999:1 - 2003:12. See Table 4.3 for further details.

Table 4.12: [T]RMSPE 1999:1 - 2003:12, 3-month forecast horizon

	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	197.06	63.52	56.75	56.26	55.97	58.38	55.60	50.80	47.81	197.06	63.52	56.75	56.26	55.97	58.38	55.60	50.80	47.81
Panel A: Individual models without macro factors																		
AR	1.09	1.17	1.17	1.16	1.17	1.10	1.04	1.03	1.02	1.07	1.14	1.14	1.14	1.14	1.07	1.05	1.02	1.00
VAR	1.18	**0.92**	1.10	1.25	1.37	1.35	1.16	1.11	1.08	1.14	1.01	1.17	1.27	1.29	1.24	1.09	1.06	1.02
NS2-AR	1.34	1.12	1.30	1.45	1.58	1.53	1.32	1.23	1.17	-	-	-	-	-	-	-	-	-
NS2-VAR	1.10	**0.91**	1.00	1.13	1.24	1.24	1.11	1.06	1.03	-	-	-	-	-	-	-	-	-
NS1	1.13	1.04	1.09	1.17	1.24	1.25	1.14	1.07	1.02	1.04	1.02	1.02	1.05	1.07	1.10	1.03	1.01	**0.99****
ATSM	1.10	**0.86**	**0.98**	1.17	1.35	1.28	1.03	1.08	1.04	1.10	**0.87**	**0.99**	1.18	1.36	1.29	1.04	1.09	1.04
Panel B: Individual models with macro factors																		
AR-X	1.01	1.01	**0.98**	**0.99**	1.02	1.00	1.01	1.02	1.01	**0.98**	**0.96**	**0.95**	**0.95**	**0.98**	**0.99**	**0.99**	1.01	1.00
VAR-X	1.02	**0.90**	1.01	1.03	1.05	1.09	1.03	1.02	0.96	1.10	**0.85**	**0.98**	1.07	1.16	1.23	1.14	1.09	**0.99**
NS2-AR-X	1.23	**0.99**	1.16	1.27	1.39	1.38	1.24	1.21	1.12	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.10	**0.82**	**0.94**	1.06	1.19	1.23	1.14	1.12	1.04	-	-	-	-	-	-	-	-	-
NS1-X	1.09	**0.92**	**0.99**	1.06	1.15	1.20	1.15	1.11	1.03	**0.99**	**0.82**	**0.90**	**0.94**	**0.98**	1.00	1.00	1.06	1.03
ATSM-X	1.07	**0.78**	**0.85**	**0.98**	1.18	1.23	1.14	1.10	1.04	1.08	**0.79**	**0.86**	**0.95**	1.18	1.24	1.15	1.11	1.12
Panel C: Forecast combinations without macro factors																		
FC-EW	1.10	**0.98**	1.07	1.16	1.25	1.22	1.08	1.06	1.02	1.04	**0.99**	1.05	1.10	1.15	1.12	1.02	1.01	**0.99**
FC-MSPE-exp	1.09	**0.96**	1.06	1.14	1.22	1.19	1.07	1.05	1.02	1.04	**0.98**	1.04	1.09	1.13	1.11	1.02	1.01	**0.99**
FC-MSPE-60	1.09	**0.97**	1.06	1.15	1.23	1.20	1.08	1.06	1.02	1.04	**0.98**	1.04	1.08	1.12	1.11	1.02	1.01	**0.99**
FC-MSPE-24	1.09	**0.97**	1.07	1.16	1.23	1.19	1.07	1.06	1.03	1.04	**0.98**	1.04	1.09	1.13	1.10	1.02	1.02	**0.99**
FC-MSPE-12	1.09	**0.96**	1.05	1.14	1.21	1.19	1.07	1.06	1.02	1.04	**0.97**	1.04	1.09	1.12	1.10	1.02	1.02	**0.99**
BMA-ALL	-	-	-	-	-	-	-	-	-	1.03	1.01	1.03	1.06	1.08	1.08	1.01	1.01	**0.99**
Panel D: Forecast combinations with macro factors																		
FC-EW-X	1.05	**0.89**	**0.96**	1.02	1.10	1.13	1.07	1.06	1.01	**0.98**	**0.85**	**0.89**	**0.93**	1.00	1.05	1.02	1.01	**0.98**
FC-MSPE-X-exp	1.04	**0.88**	**0.95**	1.01	1.08	1.11	1.07	1.06	1.01	**0.98**	**0.85**	**0.90**	**0.93**	1.00	1.03	1.01	1.02	**0.98**
FC-MSPE-X-60	1.04	**0.88**	**0.95**	1.01	1.09	1.12	1.07	1.06	1.01	**0.98**	**0.85**	**0.90**	**0.94**	1.00	1.04	1.02	1.02	**0.98**
FC-MSPE-X-24	1.04	**0.87**	**0.94**	1.01	1.09	1.12	1.07	1.06	1.01	**0.98**	**0.85**	**0.90**	**0.93**	1.00	1.03	1.02	1.02	**0.98**
FC-MSPE-X-12	1.04	**0.87**	**0.92**	1.00	1.08	1.12	1.07	1.07	1.01	**0.98**	**0.85**	**0.88**	**0.92**	**0.98**	1.03	1.02	1.02	**0.99**
BMA-X	-	-	-	-	-	-	-	-	-	**0.98**	**0.87**	**0.90**	**0.93**	**0.98**	1.02	1.01	1.01	**0.98**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	1.07	**0.91**	**0.99**	1.08	1.18	1.18	1.08	1.06	1.01	1.00	**0.89**	**0.94**	**0.99**	1.06	1.08	1.02	1.01	0.98
FC-MSPE-ALL-exp	1.06	**0.90**	**0.98**	1.06	1.15	1.16	1.07	1.06	1.01	1.00	**0.88**	**0.93**	**0.98**	1.05	1.07	1.02	1.01	0.98
FC-MSPE-ALL-60	1.06	**0.90**	**0.99**	1.07	1.16	1.17	1.08	1.06	1.01	1.00	**0.88**	**0.94**	**0.99**	1.05	1.07	1.02	1.01	0.98
FC-MSPE-ALL-24	1.07	**0.91**	1.00	1.09	1.17	1.17	1.09	1.07	1.02	1.01	**0.89**	**0.95**	1.00	1.06	1.07	1.02	1.02	0.98
FC-MSPE-ALL-12	1.06	**0.89**	**0.97**	1.06	1.14	1.16	1.08	1.07	1.02	1.00	**0.88**	**0.92**	**0.98**	1.03	1.06	1.02	1.02	0.99
BMA-ALL	-	-	-	-	-	-	-	-	-	**0.99**	**0.91**	**0.93**	**0.97**	1.02	1.05	1.01	1.01	**0.98**

Note: The table reports forecast results for a 3-month horizon for the out-of-sample period 1999:1 - 2003:12. See Table 4.3 for further details.

Table 4.13: [T]RMSPE 1999:1 - 2003:12, 6-month forecast horizon

	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	308.65	100.60	97.9	97.58	95.57	93.89	84.04	74.90	65.40	308.65	100.60	97.9	97.58	95.57	93.89	84.04	74.90	65.40
Panel A: Individual models without macro factors																		
AR	1.13	1.22	1.19	1.18	1.21	1.14	1.06	1.06	1.04	1.12	1.19	1.18	1.18	1.17	1.12	1.10	1.05	1.10
VAR	1.33	1.17	1.30	1.41	1.51	1.47	1.26	1.22	1.25	1.36	1.30	1.41	1.51	1.55	1.46	1.24	1.21	1.28
NS2-AR	1.34	1.17	1.27	1.37	1.49	1.50	1.33	1.25	1.23	-	-	-	-	-	-	-	-	-
NS2-VAR	1.11	1.02	1.07	1.15	1.23	1.24	1.09	1.04	1.06	-	-	-	-	-	-	-	-	-
NS1	1.15	1.15	1.16	1.21	1.25	1.24	1.11	1.05	1.05	1.02	1.03	1.02	1.03	1.04	1.06	1.00	**0.99**	1.02
ATSM	1.11	**0.98**	1.06	1.19	1.31	1.26	1.02	1.07	1.06	1.12	**0.98**	1.07	1.20	1.32	1.27	1.03	1.08	1.07
Panel B: Individual models with macro factors																		
AR-X	1.00	1.00	**0.98**	**0.98**	1.01	1.01	1.00	1.02	1.01	**0.96**	**0.93**	**0.93**	**0.93**	**0.96**	**0.96**	**0.97**	1.00	**0.98**
VAR-X	1.02	**0.96**	1.02	1.02	1.05	1.06	1.01	**0.99**	**0.98**	1.14	**0.84**	1.06	1.14	1.24	1.28	1.16	1.09	1.04
NS2-AR-X	1.20	1.01	1.11	1.19	1.29	1.32	1.22	1.20	1.17	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.07	**0.84**	**0.93**	1.02	1.14	1.18	1.11	1.09	1.07	-	-	-	-	-	-	-	-	-
NS1-X	1.08	**0.91**	**0.97**	1.04	1.12	1.17	1.11	1.08	1.06	**0.93**	**0.84**	**0.85**	**0.87**	**0.91**	**0.94**	**0.96**	1.00	1.02
ATSM-X	1.04	**0.81**	**0.88**	**0.99**	1.12	1.15	1.08	1.07	1.06	1.04	**0.82**	**0.89**	**0.99**	1.12	1.15	1.08	1.07	1.07
Panel C: Forecast combinations without macro factors																		
FC-EW	1.14	1.07	1.12	1.19	1.26	1.23	1.09	1.06	1.06	1.09	1.07	1.11	1.15	1.19	1.15	1.04	1.03	1.04
FC-MSPE-exp	1.12	1.06	1.11	1.17	1.23	1.20	1.08	1.05	1.05	1.07	1.05	1.09	1.12	1.15	1.13	1.04	1.02	1.04
FC-MSPE-60	1.13	1.07	1.12	1.18	1.24	1.22	1.08	1.06	1.05	1.07	1.05	1.07	1.11	1.14	1.13	1.03	1.02	1.03
FC-MSPE-24	1.13	1.07	1.13	1.19	1.25	1.22	1.09	1.06	1.06	1.08	1.05	1.09	1.13	1.16	1.14	1.04	1.03	1.04
FC-MSPE-12	1.13	1.06	1.12	1.18	1.24	1.21	1.08	1.06	1.07	1.09	1.06	1.11	1.16	1.16	1.13	1.04	1.04	1.08
BMA-ALL	-	-	-	-	-	-	-	-	-	1.04	1.04	1.05	1.06	1.09	1.07	1.01	1.00	1.01
Panel D: Forecast combinations with macro factors																		
FC-EW-X	1.03	**0.90**	**0.95**	1.00	1.07	1.10	1.05	1.04	1.03	**0.98**	**0.89**	**0.92**	**0.95**	1.01	1.03	1.00	**0.99**	**0.97**
FC-MSPE-X-exp	1.02	**0.90**	**0.95**	1.00	1.06	1.08	1.04	1.04	1.03	**0.97**	**0.88**	**0.92**	**0.95**	1.00	1.02	1.00	**0.99**	**0.97**
FC-MSPE-X-60	1.03	**0.91**	**0.97**	1.01	1.07	1.09	1.05	1.04	1.03	**0.98**	**0.89**	**0.93**	**0.96**	1.01	1.02	1.00	**0.99**	**0.97**
FC-MSPE-X-24	1.03	**0.90**	**0.95**	1.00	1.06	1.10	1.06	1.06	1.04	**0.98**	**0.89**	**0.92**	**0.96**	1.00	1.02	1.00	1.00	**0.98**
FC-MSPE-X-12	1.03	**0.89**	**0.93**	**0.99**	1.06	1.11	1.07	1.06	1.05	**0.97**	**0.88**	**0.91**	**0.95**	1.00	1.02	**0.99**	**0.99**	**0.98**
BMA-X	-	-	-	-	-	-	-	-	-	**0.97**	**0.92**	**0.93**	**0.95**	**0.98**	1.00	**0.98**	**0.98**	**0.97**
Panel E: Forecast combinations with all models																		
FC-EW-ALL	1.07	**0.96**	1.01	1.08	1.16	1.16	1.06	1.04	1.03	1.02	**0.96**	**0.99**	1.04	1.09	1.09	1.01	1.00	1.00
FC-MSPE-ALL-exp	1.06	**0.95**	1.00	1.06	1.13	1.14	1.06	1.04	1.03	1.01	**0.93**	**0.97**	1.01	1.06	1.07	1.01	1.00	1.00
FC-MSPE-ALL-60	1.07	**0.96**	1.03	1.08	1.15	1.15	1.06	1.04	1.03	1.01	**0.94**	**0.98**	1.02	1.07	1.07	1.01	1.00	0.99
FC-MSPE-ALL-24	1.09	**0.98**	1.04	1.10	1.17	1.17	1.09	1.08	1.06	1.03	**0.96**	1.01	1.05	1.10	1.09	1.02	1.02	1.01
FC-MSPE-ALL-12	1.09	**0.97**	1.03	1.09	1.16	1.17	1.09	1.09	1.09	1.04	**0.96**	1.02	1.08	1.10	1.08	1.02	1.02	1.04
BMA-ALL	-	-	-	-	-	-	-	-	-	1.00	**0.96**	**0.97**	1.00	1.03	1.04	1.00	**0.99**	**0.98**

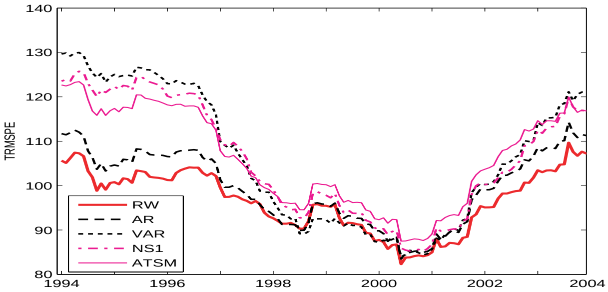
Note: The table reports forecast results for a 6-month horizon for the out-of-sample period 1999:1 - 2003:12. See Table 4.3 for further details.

Table 4.14: [T]RMSPE 1999:1 - 2003:12, 12-month forecast horizon

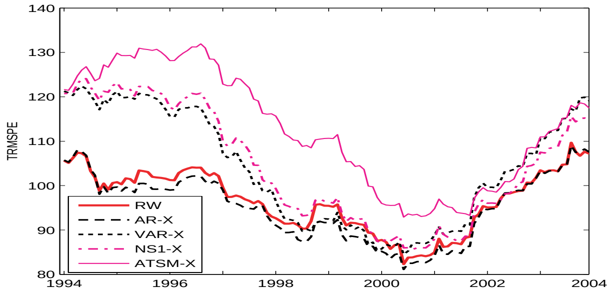
	FREQUENTIST INFERENCE									BAYESIAN INFERENCE								
Models	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y	TRMSPE	1m	3m	6m	1y	2y	5y	7y	10y
RW	479.67	166.03	169.29	170.87	165.13	149.43	118.35	102.82	86.76	479.67	166.03	169.29	170.87	165.13	149.43	118.35	102.82	86.76
Panel A: Individual models without macro factors																		
AR	1.20	1.24	1.20	1.19	1.23	1.21	1.15	1.16	1.15	1.15	1.16	1.19	1.17	1.19	1.15	1.14	1.10	1.10
VAR	1.62	1.44	1.49	1.55	1.62	1.67	1.65	1.70	1.86	1.85	1.75	1.80	1.83	1.86	1.87	1.82	1.86	2.03
NS2-AR	1.28	1.12	1.15	1.20	1.29	1.38	1.36	1.31	1.34	-	-	-	-	-	-	-	-	-
NS2-VAR	1.18	1.14	1.14	1.16	1.21	1.25	1.18	1.14	1.20	-	-	-	-	-	-	-	-	-
NS1	1.21	1.24	1.21	1.21	1.23	1.25	1.18	1.13	1.17	1.02	1.03	1.01	1.01	1.02	1.03	1.01	0.99	1.02
ATSM	1.19	1.12	1.14	1.20	1.27	1.28	1.12	1.19	1.21	1.20	1.12	1.15	1.20	1.28	1.28	1.13	1.19	1.21
Panel B: Individual models with macro factors																		
AR-X	1.03	**0.98**	**0.96**	**0.98**	1.02	1.05	1.06	1.09	1.10	**0.97**	**0.92**	**0.90**	**0.93**	**0.97**	1.00	1.01	1.04	1.03
VAR-X	1.03	**0.99**	1.01	1.02	1.04	1.06	1.02	1.02	1.04	1.37	1.19	1.23	1.27	1.35	1.43	1.43	1.44	1.51
NS2-AR-X	1.17	1.00	1.05	1.08	1.16	1.24	1.25	1.26	1.26	-	-	-	-	-	-	-	-	-
NS2-VAR-X	1.07	**0.90**	**0.95**	**0.99**	1.07	1.15	1.14	1.14	1.13	-	-	-	-	-	-	-	-	-
NS1-X	1.08	**0.93**	**0.96**	1.00	1.08	1.15	1.16	1.15	1.16	**0.93**	**0.90**	**0.89**	**0.90**	**0.92**	**0.94**	**0.95**	**0.97**	**0.99**
ATSM-X	1.05	**0.89**	**0.94**	1.00	1.07	1.12	1.11	1.13	1.13	1.05	**0.89**	**0.94**	**0.99**	1.07	1.12	1.11	1.13	1.17
Panel C: Forecast combinations without macro factors																		
FC-EW	1.20	1.16	1.16	1.19	1.24	1.26	1.20	1.18	1.21	1.19	1.17	1.19	1.20	1.23	1.22	1.15	1.15	1.17
FC-MSPE-exp	1.18	1.14	1.14	1.17	1.21	1.23	1.17	1.16	1.18	1.12	1.11	1.12	1.12	1.15	1.15	1.10	1.10	1.11
FC-MSPE-60	1.18	1.13	1.14	1.16	1.22	1.24	1.17	1.15	1.17	1.10	1.09	1.09	1.10	1.13	1.14	1.09	1.07	1.07
FC-MSPE-24	1.20	1.15	1.16	1.19	1.24	1.26	1.19	1.17	1.18	1.14	1.11	1.13	1.15	1.18	1.18	1.12	1.11	1.09
FC-MSPE-12	1.22	1.17	1.18	1.21	1.26	1.28	1.21	1.20	1.20	1.17	1.13	1.16	1.18	1.22	1.22	1.16	1.15	1.13
BMA-ALL	-	-	-	-	-	-	-	-	-	1.08	1.07	1.07	1.08	1.10	1.10	1.07	1.06	1.05
Panel D: Forecast combinations with macro factors																		
FC-EW-X	1.03	**0.92**	**0.94**	**0.97**	1.03	1.08	1.08	1.09	1.09	1.02	**0.95**	**0.96**	**0.98**	1.02	1.06	1.05	1.05	1.05
FC-MSPE-X-exp	1.03	**0.93**	**0.95**	**0.98**	1.03	1.08	1.08	1.09	1.09	1.02	**0.94**	**0.95**	**0.98**	1.02	1.05	1.05	1.05	1.05
FC-MSPE-X-60	1.05	**0.96**	**0.99**	1.01	1.05	1.09	1.09	1.09	1.10	1.01	**0.95**	**0.96**	**0.98**	1.03	1.05	1.03	1.03	1.02
FC-MSPE-X-24	1.05	**0.94**	**0.97**	**0.99**	1.04	1.09	1.11	1.12	1.12	1.04	**0.96**	**0.97**	**0.99**	1.04	1.07	1.08	1.08	1.07
FC-MSPE-X-12	1.07	**0.94**	**0.97**	1.00	1.06	1.12	1.14	1.15	1.15	1.05	**0.96**	**0.97**	**0.99**	1.04	1.08	1.09	1.11	1.13
BMA-X	-	-	-	-	-	-	-	-	-	1.00	**0.96**	**0.97**	**0.98**	1.00	1.02	1.02	1.02	1.03
Panel E: Forecast combinations with all models																		
FC-EW-ALL	1.10	1.01	1.03	1.06	1.12	1.16	1.12	1.11	1.12	1.10	1.05	1.06	1.08	1.13	1.14	1.10	1.10	1.11
FC-MSPE-ALL-exp	1.08	1.00	1.02	1.05	1.10	1.14	1.11	1.10	1.11	1.06	1.00	1.02	1.03	1.08	1.10	1.08	1.08	1.08
FC-MSPE-ALL-60	1.09	1.02	1.04	1.07	1.12	1.15	1.10	1.09	1.10	1.05	1.01	1.02	1.04	1.08	1.10	1.06	1.05	1.04
FC-MSPE-ALL-24	1.15	1.06	1.08	1.11	1.17	1.21	1.18	1.17	1.17	1.10	1.04	1.06	1.08	1.13	1.15	1.12	1.11	1.09
FC-MSPE-ALL-12	1.21	1.09	1.12	1.16	1.22	1.27	1.25	1.25	1.23	1.15	1.08	1.11	1.14	1.18	1.20	1.15	1.15	1.15
BMA-ALL	-	-	-	-	-	-	-	-	-	1.05	1.02	1.02	1.04	1.07	1.08	1.05	1.05	1.05

Note: The table reports forecast results for a 12-month horizon for the out-of-sample period 1999:1 - 2003:12. See Table 4.3 for further details.

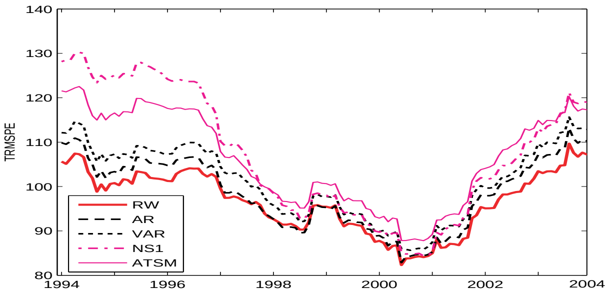
Figure 4.4: 60-month moving TRMSPE: 1-month forecast horizon



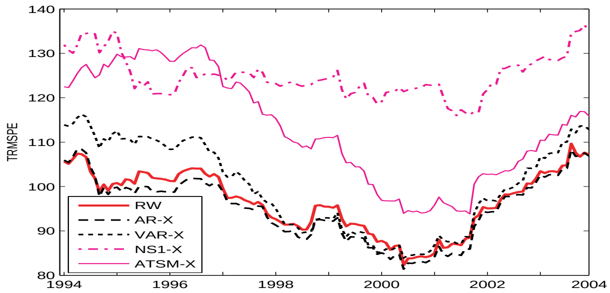
(a) classical inference



(b) classical inference

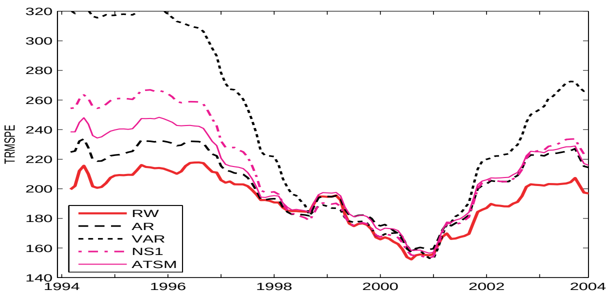


(c) Bayesian inference

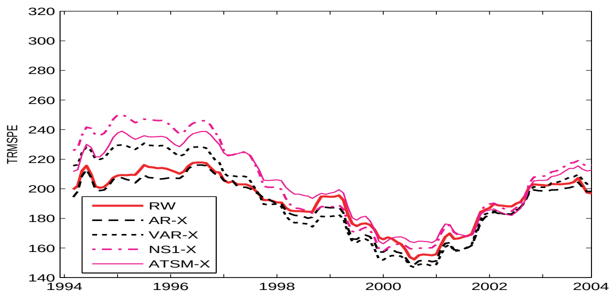


(d) Bayesian inference

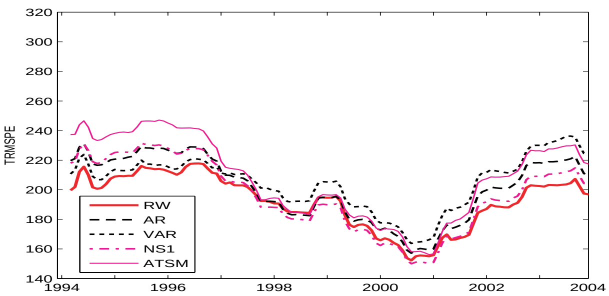
Figure 4.5: 60-month moving TRMSPE: 3-month forecast horizon



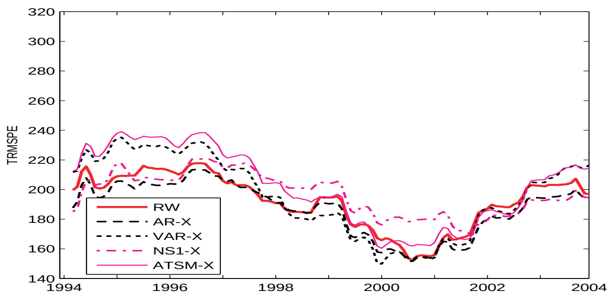
(a) classical inference



(b) classical inference



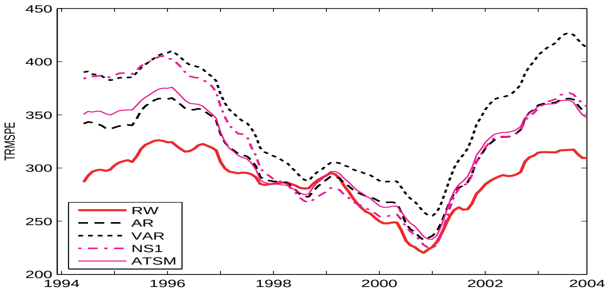
(c) Bayesian inference



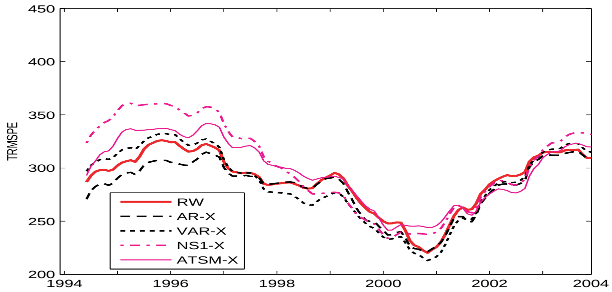
(d) Bayesian inference

Note: The figure presents the 60-month rolling window TRMSPE for individual models in the left panels and for individual models augmented with macro factors in the right panels. The TRMSPE is shown for the out-of-sample period 1994:1-2003:12 for a 1-month horizon in Figure 4.4 and a 3-month horizon in Figure 4.5. The models depicted are the Random Walk [RW], first order (Vector) Autoregressive [(V)AR], State-Space Nelson-Siegel [NS1] and the affine [ATSM] model. The affix 'X' indicates that macro factors have been added as additional explanatory variables.

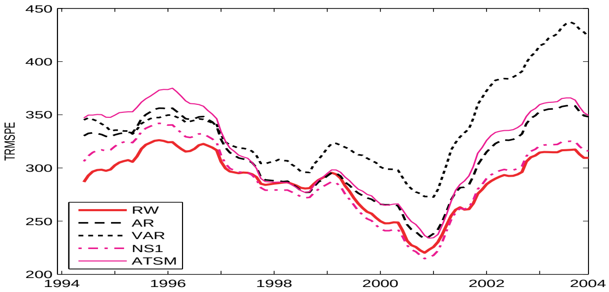
Figure 4.6: 60-month moving TRMSPE: 6-month forecast horizon



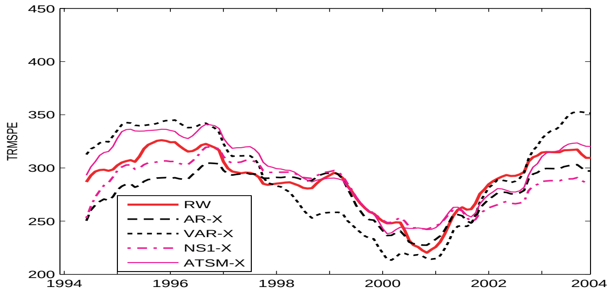
(a) classical inference



(b) classical inference

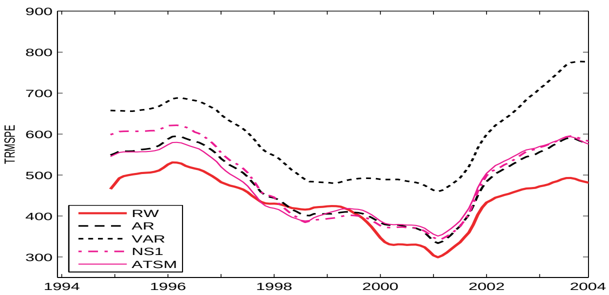


(c) Bayesian inference

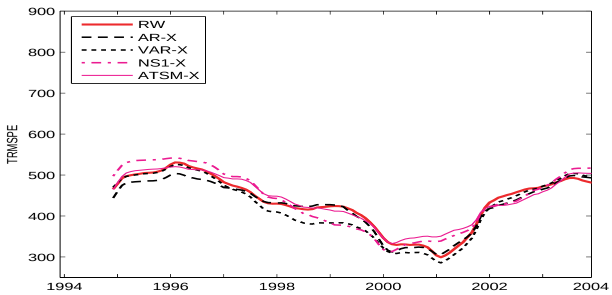


(d) Bayesian inference

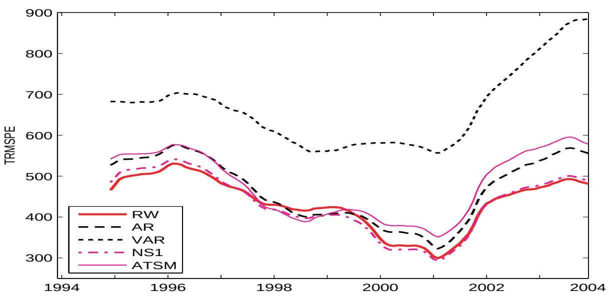
Figure 4.7: 60-month moving TRMSPE: 12-month forecast horizon



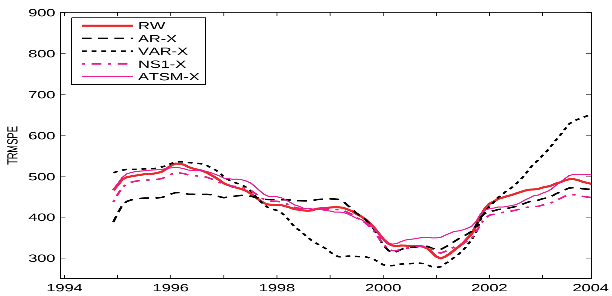
(a) classical inference



(b) classical inference



(c) Bayesian inference



(d) Bayesian inference

Note: The figure presents the 60-month rolling window TRMSPE for individual models in the left panels and for individual models augmented with macro factors in the right panels. The TRMSPE is shown for the out-of-sample period 1994:1-2003:12 for a 6-month horizon in Figure 4.6 and a 12-month horizon in Figure 4.7. See Figures 4.4-4.5 for further details.

until 1997 after which a decreasing trend sets in lasting until mid 2000. The high degree of interest rate predictability during the 1994-1998 subperiod is the cause of the decreasingly low TRMSPEs for the period 1998-2001. From 2001 onwards a sharp increase is visible in TRMSPEs indicating large forecasting errors due to the sharp decline in interest rate levels and the widening of spreads during this period. Zooming in on the performance of individual models, we notice that the random walk is one of the best models at the beginning and at the end of the forecasting period. During the 1998-2001 period the random walk tends to be outperformed by the AR-X, VAR-X and NS1-X models. An opposite pattern is visible for the ATSM model which performs well only in the middle of the out-of-sample period.

The main point to take from these graphs is that the performance of individual models varies substantially over time and establishing a clear-cut ordering of the models which holds across the entire 1994-2003 period seems infeasible. Therefore, believing in a single forecasting model may be dangerous. In the next section, we therefore discuss several forecast combination techniques.

4.5 Forecast combination

Our subsample and rolling TRMPSE analysis reveals that it is seems impossible to identify a single model that consistently outperforms the random walk across all subperiods. The forecasting ability of individual models varies considerably over time. It seems that each model may play a complementary role in approximating the data generating process, at least during subperiods. Model uncertainty is troublesome if one has hopes of obtaining a single model for forecasting or investment purposes. A worthwhile endeavor for cushioning the effects of model uncertainty is to *combine* the forecasts of different models. In this section we examine several forecast combination schemes. Two combination methods are standard approaches and can be applied to combine frequentist as well as Bayesian forecasts. We also investigate a third combination method which is a truly Bayesian approach that can only be applied to Bayesian forecasts. We first discuss the different methods and then move on to examine the forecast combination results in comparison to the results of the individual models.

4.5.1 Forecast combination: schemes

Scheme 1: Equally weighted forecasts

The first forecast combination method assigns an equal weight to the forecasts from all individual models. Assuming we are combining forecasts from M different models, each weight is the same and equal to $w_{T+h,m}^{(\tau_i)} = 1/M$ for $m = 1, \ldots, M$. The equally weighted combined forecast for a h-month horizon for any maturity τ_i is therefore given by $\widehat{y}_{T+h}^{(\tau_i)} = \sum_{m=1}^{M} w_{T+h,m}^{(\tau_i)} \widehat{y}_{T+h,m}^{(\tau_i)}$ which we denote as Forecast Combination - Equally Weighted (**FC-EW**). As explained in Timmermann (2006) this method is likely to work well if forecast errors from different models have similar variances and are highly correlated, which is certainly the case here.

Scheme 2: Inverted MSPE-weighted forecasts

The second forecast combination scheme we examine uses weights that take into account historical relative performance. Model weights are based on each model's (inverted) MSPE relative to those of all other models, computed over a window of the previous υ months and we denote these by Forecast Combination - MSPE (**FC-MSPE**)[21]. The weight for model m is computed as $w_{T+h,m}^{(\tau_i)} = \frac{1/\text{MSPE}_{T+h,m}^{(\tau_i)}}{\sum_{m=1}^{M}(1/\text{MSPE}_{T+h,m}^{(\tau_i)})}$ where $\text{MSPE}_{T+h,m}^{(\tau_i)} = \frac{1}{\upsilon}\sum_{r=1}^{\upsilon}(\widehat{y}_{T+h-r|T-r,m}^{(\tau_i)} - y_{T+h-r}^{(\tau_i)})^2$. A model with a lower MSPE is given a relatively larger weight than a worse performing model, see Timmermann (2006) for discussion and Stock and Watson (2004) for an application to forecasting GDP growth[22]. Which value should be used for υ is difficult to determine *a priori*. Using a smaller window will make weights more responsive to changes in models' forecasting accuracy but it will also make them more noisy. The optimal choice of υ will therefore need to be determined empirically. Here we use four different windows to compute model weights. We use an expanding window where υ is initially set equal to 60 months but which increases with every new yield realization that becomes available and we denote the resulting combination forecast as **FC-MSPE-exp**. We also apply moving windows of different length, in particular $\upsilon = 12$, 24 and 60 months. We denote these by **FC-MSPE-12**, **FC-MSPE-24** and **FC-MSPE-60** respectively.

Scheme 3: Bayesian predictive likelihood

The third and final combination scheme we consider is a purely Bayesian model averaging scheme, which we denote by **BMA**[23], and is based on the predictive likelihood approach

[21]Note that whereas in the tables we report results for the Root MSPE, Timmermann (2006) argue that it is better to use the MSPE to construct model weights.

[22]The weights applied in this and the previous forecast combination scheme are always bounded between zero and one. Other approaches for which this does not necessarily need to be the case, in particular OLS-based weights (see again Timmermann, 2006), proved to be problematic here due to multicollinearity problems between the different forecasts. This resulted in often extreme (offsetting) weights we therefore did not further pursue these approaches.

[23]In the remainder of the text, we often refer to this third scheme as forecast combination. With a slight abuse of denotation we share BMA in the class of forecast combination methods which, strictly speaking, is incorrect since BMA *averages* models instead of *combining* models.

proposed by Geweke and Whiteman (2006). The probability of the *realized* value at time $T + h$ is evaluated under the Bayesian predictive density for $T + h$ for a given model conditional on the information at time T. The resulting probability is called predictive likelihood. Geweke and Whiteman (2006) apply these probabilities to average individual models. The realized value will fall near the center of the predictive density of a given model if this density is accurate. The particular model then receives a large weight relative to a model for which the realization ends up far out in the tail of its predictive density.

The approach of Geweke and Whiteman (2006) is an alternative to the most commonly used BMA method based on the marginal likelihood, see for example Madigan and Raftery (1994). We choose the predictive likelihood BMA for three reasons. Firstly, the predictive likelihood is an out-of-sample performance measure, on contrary the marginal likelihood is an in-sample fitting measure. Secondly, the marginal likelihood of highly nonlinear models, such as the Nelson-Siegel and affine models, cannot be derived analytically and may be very difficult to compute by Monte Carlo simulation. Thirdly, Eklund and Karlsson (2007) show, in a simulation setting and in an empirical application to Swedish inflation, that model weights based on the predictive likelihood have better small sample properties and result in better out-of-sample performance than weights based on the traditional marginal likelihood measure.

Whereas we refer to the appendix for specific details, we do want to briefly discuss a major difference between our forecast combination approach and that of Eklund and Karlsson (2007). Unlike in their study, we do not apply the system of updating and probability forecasting *prequential*, as defined by Dawid (1984). We compute the predictive density for month $T+h$ using information up until month T and we evaluate the *realized* value for time $T+h$ using the same density. The resulting probability is then used to compute the weight for model m in constructing the forecast for $T + 2h$ made at time $T + h$. Eklund and Karlsson (2007) on the other hand evaluate the fit of the predictive density over a small number of observations, by means of the predictive likelihood, and then update the probability density for the forecasts. The latter approach results in weights which are based more on the fit of the model, even when using out-of-sample data, than on the probability of out-of-sample realized values. In an unreported simulation exercise we find that our approach reacts faster to out-of-sample uncertainty and instability since it is not constrained to give more probability to the model which provide the best fit of predicted values.

4.5.2 Forecast combination results

A important question to answer when combining forecasts is which models should be included. Here we combine forecasts using three different sets of models. First we include only those specifications that do not incorporate macro factors ($M = 7$ for the models estimated with frequentist methods and $M = 5$ for the Bayesian counterpart); second, we use only those model specifications that do incorporate macro factors (again $M = 7$ for the models estimated with frequentist methods and $M = 5$ for the Bayesian counterpart) and finally, we simply combine all specifications ($M = 13$ and $M = 9$ respectively)[24]. By again making the distinction between models with and without macro factors we can assess the added value of including macroeconomic information also for the combined forecasts, just like we did for the individual models. The only model that is always included in the forecast combinations is the random walk.

Full sample 1994:1 - 2003:12

The results of the forecast combination methods for the 1994-2003 period are reported in Panels C-E of Tables 4.3-4.6. The following main overall conclusions can be drawn. Firstly, it holds for all horizons that forecast combinations methods are a valuable alternative compared to selecting any individual model, especially when combining forecasts from models estimated with Bayesian methods. The reported TRMSPE numbers show that the best combination scheme always outperforms the best individual model as well as the random walk, although the differences are not statistically significant. Secondly, Panels C-E show that combining forecasts works increasingly well for longer forecast horizons. Indeed, for the 6-month and 12-month horizons, the best combination scheme outperforms the random walk and the best individual model by several percentage points in terms of relative RMSPEs. Thirdly, results are particularly encouraging for long maturities. All the individual models tend to forecast maturities beyond 5 years rather poorly, with some exceptions such as the VAR-X and NS1-X models. This is not the case, however, for the combination schemes which outperform the random walk by up to 7% for the 6-month horizon (FC-MSPE-X) and 8-9% for the 12-month forecast horizon (again FC-MSPE-X). This is an important result as other studies have documented the difficulty of accurately forecasting long maturities with individual models. The claim that individual models pro-

[24] Many other subsets can of course be selected. Aiolfi and Timmermann (2006) suggest filtering out the worst performing model(s) in an initial step. Preliminary analysis suggests that doing so does not lead to much improvement in forecasting performance in our case. However, a more thorough selection procedure than simply including all available models as applied here, will most likely lead to better results for the forecast combination methods. Although this is a very interesting issue to examine in more detail, our main point here is that want to show the benefits of combining forecasts as an alternative to putting all one's eggs in a single model basket.

vide complementary information definitely seems to hold for longer maturities. Fourthly, averaging models with macro factors is the superior combination approach. When we compare the forecast combinations with different model sets, in particular models with and without macro factors, there seems to be no doubt that combining forecasts from only the models that include macro factors provides the most accurate results. When models with macro factors are averaged (Panels D), the resulting statistics are almost always below those of the Random Walk, irrespective of the considered horizon, and this is true independently of the averaging scheme used. On contrary, forecasts from combining models without macro factors (Panels C) always have relative RMSPEs above one. Including all models in the averaging strategy therefore also does not seem to be the most favorable approach in particular not with a longer forecast horizon. Finally, comparing the different forecast combination schemes in more detail, we observe that MSPE-based weights work better than giving each model an equal weight. Differences are most pronounced for long maturities and a long forecast horizon. For example, for a 12-month horizon with the 10-year maturity using Bayesian inference, the relative MSPE of FC-EW-X is 0.98 whereas that of FC-MSPE-X-exp is 6% lower at 0.92. With respect to the length of window that should be used to compute the MSPE weights, we find that weights that are based on the relative performance over a long history give the most accurate forecasts. Using an expanding window or a 60-month rolling window works very well whereas using a shorter history deteriorates the combination results.

The BMA scheme gives very similar results as the equal weight scheme. Bayesian model averaging has the attractive feature of being able to assign near-zero weights to, and thereby effectively eliminating, the worst performing models. Although BMA outperforms the FC-MSPE with $\upsilon = 12$ and 24, like these schemes it assigns probability to models using only the very recent historical performance. Our results indicate that a long history is important to accurately assign weights to models.

By analyzing the forecast combination results for the two five-year sub periods we can judge the robustness of the above conclusions.

Subsample 1994:1 - 1998:12

For this period, which is characterized by a high level of predictability in general and with some individual models performing particularly well, forecast combinations are still attractive as reported in Tables 4.7-4.10. Improvements with respect to the random walk are statistically significant, often even at the 99% confidence level. For short forecast horizons, some individual models, mainly the NS2-AR, outperform the forecast combination schemes. For the 6-month and 12-month horizons, forecast combinations with macroeconomic information, based on the MSPE-weights using a long historical window are the most accurate forecasting methods. It is interesting that whereas for the individual models (except for the VAR model) adding macro factors worsens forecasting performance, for the forecast combination methods adding macro factors is very beneficial.

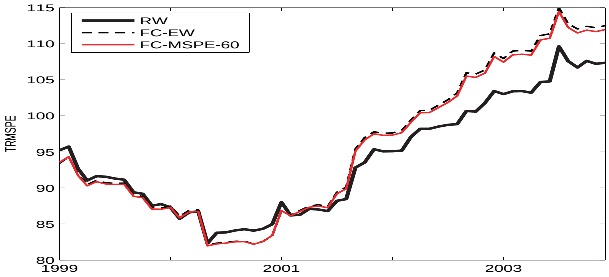
Panel E of Table 4.9 and 4.10 shows that for this subsample combining all models seems to work somewhat better than just the macro models, as judged by the TRMSPE. However, this outperformance is achieved through the short and medium maturities which, for the frequentist results, can be explained by the stellar performance of the NS2-AR model. Nevertheless, the combination methods that average only over models with macro factors are by far the most accurate for long maturities. We find that the MSPE decreases by around 20% compared to the random walk and over 10% compared to the best individual model for the 12-month forecast horizon.

Subample 1999:1 - 2003:12

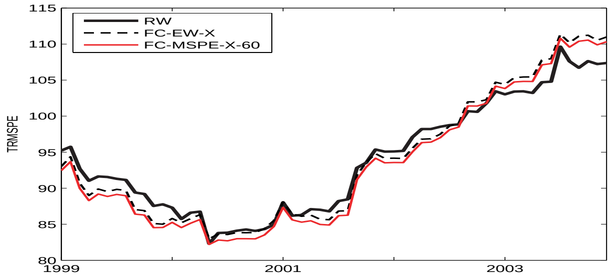
Our analysis in section 5.2.1 shows that the Bayesian Nelson Siegel model with macro factors forecasts very accurately in this subsample. Forecast combinations provide similar results for short horizons, but results are worse for $h = 6$ and $h = 12$. The MSPE combination scheme with only macro factors and a long history to base the weights on still is the best forecast combination approach. It again outperforms nearly all the individual models but is still less accurate than the random walk. The somewhat disappointing results for this forecast combination scheme can, however, be explained by the way model weights are determined. One of the best performing individual models in the 1994-1998 subsample is the VAR-X model. With either an expanding window or a 60-month moving window, the VAR-X model will initially receive a large weight relative to other models during the 1999-2003 period with the MSPE combination scheme. However, the VAR-X model has low predictability in this subsample which negatively influences the results of the combination methods. As the MSPE combination scheme is solely based on past performance, it cannot account for structural changes in the forecasting performance of individual models. Using a smaller moving window $\upsilon = 12,\ 24$ does not help although the results for BMA do suggest that a shorter history may be worthwhile[25]. More accurate combination schemes would ideally be able to account for structural changes.

[25]The potential problem with the MSPE-based weight scheme is that the squared forecasts errors are all given a weight of one when summing these to compute the MSPE. To assign more weight to the most recent forecast errors we therefore experimented with a weighted MSPE as suggested by Diebold and Pauly (1987) and we computed the weighted MSPE as follows: WMSPE $= \frac{1}{\upsilon}\sum_{r=1}^{\upsilon}\lambda^{r-1}(\widehat{y}^{(\tau_i)}_{T+h-r|T-r} - y^{(\tau_i)}_{T+h-r})^2$. The factor λ^{r-1} introduces exponentially decreasing weights as 1, λ, λ^2, ..., starting from the most recent forecast error. We set $\lambda = 0.9439$ such that the 12 most recent forecasts receive 50% of the total weight given. Although this method does indeed give smaller weights to the VAR-X model, the overall forecasting performance of the MSPE-weighted scheme did not improve.

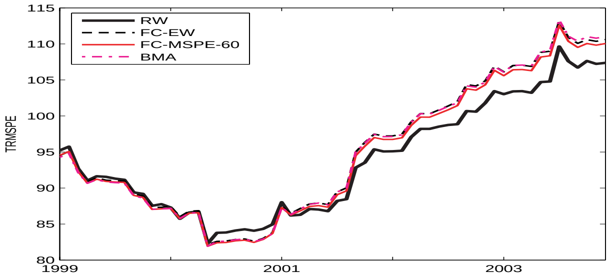
Figure 4.8: 60-month moving TRMSPE: 1-month forecast horizon



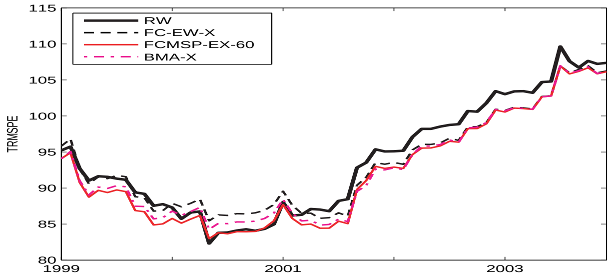
(a) classical inference



(b) classical inference

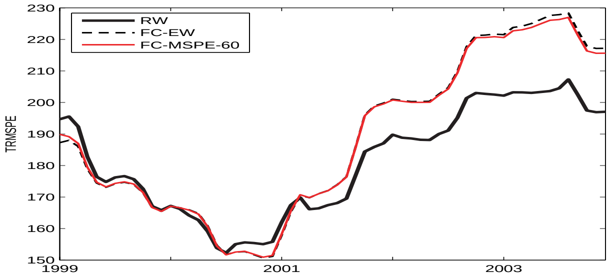


(c) Bayesian inference

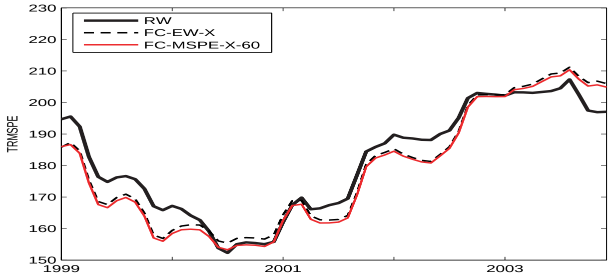


(d) Bayesian inference

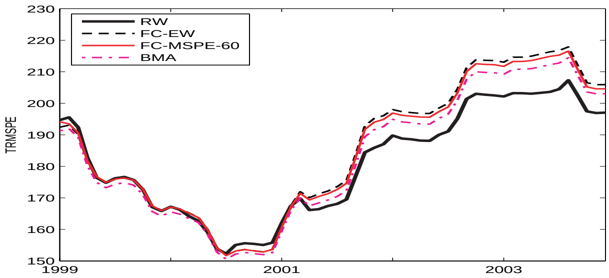
Figure 4.9: 60-month moving TRMSPE: 3-month forecast horizon



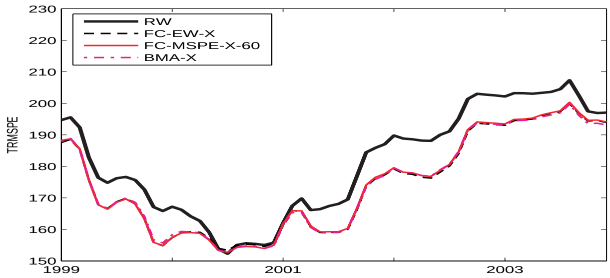
(a) classical inference



(b) classical inference



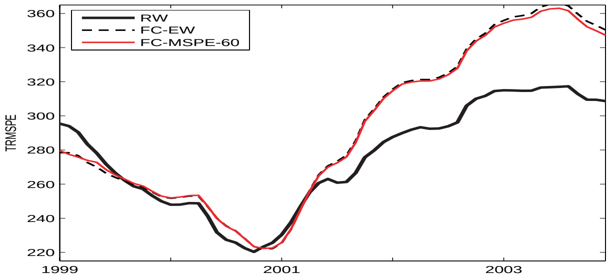
(c) Bayesian inference



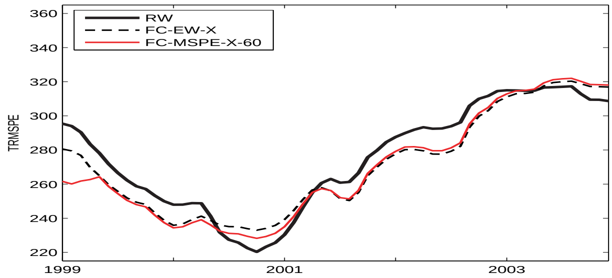
(d) Bayesian inference

Note: The figure presents the 60-month moving average TRMSPE for forecast combination methods using individual models without macro factors (left panel) and with macro factors (right panel). The TRMSPE is shown for the out-of-sample period 1999:1-2003:12 for a 1-month horizon in Figure 4.8 and a 3-month horizon in Figure 4.9. Results are depicted for the Random Walk [RW], combined forecasts using equal weights [FC-EW], MSPE-based weights based on a moving window of the last 60 forecasts [FC-MSPE-60] and combined forecasts using the Bayesian model averaging approach [BMA]. The affix 'X' indicates that only individual models with macro factors are combined.

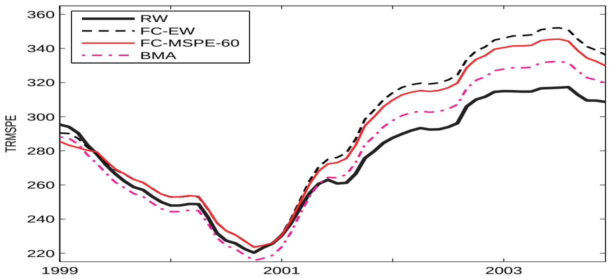
Figure 4.10: 60-month moving TRMSPE: 6-month forecast horizon



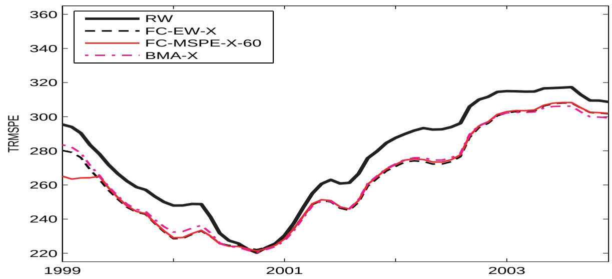
(a) classical inference



(b) classical inference

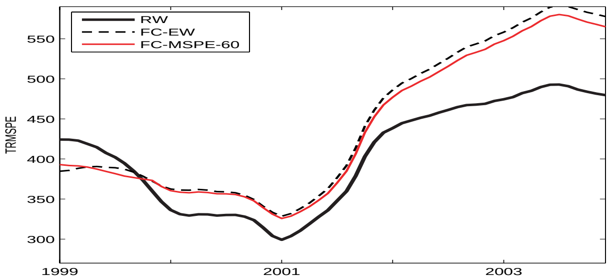


(c) Bayesian inference

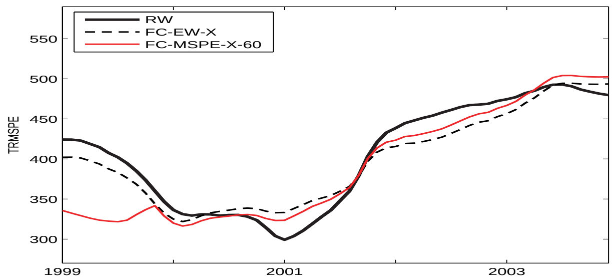


(d) Bayesian inference

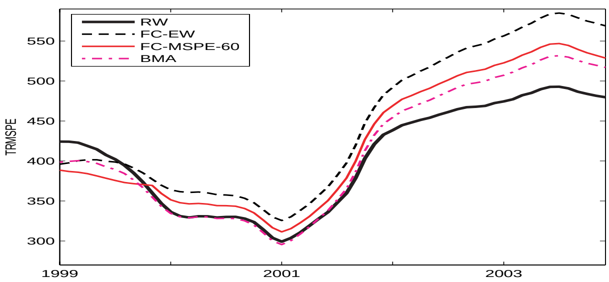
Figure 4.11: 60-month moving TRMSPE: 12-month forecast horizon



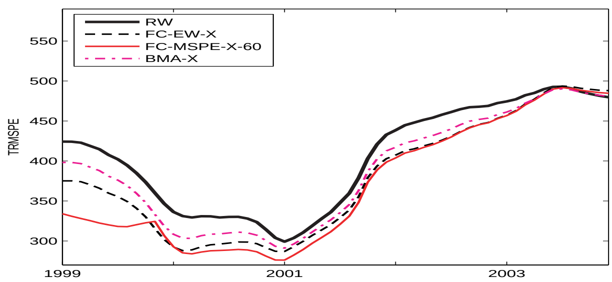
(a) classical inference



(b) classical inference



(c) Bayesian inference



(d) Bayesian inference

Note: The figure presents the 60-month moving average TRMSPE for forecast combination methods using individual models without macro factors (left panel) and with macro factors (right panel). The TRMSPE is shown for the out-of-sample period 1999:1-2003:12 for a 6-month horizon in Figure 4.10 and a 12-month horizon in Figure 4.11. See Figures 4.8-4.9 for further details.

Rolling TRSMPE

A valid question to ask is to what extent our results for the forecast combination schemes are also sample specific. An answer to this question can be given by considering Figures 4.8-4.10. These graphs show the 60-month rolling TRMSPE for the equally-weighted and MSPE-weighted combination schemes for the period 1999-2003[26], without macro factors (left panels) and with macro factors (right panels) and for frequentist (Panels [a] and [b]) as well as Bayesian estimation methods (Panels [c] and [d]).

The graphs show that the forecast combination schemes which incorporate macro factors always outperform the schemes that do not incorporate macroeconomic information, irrespective of the forecast horizon and the estimation method[27]. What is most striking though is that the averaging schemes with macro factors outperform the random walk for nearly every five-year subperiod, except for a few samples ending in either the second half of 2000 or at the end of 2003. This is particularly true when model forecasts are constructed with Bayesian techniques. The random walk TRMSPE lines in panel (d) of Figures 4.9-4.11 are clearly above those of the FC-MSPE-X-60 scheme which is the best performing combination method. Consequently, the performance of the forecast combinations is very stable across time and indeed much more stable than for individual models. In that respect, our choice of the second subsample is even somewhat unfortunate as the reported results for this sample do not do justice to the combination schemes.

4.6 Conclusion

This Chapter addresses the task of forecasting the term structure of interest rates. Several recent studies have shown that significant steps forward are being made in this area. We contribute to the existing literature by assessing the importance of incorporating macroeconomic information, parameter uncertainty, and, in particular, model uncertainty. Our results show that these issues are worth addressing since they improve interest rate forecasts.

We examine the forecast accuracy of a range of models with varying degrees of complexity. We assess model forecasts over a ten-year out-of-sample period, using the entire period as well as several subperiods to show that the predictive ability of individual models varies over time considerably. Models that incorporate macroeconomic variables seem more accurate in subperiods during which the uncertainty about the future path of interest rates is substantial. As an example we mention the period 2000-2003 when spreads

[26]We need the forecasts of the first subsample to initialized the rolling TRMSPE statistics.

[27]Note that the rolling TMSPEs seem to be more stable over time when the forecast horizon lengthens which is counterintuitive. However, this is only due to the scaling of the vertical axes in the graphs.

were high. Models without macro information do particularly well in subperiods where the term structure has a more stable pattern such as in the early 1990s.

The fact that different models forecast well in different subperiods confirms ex-post that alternative model specifications play a complementary role in approximating the data generating process. Our subsample results provide a strong claim for using forecast combination techniques as opposed to believing in a single model. Our model combination results show that recognizing model uncertainty and mitigating the likely effects, leads to substantial gains in interest rate forecastability. We show that combining forecasts of models that incorporate macro factors are superior to forecasts of any individual model as well as the random walk benchmark. Additionally, the outperformance of the optimal combination scheme which assigns weights to models based on the relative historical performance over a long sample is very stable over time. We obtain the largest gains in forecastability for long maturities.

We feel that our results open up exciting avenues for further research. In this study we have only considered very generic models, in particular in our use of a three-factor Gaussian affine model. It would therefore be interesting to expand the model consideration set with more sophisticated models such as the FAVAR models of Mönch (2006a) or the structural model by Hordahl *et al.* (2006) both of which have been found to forecast well. More sophisticated ways of combining forecasts are worth addressing as well, see e.g. Guidolin and Timmermann (2007) who use a combination scheme with time-varying weights where weights have regime switching dynamics. In terms of incorporating parameter uncertainty, much more work can be done on the use of sensible informative priors. As an example we mention the use of adaptive priors that could take into account likely changes in yield dynamics due to clear political or economic reasons. Other, technical, issues that could be addressed are more specifically related to estimation and forecasting procedures. For example, changes in yield dynamics could also be accounted for by using rolling estimation windows instead of the expanding window which we have used here.

Appendix: Estimation details

4A Individual models

In this appendix we provide details on how we perform inference on the parameters of the models in Section 3. We discuss each model separately and we distinguish between frequentist and Bayesian inference.

4A.1 AR model

Frequentist Inference

We estimate the parameters $(c^{(\tau_i)}, \phi^{(\tau_i)}, \psi^{(\tau_i)})$ using standard OLS. Given the parameter estimates we construct iterated forecasts as

$$\widehat{y}_{T+h}^{(\tau_i)} = \hat{c}^{(\tau_i)} + \hat{\phi}^{(\tau_i)} \widehat{y}_{T+h-1}^{(\tau_i)} + \hat{\psi}^{(\tau_i)\prime} \widehat{X}_{T+h} \tag{4A.1}$$

with $\widehat{y}_T^{(\tau_i)} = y_T^{(\tau_i)}$. We construct forecast both with and without the macroeconomic factors. The forecasts of the macro factors, $\widehat{X}_{T+h}$, are iterated forecasts constructed from the VAR(3) macro model.

Bayesian Inference

For the Bayesian inference, we use a Normal-Gamma conjugate prior for the parameters $(c^{(\tau_i)}, \phi^{(\tau_i)}, \psi^{(\tau_i)}, \sigma^{2(\tau_i)})$,

$$(c^{(\tau_i)}, \phi^{(\tau_i)}, \psi^{(\tau_i)}, \sigma^{2(\tau_i)})' \sim NG(\underline{b}, \underline{v}, \underline{s^2}, \underline{\nu}) \tag{4A.2}$$

The marginal posterior densities of the parameters and the predictive density of $y_{T+h}^{(\tau_i)}$, conditional on $y_T^{(\tau_i)}$ and X_{T+h}, can be derived using standard Bayesian results, see for example Koop (2003).

4A.2 VAR model

Frequentist Inference

We estimate the equation parameters (c, Φ, Ψ) in (4.5) using equation-by-equation OLS. Forecasts are obtained as

$$\widehat{Y}_{T+h} = \widehat{c} + \widehat{\Phi}\widehat{F}_{T+h-1} + \widehat{\Psi}\widehat{X}_{T+h} \tag{4A.3}$$

We construct the yield factor forecasts, $\widehat{F}_{T+h-1}$, by first calculating the principal component factor loadings using data only up until month T and then multiplying these with the iterated yields forecasts.

Bayesian Inference

We apply direct simulation to draw inference on VAR model. Note that this is a novel approach as the literature commonly uses MCMC simulation algorithms. Direct simulation is faster and more precise since truly independent draws are used. Our derivation is based on Zellner (1971), who provides all the necessary computations with diffuse priors, and we extend the analysis to include informative priors[28].

[28] We present the main results. Details of the derivations are available upon request.

Prior Specification We apply informative prior densities for the parameter matrices $\Pi = [c\ \Phi\ \Psi]$ and S in (4.5). For computational tractability we select the following conjugate priors:

$$\Pi|S \sim MN(\underline{B}, S \otimes \underline{V}) \tag{4A.4}$$

and

$$S \sim IW(\underline{S}, \underline{\mu}) \tag{4A.5}$$

where MN indicates the mactrivariate normal distribution with mean $\underline{B}$ and variance matrix $S \otimes \underline{V}$, and where IW indicates the Inverted Wishart distribution.

Posterior Simulation The likelihood function of Y_T for the VAR is given by

$$p(Y_T|F_{T-1}, X_T, \Pi, S) = (2\pi)^{-TN/2}|S|^{-T/2}\exp[-\frac{1}{2}tr(S^{-1}(Y_T - Z_T\Pi)'(Y_T - Z_T\Pi))] \tag{4A.6}$$

where $Z_T = (e_N, F'_{T-1}, X'_T)$ and e_N is a $(N \times 1)$ vector of ones. If we combine (4A.6) with the prior densities in (4A.4)–(4A.5) we obtain the joint posterior density for (Π, S) as

$$\begin{aligned} &p(\Pi, S|Y_T, F_{T-1}, X_T) = p(Y_T|F_{T-1}, X_T, \Pi, S)p(\Pi|S)p(S) \\ &\propto |S|^{-(T+N+\nu+1)/2}\exp(-\tfrac{1}{2}tr(S^{-1}(\underline{S} + (Y_T - Z_T\Pi)'(Y_T - Z_T\Pi) + (\Pi - \underline{B})'\underline{V}^{-1}(\Pi - \underline{B})))) \end{aligned} \tag{4A.7}$$

where $\nu = G + \underline{\nu}$ with G the number of columns of Π. If we define $W_T = (Y_T, \underline{V}^{-1/2}\underline{B})'$, $V_T = (Z_T, \underline{V}^{-1/2})$ and apply the decomposition rule and the Inverted Wishart integration step, the posterior density for Π, conditional on (Y_T, F_{T-1}, X_T), will be a generalized t-distribution with location parameter $\widehat{\Pi} = (V'V)^{-1}V'W$, scale parameters $\underline{S} + (W_T - V_T\widehat{\Pi})'(W_T - V_T\widehat{\Pi})$ and $(Z'_T Z_T + \underline{V}^{-1})$ and $T + \underline{\nu}$ degrees of freedom. That is,

$$p(\Pi|Y_T, F_{T-1}, X_T) \propto |\underline{S} + (W_T - V_T\widehat{\Pi})'(W_T - V_T\widehat{\Pi}) + (\Pi - \widehat{\Pi})'(Z'_T Z_T + \underline{V}^{-1})(\Pi - \widehat{\Pi})|^{-(T+\nu)/2} \tag{4A.8}$$

The posterior density of S conditional on (Y_T, F_{T-1}, X_T) is:

$$S|Y_T, F_{T-1}, X_T \sim IW(\underline{S} + (W_T - V_T\widehat{\Pi})'(W_T - V_T\widehat{\Pi}), T + \nu) \tag{4A.9}$$

Forecasting The predictive density conditional on (Y_T, X_T) and (F_{T+h-1}, X_{T+h}) is defined as:

$$\begin{aligned} &p(Y_{T+h}|Y_T, X_T, F_{T+h-1}, X_{T+h}) = \int\int p(Y_{T+h}, \Pi, S|Y_T, X_T, F_{T+h-1}, X_{T+h})d\Pi dS \\ &= \int\int p(Y_{T+h}|F_{T+h-1}, X_{T+h}, \Pi, S)p(\Pi, S|Y_T, X_T)d\Pi dS \end{aligned} \tag{4A.10}$$

By applying the inverted Wishart step to (4A.10), and integrating with respect to Π, we have:

$$\begin{aligned} &p(Y_{T+h}|Y_T, X_T, F_{T+h-1}, X_{T+h}) \propto [\underline{S} + (W_T - V_T\widehat{\Pi})'(W_T - V_T\widehat{\Pi}) + \\ &(Y_{T+h} - Z_{T+h}\widehat{\Pi})'(I - Z_{T+h}L^{-1}Z'_{T+h})(Y_{T+h} - Z_{T+h}\widehat{\Pi})]^{-(T+\nu+h)/2} \end{aligned} \tag{4A.11}$$

where $L = (Z'_{T+h}Z_{T+h} + Z'_T Z_T + \underline{V}^{-1})$, $Z_{T+h} = (I_h, F_{T+h-1}, X_T)$ with I_h a $(h \times h)$ identity matrix.

The predictive density of Y_{T+h} conditional on $(Y_T, X_T, F_{T+h-1}, X_{T+h})$ is thus a generalized t-distribution with location parameter $Z_{T+h}\widehat{\Pi}$, scale parameters $\underline{S} + (W_T - V_T\widehat{\Pi})'(W_T - V_T\widehat{\Pi})$ and $(I_N - Z_{T+h}L^{-1}Z'_{T+h})$, and $T + \nu$ degrees of freedom. Following Zellner (1971) we rewrite (4A.11) as:

$$p(Y_{T+h}|Y_T, X_T, F_{T+h-1}, X_{T+h}) = p(Y_{T+1}|F_T, X_{T+1}) \times \ldots \times p(Y_{T+h}|F_T, X_{T+1}, ..., F_{T+H-1}, X_{T+h}) \tag{4A.12}$$

F_{T+h-1} and X_{T+h} are generated from their predictive densities conditional on past values, independently from Y_{T+h}. Therefore, we substitute these densities in (4A.12) and we apply direct simulation to draw the predictive density of Y_{T+h}, conditional on Y_T and X_T,

$$p(Y_{T+h}|Y_T, X_T) = \iint p(Y_{T+h}|Y_T, X_T, F_{T+h-1}, X_{T+h})p(F_{T+h-1}|F_{T+h-2})p(X_{T+h}|X_{T+h-1})dF_{T+h-1}dX_{T+h} \tag{4A.13}$$

Note that we integrate with respect to the predictive density of the macroeconomic factors X_{T+h} given X_T.

4A.3 Nelson-Siegel model

Frequentist Inference

With the frequentist approach we estimate the Nelson-Siegel model using the two-step approach of Diebold and Li (2006) and the one-step approach of Diebold *et al.* (2006).

In the two-step approach we fix λ to 16.42, which, as shown in Diebold and Li (2006), maximizes the curvature factor loading at a 30-month maturity. For every month we then estimate the vector of β's by applying OLS on the cross-section of 18 maturities. From this first step we obtain time-series for the three factors, $\{\beta_t\}_{t=1}^T$. The second step consists of modelling the factors in (4.7) by fitting either separate AR(1) models or a single VAR(1) model.

In the one-step approach we estimate the unknown parameters and latent factors by means of the Kalman Filter using the prediction error decomposition for the State-Space model in (4.6)-(4.7). For each sample in the recursive estimation procedure, we first run the two-step approach with a VAR(1) specification for the state vector to obtain starting values. The unconditional mean and covariance matrix of $\{\beta_t\}_{t=1}^T$ are used to start the Kalman Filter. We discard the first 12 observations when evaluating the likelihood. All variance parameters of the diagonal matrix H and the full matrix Q are initialized to 1. The covariance terms in Q are initialized to 0. In the optimization procedure, we maximize the likelihood using the standard deviations as parameters to ensure positive estimates for the variance parameters. Finally, λ is initialized to 16.42.

Iterated forecasts for the factors are obtained as

$$\widehat{f}_{T+h} = \widehat{a} + \widehat{\Gamma}\widehat{f}_{T+h-1} \tag{4A.14}$$

where $\widehat{f}_{T+h} = (\widehat{\beta}_{1,T+h}, \widehat{\beta}_{2,T+h}, \widehat{\beta}_{3,T+h}, \widehat{M}_{T+h}, \widehat{M}_{T+h-1}, \widehat{M}_{T+h-2})$. These are then inserted in the measurement equation to compute interest rate forecasts:

$$\widehat{y}^{(\tau_i)}_{T+h} = \widehat{\beta}_{1,T+h} + \widehat{\beta}_{2,T+h}\left(\frac{1-\exp(-\tau_i/\widehat{\lambda})}{\tau_i/\widehat{\lambda}}\right) + \widehat{\beta}_{3,T+h}\left(\frac{1-\exp(-\tau_i/\widehat{\lambda})}{\tau_i/\widehat{\lambda}} - \exp(-\tau_i/\widehat{\lambda})\right) \tag{4A.15}$$

Bayesian Inference

The joint posterior densities for parameters of the Nelson-Siegel and affine models do not have a known closed-form expression. Therefore, we cannot analytically compute marginal densities for model parameters nor marginal predictive densities. We use Monte Carlo methods instead.

Prior Specification The model parameters are summarized by $\theta = (\lambda, \sigma^2, a, \Gamma, Q)$, where σ^2 is a (18×1) vector containing the diagonal elements of the measurement equation covariance matrix H. To facilitate the posterior simulation we use independent conjugate priors for the model parameters. For the variance parameters $\sigma^{(\tau_i)}$ we take the Inverted Gamma-2 prior

$$\sigma^{2^{(\tau_i)}} \sim \text{IG-2}(\underline{\nu}^{(\tau_i)}, \underline{\delta}^{(\tau_i)}) \tag{4A.16}$$

For the non-zero blocks in the state equation covariance matrix, Q_1 and Q_2, we assume Inverted Wishart distributions,

$$Q_1 \sim \text{IW}(\underline{\mu}_1, \underline{\Delta}_1) \tag{4A.17}$$

$$Q_2 \sim \mathrm{IW}(\underline{\mu}_2, \underline{\Delta}_2) \tag{4A.18}$$

For the linear regression parameters we assume a matricvariate Normal distribution,

$$[a, \Gamma] \sim MN(\underline{\Gamma}, Q \otimes \underline{V}_{\Gamma}) \tag{4A.19}$$

Finally for λ we assume a uniform distribution,

$$\lambda \sim U(\underline{a}_{\lambda}, \underline{b}_{\lambda}) \tag{4A.20}$$

We choose the parameters $\underline{a}_{\lambda}$ and $\underline{b}_{\lambda}$ to reflect the prior belief about the shape of the loading factors.

Posterior Simulation We obtain posterior results by using the Gibbs sampler of Geman and Geman (1984) with the data augmentation technique of Tanner and Wong (1987). The latent variables $B_T = \{\beta_{1,t}, \beta_{2,t}, \beta_{3,t}\}_{t=1}^{T}$ are simulated alongside the model parameters θ.

The complete data likelihood function is given by

$$p(Y_T, F_T|\theta) = \prod_{t=1}^{T}\prod_{i=1}^{18} p(y_t^{(\tau_i)}|f_t, \lambda, \sigma^{2(\tau_i)}) p(f_t|f_{t-1}, a, \Gamma, Q) \tag{4A.21}$$

where $Y_T = \{y_t^{(\tau_1)}, \ldots, y_t^{(\tau_N)}\}_{t=1}^{T}$ and where $F_T = \{\beta_{1,t}, \beta_{2,t}, \beta_{3,t}, M_t, M_{t-1}, M_{t-2}\}_{t=1}^{T}$. The terms $p(y_t^{(\tau_i)}|f_t, \lambda, \sigma^{2(\tau_i)})$, and $p(f_t|f_{t-1}, a, \Gamma, Q)$ are Normal density functions which follow directly from (4.6)–(4.7). When we combine (4A.21) with the prior densities $p(\theta)$ in (4A.16)–(4A.20) we obtain the posterior density

$$p(\theta, B_T|Y_T, M_T, M_{T-1}, M_{T-2}) \propto p(Y_T, F_T|\theta) p(\theta) \tag{4A.22}$$

We compute the full conditional posterior density for the latent regression parameters B_T using the simulation smoother as in Carter and Kohn (1994, Section 3) and we use the Kalman smoother to derive the conditional mean and variance of the latent factors. For the initial value β_0 we choose a multivariate normal prior with mean zero.

To sample the θ parameters (excluding λ), we use standard results. Hence, the variance parameters $\sigma^{(\tau_i)}$ are sampled from inverted Gamma-2 distributions, the matrix Q_1 is sampled from an Inverted Wishart distribution, and the parameters (a_1, Γ_1) are sampled from matricvariate Normal distributions, where (a_1, Γ_1) are the non-zero blocks of a and Γ respectively. In our framework the macro variables have a VAR(3) structure independent from the latent factors. Therefore, we simulate a_2, Γ_2, and Q_2 from their marginal densities, respectively generalized t-distributions and an Inverted Wishart distribution to improve the speed of convergence.

Finally, the posterior density for λ, conditional on (Y_T, F_T, H) is:

$$p(\lambda|Y_T, F_T, H) \propto \prod_{t=1}^{T}\prod_{i=1}^{N} p(y_t^{(\tau_i)}|f_t, \lambda, \sigma^{2(\tau_i)}) p(\lambda) \tag{4A.23}$$

Equation (4A.23) is not proportional to any known density. Therefore, λ has to be drawn by applying MCMC methods. We use the Griddy Gibbs algorithm. The Griddy Gibbs sampler was developed by Ritter and Tanner (1992) and is based on the idea to construct a simple approximation of the inverse cumulative distribution function of the target density on a grid of points[29]. More formally and referring to equation (4A.23), we perform the following steps:

- We evaluate $p(\lambda|Y_T, F_T, H)$ at points $V_i = v_1, ..., v_n$ to obtain $w_1, ..., w_n$;
- We use $w_1, ..., w_n$ to obtain an approximation to the inverse cdf of $p(\lambda|Y_T, F_T, H)$;
- We sample a uniform (0,1) deviate and we transform the observation via the approximate inverse cumulative density function.

[29]Mönch (2006a) applies a random walk Metropolis Hastings algorithm to draw λ. We choose the Griddy-Gibbs since the space of λ is well defined and only the cumulative density function needs to be estimated in these grid points.

Forecasting The h-step ahead predictive density of Y_{T+h}, conditional on Y_T and F_T, is given by

$$p(Y_{T+h}|Y_T, F_T) = \iint p(y_{T+h}^{(\tau_i)}|f_{T+h}, \lambda, \sigma^{2(\tau_i)})p(f_{T+h}|f_{T+h-1}, a, \Gamma, Q)\times$$
$$p(\theta, B_T|Y_T, M_T, M_{T-1}, M_{T-2})df_{T+h}\, d\theta \quad (4A.24)$$

where $p(y_{T+h}^{(\tau_i)}|f_{T+h}, \lambda, \sigma^{2(\tau_i)})$ and $p(f_{T+h}|f_{T+h-1}, a, \Gamma, Q)$ follow directly from the state space system and where $p(\theta, B_T|Y_T, M_T, M_{T-1}, M_{T-2})$ is the posterior density.

Simulating Y_{T+h} from the h-step ahead distribution (4A.24) is straightforward. In each step of the Gibbs sampler, we use the simulated values of (a, Γ, Q) to draw the out-of-sample values of f_{T+h}. Then, f_{T+h}, in combination with the current Gibbs draws of H and λ, provides a simulated value for $y_{T+h}^{(\tau_i)}$.

4A.4 Affine model

Frequentist Inference

To estimate the affine model we assume that yields of every maturity are contaminated with measurement error. We estimate the parameters in the resulting State-Space model by applying the two-step approach used in Ang *et al.* (2006b). We make the latent factors Z_t observable by extracting the first three principal components from the panel of yields of different maturities. The first step of the estimation procedure consists of estimating the equation and variance parameters of the state equations (4.23). In the second step we estimate the remaining parameters $(\delta_0, \delta_1, \lambda_0, \lambda_1)$. We first estimate (δ_0, δ_1) by applying OLS to the short rate equation (4.13) where we use the 1-month yield as the observable short rate. We then estimate the risk premia parameters (λ_0, λ_1) by minimizing the sum of squared yields errors in the measurement equations (4.22), giving the parameter estimates from the first step, $(\widehat{\mu}, \widehat{\Psi}, \widehat{\Sigma})$ and the short rate parameters $(\widehat{\delta}_0, \widehat{\delta}_1)$. In the second step we initialize all risk premia parameters to zero. Common approaches for obtaining starting values for the risk premia parameters by first estimating either λ_0 or λ_1 in a separate step yielded unsatisfactory results.

Yield forecasts are generated by forward iteration of the state equations

$$\widehat{f}_{T+h} = \widehat{\mu} + \widehat{\Psi}\widehat{f}_{T+h-1} \quad (4A.25)$$

where $\widehat{f}_{T+h} = (\widehat{Z}_{1,T+h}, \widehat{M}_{T+h-1}, \widehat{M}_{T+h-2})$. With the estimated parameters substituted in $a^{(\tau_i)}$ and $b^{(\tau_i)}$ we then construct interest rate forecasts as

$$\widehat{y}_{T+h}^{(\tau_i)} = \widehat{a}^{(\tau_i)} + \widehat{b}^{(\tau_i)}\widehat{f}_{T+h} \quad (4A.26)$$

Bayesian Inference

Bayesian inference on model (22)-(23) is very complex due to the high degree of nonlinearity and, above all, the large set of yields we model. It is particularly difficult to define the space of the short rate parameters. The likelihood is very sensitive to these parameters and small perturbations give very different and unrealistic results. Therefore, an estimation approach similar to Ang *et al.* (2006a) may not be the optimal solution. We opt for a normal approximation of the full posterior density around frequentist parameter estimates. This choice implies that the predictive density of Y_{T+h}, conditional on Y_T and $F_t = \{f_t\}_{t=1}^T$, can be derived without having to compute posterior densities.

Forecasting The h-step ahead predictive density of $y_{T+h}^{(\tau_i)}$, conditional on Y_T and F_T, is given by

$$p(y_{T+h}^{(\tau_i)}|Y_T, F_T) = \iint p(y_{T+h}^{(\tau_i)}|f_{T+h}, a^{(\tau_i)}, b^{(\tau_i)}, \sigma^{2(\tau_i)})p(f_{T+h}|f_{T+h-1}, \mu, \Psi, Q)p(\theta|Y_T, F_T)df_{T+h}\, d\theta \quad (4A.27)$$

where $p(y_{T+h}^{(\tau_i)}|x_{T+h}, a^{(\tau_i)}, b^{(\tau_i)}, \sigma^{2(\tau_i)})$, and $p(f_{T+h}|f_{T+h-1}, \mu, \Psi, Q)$ are the conditional predictive densities and where $p(\theta|Y_T, X_T)$ is the posterior density for the parameter vector $\theta = (\mu, \Psi, H, Q, a, b, \lambda_0, \lambda_1)$.

As we discussed in the previous paragraph we approximate $p(\theta|Y_T, X_T)$ in (4A.27), with a normal distribution around frequentist estimates: $q(\widehat{\theta}|Y_T, X_T)$. Since f_{T+h} can be drawn independently of Y_{T+h}, we use direct simulation to compute the predictive density of Y_{T+h} conditional on (Y_T, F_T):

$$p(y_{T+h}^{(\tau_i)}|Y_T, X_T) = \iint p(y_{T+h}^{(\tau_i)}|f_{T+h}, \widehat{a}^{(\tau_i)}, \widehat{b}^{(\tau_i)}, \widehat{\sigma^2}^{(\tau_i)})p(f_{T+h}|f_{T+h-1}, \widehat{\mu}, \widehat{\Psi}, \widehat{Q})df_{T+h}d\theta \qquad (4A.28)$$

4B Bayesian Model Averaging

We denote the predictive density of $y_{T+h}^{(\tau_i)}$, given M individual models and conditional on the time T information set, by D_T. This density is given by

$$p(y_{T+h}^{(\tau_i)}|Y_T, D_T) = \sum_{i=1}^{M} P(m_j^{(\tau_i)}|Y_T, D_T)p(y_{T+h}^{(\tau_i)}|Y_T, D_T, m_j^{(\tau_i)}) \qquad (4B.1)$$

for $j = 1, ..., M$ and where $P(m_j^{(\tau_i)}|Y_T, D_T)$ is the posterior probability of model m_j for maturity τ_i, conditional on data at time T, and where $p(y_{T+h}^{(\tau_i)}|Y_T, D_T, m_j^{(\tau_i)})$ is the model m_j predictive density of $y_{T+h}^{(\tau_i)}$, conditional on Y_T and D_T. The posterior probability of model m_j for maturity τ_i is computed as:

$$P(m_j^{(\tau_i)}|Y_T, D_T) = \frac{p(y_{T,o}^{(\tau_i)}|Y_T, D_T, m_j^{(\tau_i)})P(m_j^{(\tau_i)})}{\sum_{s=1}^{k} p(y_{T,o}^{(\tau_i)}|Y_T, D_T, m_s^{(\tau_i)})P(m_s^{(\tau_i)})} \qquad (4B.2)$$

where $P(m_j^{(\tau_i)})$ is the prior probability of model m_j for maturity τ_i. The predictive likelihood value for model m_j, $p(y_{T,o}^{\tau_i}|Y_T, D_T, m_j^{(\tau_i)})$, is computed by substituting the realized value $y_{T,o}^{(\tau_i)}$ in the predictive density $p(y_T^{(\tau_i)}|Y_T, D_T, m_j^{(\tau_i)})$. We average individual models independently for every maturity.

4C Prior specification

In the literature uninformative priors or diffuse informative priors are often chosen to derive posterior densities that depend only on data information (the likelihood). We do not follow this approach as we apply informative priors in our estimation and forecasting procedures. There are several motivations to do so. Firstly, for nonlinear models such as the Nelson-Siegel and affine models, it is very difficult to determine when a prior is non-informative. Secondly, the simulation algorithm might get stuck in some (nonsensical) regions of the parameter space and it may require a substantial number of simulations to converge, thereby enormously increasing estimation time. Thirdly, we believe that market agents will to some degree always have prior information which can be partially incorporated in our models when forecasting interest rates. Finally, we want to study and underline differences between frequentist and Bayesian inference in forecasting yields, and the use of priors is one, if not the main difference between the two approaches.

We briefly discuss the specification of the prior densities for the parameters of the models presented in Appendix A. We start with the AR model and the Normal-Gamma conjugate prior in (4A.2) for parameters $(c^{(\tau_i)}, \phi^{(\tau_i)}, \psi^{(\tau_i)}, \sigma^{(\tau_i)})$. We choose $\underline{v} = 0.01$ to have a prior density for the vector $(c^{(\tau_i)}, \phi^{(\tau_i)}, \psi^{(\tau_i)})$ concentrated around the mean value. We choose the mean vector value $\underline{b}$ by calibrating it to the initial in-sample data (1970:1-1993:12) and to prevent unit root type behavior. The prior for $\sigma^{(\tau_i)}$ is less informative with $\underline{\nu}$ fixed to 20 and $\underline{s}^2$ again calibrated to in-sample data.

The calibration of the prior for the VAR model is more complex due to the high dimensionality of Π and S. Therefore, we relax our prior assumption and we choose a wider region for $\underline{V}$ and $\underline{\nu}$ in (4A.4)–(4A.5). $\underline{B}$ is again calibrated with initial in-sample data and the resulting values imply plausible factor loadings of the yield factors.

The order of prior information in the Nelson-Siegel model is comparable to the VAR model. For the parameter λ we choose the following prior density:

$$\lambda \sim U(3.34, 33.45) \qquad (4C.1)$$

By restricting λ in the interval [3.34, 33.45] we can make sure that the loading on the curvature factor $\beta_{3,t}$ is at its maximum for a maturity between 6 months and 5 years.

For the ATSM model we do not apply prior densities. We use a normal approximation of the conditional predictive density around maximum likelihood parameter estimates. Indeed, because we consider a large number of maturities, which results in a substantial number of parameters that need to be estimated, the speed of convergence of MCMC algorithms such as the Gibbs sampler is very slow. Moreover, some parameters do not converge at all, and unrealistic values are simulated. However, we believe the approximation is satisfactory. We therefore still account for parameter uncertainty in the affine models.

Finally, in Bayesian model averaging we apply the same uninformative prior probability to each model, $P(m_j^{(\tau_i)}) = 1/M$.

Chapter 5

The power of weather

Some empirical evidence on predicting day-ahead power prices through weather forecasts

5.1 Introduction

The heatwave in Europe during August 2003 (the warmest summer in Europe since 1500), resulted in extremely high prices in several power markets, as France, Germany and the Netherlands, see for example Figure 5.1. The fact that not the 15000 casualties due to the heatwave but the technical problems of electricity supply experienced by Électricité de France (EDF), the main power supplier in France, were on the top of the agenda of the French Cabinet meeting held on August 11, 2003, illustrates the tremendous importance of the functioning of the power system to our society.

A decade ago, the electricity industry was vertically integrated, prices were regulated and reflected the short-term production costs. Hence, back then the electricity price did not reflect the temporal effects such as seasonality, weather and business activity. But this all changed when many governments worldwide, started reforming their electricity industry as of the middle ninety-nineties. Currently, the economic law of demand and supply determines the price on market places where electricity can be traded on spot or forward (i.e. hour-ahead, day-ahead, or month-ahead).

Many studies have documented these stylized facts from examining the prices observed at day-ahead markets[1], which are by far the most liquid power wholesale markets, see Escribano *et al.* (2002), Lucia and Schwartz (2002) and Koopman *et al.* (2007). Bunn and Karakatsani (2003), provide a thorough review of the stochastic price models presented in these studies and classify these into three groups, being random walk models, basic mean-reversion models, and extended mean-reversion models that incorporate time-varying parameters (to control for seasonality and volatility patterns). They conclude that

[1] On these markets, hourly prices are quoted for delivery of electricity on certain hours on the next day.

the idiosyncratic price structure has not been accurately described. Furthermore, the results reported in these studies are often obtained from in-sample tests, hence they do not resolve the issue of the out-of-sample predictive value of power models.

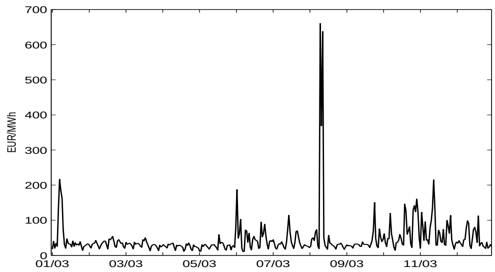
Only a few studies have recognized the need for modelling weather directly, however, mainly addressed the weather effect on electricity sales (see for example the special issue of Journal of Econometrics 1979). Moral-Carcedo and Vicns-Otero (2005) also study temperature effects on the variability of daily electricity demand in Spain, and document empirical evidence of a non-linear relationship between variations in temperature and the demand response. Knittel and Roberts (2005) test stochastic price models on hourly-ahead electricity prices obtain from the Californian market. They compare price models that incorporate seasonal and temperature variables with models that do not include these variables, and provide preliminary evidence that first-mentioned models significantly outperform in terms of forecasting accuracy.

In this study we shed light on forecasting performance of stochastic price models in day-ahead electricity markets, by studying the impact of weather variables to anticipate future electricity price movements. We examine six stochastic price models to forecast day-ahead prices of two of the most active power exchanges in the world: the Nordic Power Exchange and the Amsterdam Power Exchange[2]. Three of these forecasting models extend Knittel and Roberts (2005) by including weather variables (temperature, precipitation and wind speed) as explanatory variables for variation in day-ahead electricity prices. To our best knowledge, we are the first to use next-day weather forecasts in our analysis rather than real weather data. This is consistent with the market microstructure of day-ahead markets, which is such that agents submit their bids and offers for delivery of electricity in all hours of the next day (for a certain market closure time). Furthermore, we implement specific models for different power markets, due to the heterogeneity in weather conditions and production plants.

From these six tested models, an extended ARMA model (that includes power transformations) of next-day weather forecasts yields the best forecasting results for predicting day-ahead prices. Also, this model has substantial explanatory power to anticipate on ex-ante price jumps. As adverse climate conditions tend to lead to sharp increases in demand, hereby causing supply shortages in electricity, we carefully explore the causal relationship between weather and prices. The reported evidence indicates that the next-day weather forecasts influence day-ahead prices, both via the demand and supply side. In particular, when production is less related to weather as in the case of the Dutch market (opposed to the Scandinavian hydro market), weather forecasts have a negligible influence

[2]See for example Geman (2005).

Figure 5.1: APX 2003 Prices



Note: The figure presents the daily electricity prices in the APX market over 2003.

on day-ahead prices. Finally, the empirical results show that a GARCH model specification including weather variables provides accurate price forecasts, which contradicts with earlier preliminary evidence that "standard" GARCH models would predict electricity poorly (Knittel and Roberts, 2005).

The remainder of this Chapter is structured as follows. Section 5.2 introduces the day-ahead power markets. Section 5.3 presents the data. Section 5.4 describes the forecasting models. Section 5.5 discusses the empirical results. Section 5.6 concludes.

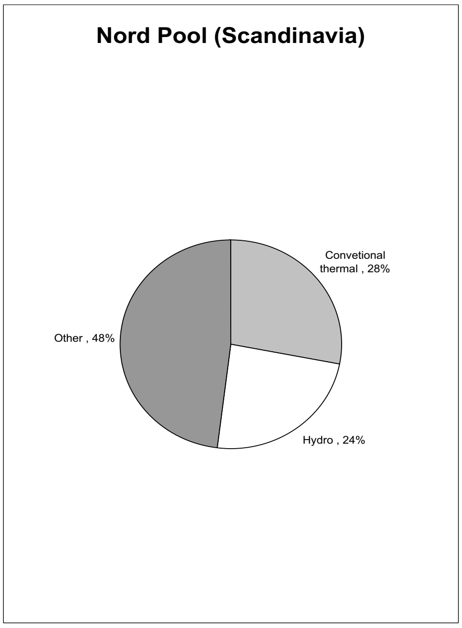
5.2 Day-ahead power markets

On 1 January 1991, the Norwegian government imposed a deregulation process on its electricity industry that resulted in the establishment of the first national power market for short-term delivery of power (real-time and day-ahead[3]) in the world, the Nordic Power Exchange (NPX). Two years later, in 1993, the range of products was extended with forward and futures contracts that have longer maturity horizons. Another few years later, Sweden joined the NPX (1996), soon followed by Finland (1998), West-Denmark (1999) and East-Denmark (2000). From 2003 all customers of Scandinavian electricity markets may trade freely in the market. The NPX, now also named Nord Pool ASA, is considered as the most liquid wholesale market worldwide. Nord Pool ASA constitutes of a day-ahead market (Elspot), a financial market (Elbas), and a clearing service. In the remainder, we mainly focus on the Elspot market. For more details on Nord Pool ASA we refer to NordPool (2004).

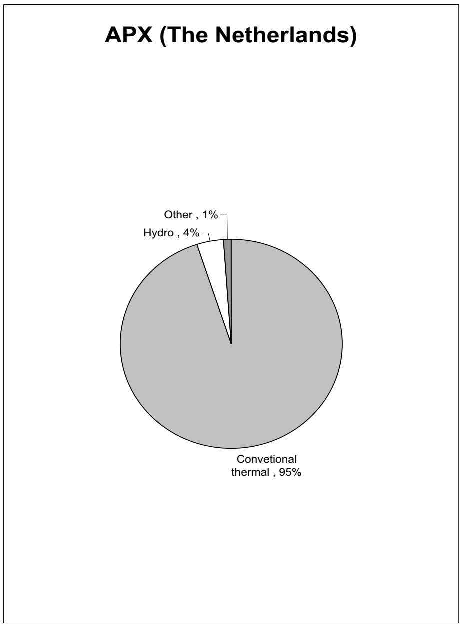
Another country that liberalized its electricity markets from an early stage onwards is The Netherlands. In 1999, the second electronic power exchange in Europe for day-ahead electricity contracts was founded here, being the Amsterdam Power Exchange (APX). The

[3]We remember from section 5.1 that day-ahead means that prices are quoted at day t for delivery of electricity on certain hours on the day $t + 1$.

Figure 5.2: Producer plants



(a)



(b)

Note: The figure presents capacity figure of EU countries with most active wholesale power markets.
Conventional thermal fuelled capacity: oil, gas, coal.
Hydro fuelled capacity: reservoir, river.
Other fuelled capacity: wind, solar.

APX is composed out of a day-ahead market and a financial market. For more details on APX we refer to www.apxgroup.com.

In Figure 5.2 some descriptive statistics of these two markets are listed. The Nord Pool market is largely dependent on electricity that is generated by renewable sources. In particular, hydro-plants, which use water stored in reservoirs or lakes, are dominant in Norway and partly Sweden; wind-plants, which use wind to produce electricity, are dominant in Denmark. In the APX market oil, coal, gas or a combination of these fuels is used to generate electricity.

Electricity prices are affected by regional and temporal influences due to the transportation and transmission limits of electricity. This statement is particular important in the Nord Pool market. For instance, when a power plant falls out in the eastern part of Sweden this only affects the power supply in the surrounding region. Hence, this will not affect power supply in the western part of Sweden and the rest of the market. Similarly, rainfall in the southern part of Norway, will potentially affect the regional demand and/or supply curve, but not the bidding curves in other regions. Nord Pool faces the problem by allowing to split the market in several bidding and prices areas. Therefore, we take into account the Nord Pool bidding area prices separately, rather than examining the Elspot system price (which is a weighted average of the bidding prices in all Nord Pool bidding areas). We examine two out of the eleven bidding areas in the Nord Pool, being the Oslo area and Eastern Denmark area. It is interesting to note that these areas are the most densely populated areas in Scandinavia.

5.3 Data

5.3.1 Electricity prices

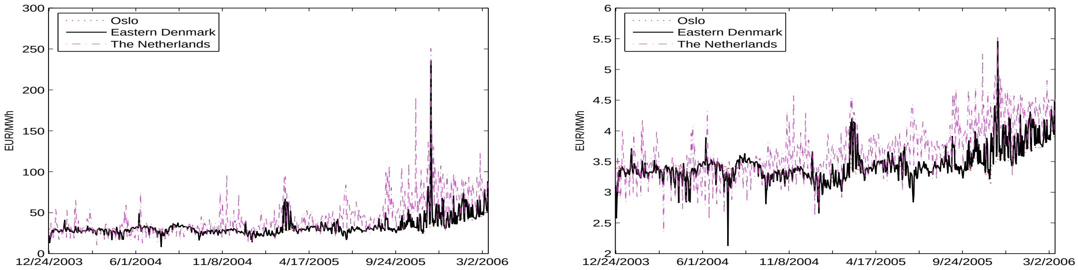
The data set used in this study consists of day-ahead prices in EUR/MWh for Oslo, Eastern Denmark and the Netherlands from the period December 24, 2003 to March 14, 2006. Oslo and Eastern Denmark are two bidding areas of Nord Pool market; Dutch electricity prices are obtained from APX market[4]. Nord Pool provides bidding area prices both in the local currency and in EUR. We choose EUR to compare directly to APX prices. Daily prices are computed as the arithmetic mean of the available 24 hourly prices series on the physical market of each country.

As in Wilkinson and Winsen (2002) and Lucia and Schwartz (2002) we start from a statistical analysis of the data we have[5]. Figure 5.3 plots the time series, the log transformations and the histograms of the daily day-ahead electricity prices; Table 5.1 gives some important descriptive statistics. A first casual look discloses an erratic behavior of the prices. The series follow a small positive increasing trend with several spikes. Interestingly, Oslo prices in Oslo have more negative spikes than positive spikes, which is not surprising when one notices that this market is characterized by hydro produced (hence, considered as 'storable') electricity. Thermal-derived electricity (meaning electricity produced from gas, coal and or nuclear fuel plants) on the other hand is non-storable, and you would expect a higher frequency of occurrence in these markets. Also, negative spikes in thermal markets typically only occur in night hours when electricity is offered at discount prices to avoid costs for ramping up and down later (Bunn and Karakatsani, 2003). As this argument does not explain the negative price spikes in hydro markets, we may conclude that supply tends to exceed demand throughout the day in Oslo, and demand is not forecasted accurately by suppliers. The price levels observed for Eastern Denmark

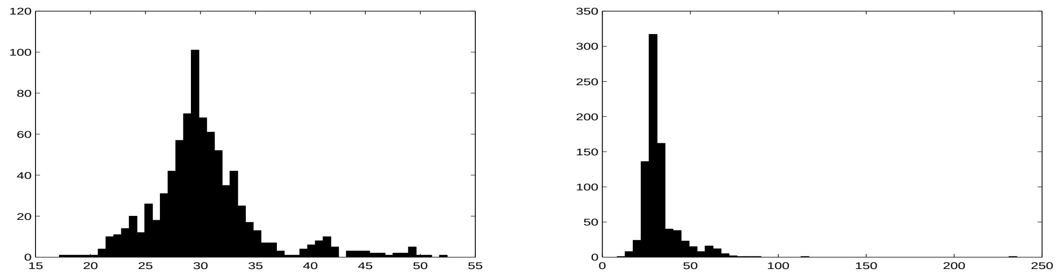
[4]Electricity prices may be available for a longer sample, but weather forecasts are available to us only for this sample.

[5]We briefly discuss some stylized facts; we refer for a more detailed analysis, for example, to Lucia and Schwartz (2002) and Pilipovic (1997).

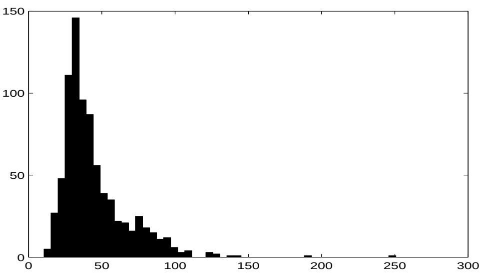
Figure 5.3: Prices



(a)



(b)



(c)

Note: The graphs in this figure present in Panel a) prices (in the left panel) and log prices (in the right panel) of daily electricity prices in Oslo, Eastern Denmark and the Netherlands; in Panel b) histograms of daily electricity prices in Oslo (in the left panel) and Eastern Denmark (in the right panel) markets; and in Panel c) histograms of daily electricity prices in the Netherlands market.

Table 5.1: Descriptive statistics

	Oslo	Eastern Denmark	The Netherlands
Mean	30.431	32.852	44.741
St dev	4.9647	12.735	22.910
Min	17.162	8.3896	10.519
Max	52.450	235.71	250.69
Skewness	1.3188	6.2948	2.3118
Kurtosis	0.3058	79.129	7.8645
Working days	30.904	34.805	49.684
No working days	29.354	28.404	33.282
ρ_1	0.913	0.766	0.754
ρ_7	0.736	0.629	0.774
ρ_{14}	0.544	0.490	0.700

Note: The table reports descriptive statistics on electricity prices in Oslo, Eastern Denmark and the Netherlands. Lines working days and no working days give the sample average prices on working days and no working days (weekends and holidays) respectively. Lines ρ_1, ρ_7 and ρ_{14} give the 1^{st}, 7^{th} and 14^{th} sample autocorrelation.

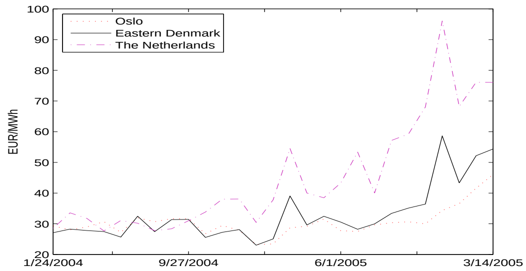
and The Netherlands are substantially higher. The histograms provide similar evidence. Eastern Denmark and the Netherlands prices are highly non-normally distributed; their volatility is very high such as the kurtosis; their skewness is positive. Oslo has a more regular distribution, but a Jarque-Bera test rejects the null hypothesis of normality for each of the three series. The series are characterized by a weekly pattern. From table 1 we can observe that prices are lower during the weekend than on business days. Yearly patterns, well documented in other studies, are more difficult to notice since the series are not very long. Still, differences among seasons can be seen in Figure 5.4. Electricity prices are very persistent and possible close to non-stationary. Table 5.1 shows that the sample autocorrelations are high up to 14-day lags. From Figure 5.3, we can observe another stylized fact, being volatility clustering. Dramatic spikes tend to occur in clusters, mainly as result of consecutively exceeding the system capacity.

In our application we use log prices and not the level. The log transformation reduces the spike behavior of the prices and makes moments of the distribution of electricity prices more similar to standard distributions, in particular for Eastern Denmark and the Netherlands log prices.

5.3.2 Weather forecasts

We continue the data analysis by focusing on weather forecasts. Forecasts on the daily average temperature in degrees Celsius, total precipitation in mm, and wind power in

Figure 5.4: Monthly average prices



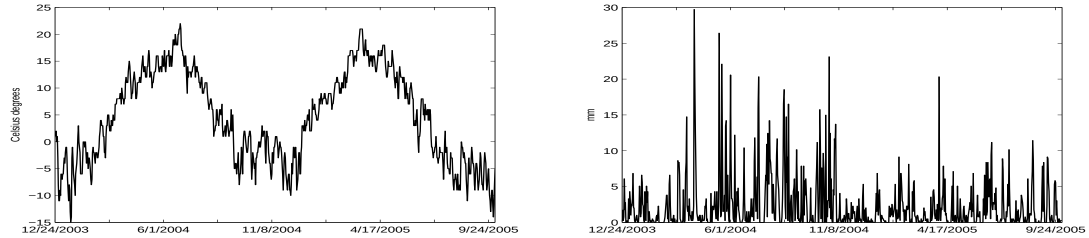
Note: The figure presents the monthly average electricity prices in Oslo, Eastern Denmark and the Netherlands.

m/s are applied. Data is obtained from the EHAMFORE index, which is provided by Meteorlogix (www.meteorlogix.com)[6]. We assume that market operators use the weather forecasts provided by Meteorlogix in their decisions. We think that this assumption is quite realistic considering the market share of Meteorlogix in providing real-time information services in the agriculture, energy, and commodity trading markets, and Bloomberg in providing data to operators. Weather forecasts refer to a square area around the measurement station, , hence our analysis does not cover the entire country of the markets under consideration. Also, the combination of different stations could be applied. However, as data from minor cities is scarce, the weather forecast errors may arise introducing further noise in the forecasting process. Finally, the areas that we study are small and homogenous in term of weather. Therefore, we only use weather forecasts for Oslo, Copenhagen and Amsterdam. The weather around Oslo may well approximate the weather in the area on the south of Oslo along the sea cost where most of the electricity for south-east Norway is produced. The weather in the area of Copenhagen may be a proxy for the weather of Zealand, the main island in Eastern Denmark. Finally, Amsterdam is located in the middle of the Netherlands.

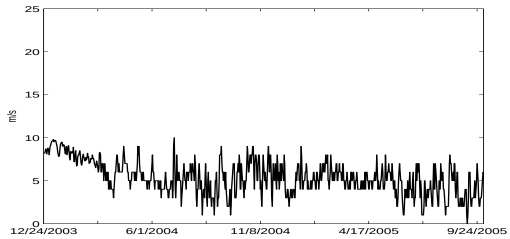
Figures 5.5-5.7 plot the three variables for each country. Temperatures have highly seasonal patterns, with lower values for Oslo and higher for the Netherlands. Precipitation figures are low in Eastern Denmark compared to Oslo and The Netherlands. The level of wind is particular high in Eastern Denmark and The Netherlands. The wind forecasts on all the three countries have a quite stable pattern in the initial months of 2004, because the meteorologic institute applies a different forecasting model on those months. We decide to keep these forecasts to extend the sample period as much as we can. Meanwhile market operators got these weather forecasts as information to make decisions at that

[6]Data from the EHAMFORE index are available in Bloomberg.

Figure 5.5: Weather variables: Oslo



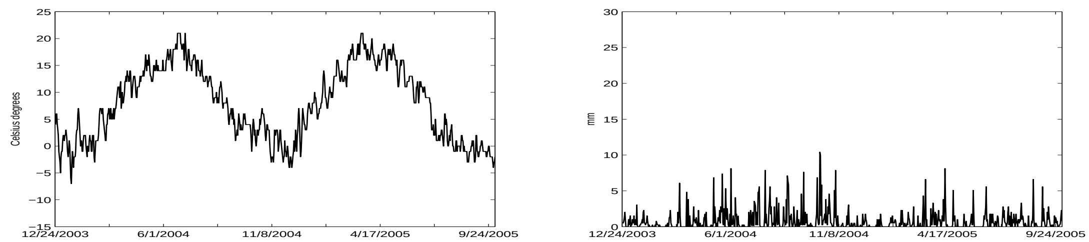
(a)



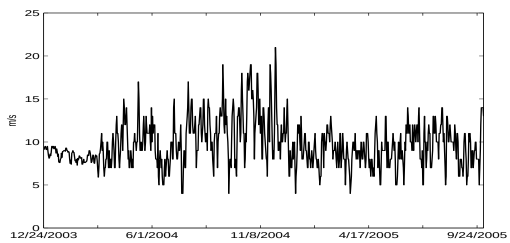
(b)

Note: See note in figure 5.6.

Figure 5.6: Weather variables: Eastern Denmark



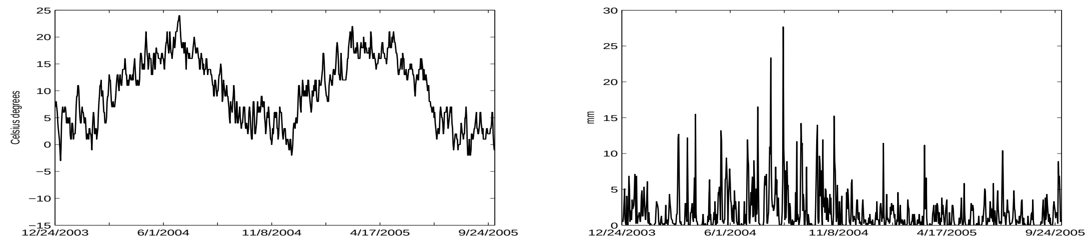
(a)



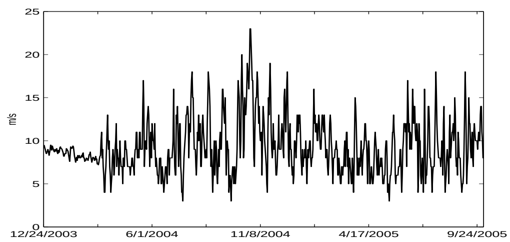
(b)

Note: The graphs in figures 5.5 and 5.6 present in Panel a) the forecasts on the daily average temperature (in the left panel), and total precipitation (in the right panel); in Panel b) the forecasts on wind speed respectively in Oslo and Copenhagen.

Figure 5.7: Weather variables: The Netherlands



(a)



(b)

Note: The graphs in this figure present in Panel a) the forecasts on the daily average temperature (in the left panel), and total precipitation (in the right panel); in Panel b) the forecasts on wind speed in Amsterdam.

moment. Some graphical relations between the forecasted weather variables and electricity prices may be identified. For example, high precipitation in Oslo at the end of May 2004 or in October 2004 corresponds to low prices; few days of very low temperature in Oslo in February 2005 correspond to high prices; strong wind in Eastern Denmark at the end of 2004 and beginning of 2006 is associated with low prices. However, even if the real weather was the weather forecasts, a graphical analysis would not be satisfactory because the relation between weather variables and electricity prices is possibly highly nonlinear as we will discuss in Section 5.5.1. Therefore, we try to find specific models to interpret the weather influences.

5.4 Forecasting models

Knittel and Roberts (2005) shows that traditional time series approaches such as ARMA models provide more accurate results in forecasting electricity prices than their continuous counterparts. Starting from these findings we built several models that may cope with the stylized facts of electricity prices.

5.4.1 Model 1: ARMA

The first model is a traditional time series approach to model electricity prices, the autoregressive moving average (ARMA) model (Hamilton (1994)). The ARMA(p, q) model implies that the current value of the investigated process (say, the log price) P_t is expressed linearly in terms of its past p values (autoregressive part) and in terms of the q previous values of the process ϵ_t (moving average part):

$$\phi(L)P_t = \theta(L)\epsilon_t \tag{5.1}$$

where $\phi(L)$ and $\theta(L)$ are the autoregressive and moving average polynomials in the lag operator L respectively, defined as:

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p \tag{5.2}$$

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q \tag{5.3}$$

and where ϵ_t is an independent and identically distributed (iid) noise process with zero mean and finite variance σ. The motivation of an ARMA process follows from the correlogram. Table 5.1 shows high correlation between the current price and the previous days' prices.

The ARMA modelling approaches assume that the time series under study is (weakly) stationary. If it is not, a transformation of the series to stationarity is necessary, such as first differentiating. The resulting model is known as the autoregressive integrated moving-average model (ARIMA). We do not work with first difference prices for several reasons. Firstly, the Dickey Fuller test on the series rejects the null hypothesis of non-stationary. Secondly, as our primary objective of this study is to accurately model and predict electricity prices we argue that a first-difference transformation of the price could eliminate important stylized facts such as price trends. Thirdly, the empirical evidence provided in literature is in favour of level of prices. For example, Lucia and Schwartz (2002) find that models based on levels and log levels provide more accurate results than models based on first differences and log first differences in forecasting Nord Pool electricity prices. And Weron and Misiorek (2005) find that in terms of out-of-sample statistics, ARMA models do better than ARIMA models in forecasting electricity prices.

5.4.2 Model 2: ARMAX

The second model is an extension of model 1. ARMA models apply information related to the past of the process and do not use information contained in other pertinent time series. However, as the data analysis shows, electricity prices are generally governed by various fundamental factors, such as seasonality and load profiles. The ARMAX(p, q) can be written as:

$$\phi(L)(P_t - X_t) = \theta(L)\epsilon_t \tag{5.4}$$

where $X_t = \sum_{i=1}^{k} \psi_i x_{i,t}$, where $x_t = (x_1, x_2, ...x_k)'$ is the $(k \times 1)$ vector of explanatory variables at time t, and where $\psi = (\psi_1, \psi_2, ..., \psi_k)'$ is a $(k \times 1)$ vector of coefficients. Following Lucia and Schwartz (2002) we use three explanatory variables: a dummy with values 0 on working days and 1 on holidays, a seasonal dummy given by the combination of the two variables $sin(2\pi t/365.25)$ and $cos(2\pi t/365.25)$. These dummy variables may be interpreted as proxies for load profiles (higher demand on working days), and proxies for weather effect (higher demand on cold and warm seasons).

In the empirical application, the ARMAX model will be our benchmark.

5.4.3 Model 3: ARMAXW

Averse weather conditions may change the demand for electricity, and may also affect the production. Low level of precipitation and/ or wind speed may cause reduction on the supply of energy, in particular in electricity markets which depend on renewable producer plants, such as Norway and Denmark. Furthermore, producer plants may study future weather conditions to estimate demand and plan their supply optimally.

The third model is an extension of model (5.4) and is built following the previous reasoning. Forecasts on the average daily temperature in degrees Celsius, precipitation in mm and wind speed in m/s are applied as further explanatory variables. The model is:

$$\phi(L)(P_t - X_t - W_t) = \theta(L)\epsilon_t \tag{5.5}$$

where $W_t = \sum_{j=1}^{l} \varphi_j w_{j,t}$, where $w_t = (w_{1,t}, w_{2,t}, ..., w_{l,t})'$ is the $(l \times 1)$ vector of weather forecast variables at time t, and where $\varphi = (\varphi_1, \varphi_2, ..., \varphi_l)'$ is a $(l \times 1)$ vector of coefficients. This model includes deterministic components that account for genuine regularities in the behavior of electricity prices and stochastic components that come from weather shocks.

Knittel and Roberts (2005) apply a similar model for forecasting California electricity prices, where the set of weather variables is composed by the level, the square and the cubic of realized temperature. But the price of tomorrow depends on the weather of tomorrow and not on the weather of today[7]. Therefore, we use weather forecasts and not realized values. Moreover, we add a variable to measure wind speed since wind may play a role both in the feeling of the people - people feel colder with stronger wind - and in the

[7] We find the same evidence in an unreported exercise on forecasting Oslo electricity prices. Results are available upon request.

supply of wind power plants. We also introduce an explanatory variable to model another dimension of the weather, being precipitation, allowing us to approximate the supply in hydro dominated plants. As in Knittel and Roberts (2005) we allow nonlinearity in the relation between prices and weather variables by including the level, the square and the cubic of the temperature forecasts, and the level and the square of the precipitation and wind forecasts[8].

For hydro markets it may be a different story: the water reservoir is often more important to plan production than the amount of precipitation, see e.g. Koopman *et al.* (2007) and Deng (2004). We emphasize that in this Chapter we work with local prices, and local water reservoirs are in most cases not observable. This is particularly true when the number of electricity producers is high, which applies for Oslo. Furthermore, we think that hydroelectric plants incorporate forecasted future precipitations in their strategic decisions of the amount of water to store.

5.4.4 Model 4: ARMAX-GARCH

ARMA models assume homoscedasticity, i.e. constant variance and covariance function, but the preliminary data analysis has disclosed that electricity prices exhibit volatility clustering. The fourth model extends model 2 by assuming a time varying conditional variance of the noise term. The heteroskedasticity is modelled by a generalized autoregressive conditional heteroskedastic GARCH(r, s) model (Bollerslev (1986)). Relaxing the assumption of homoscedasticity may change the parameter estimates of model 2, and consequently the out-of-sample forecast of the investigated process.

The model is:

$$\phi(L)(P_t - X_t) = \theta(L)\epsilon_t \tag{5.6}$$

$$\epsilon_t = \nu_t h_t^{1/2} \quad with \;\; h_t = \alpha_0 + \sum_{i=1}^{s} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{r} \beta_j h_{t-j} \tag{5.7}$$

where ϵ_t is an independent and identically distributed (iid) noise process with zero mean and conditional time varying variance h_t, and the coefficients have to satisfy $\alpha_i \geq 0$ for $1 \leq i \leq s$, $\beta_j \geq 0$ for $1 \leq j \leq r$, and $\alpha_0 > 0$ to ensure that the conditional variance is strictly positive.

5.4.5 Model 5: ARMAXW-GARCH

Following the same reasoning for model (5.6)-(5.7), model 3 can be extended by assuming a noise process with a time varying conditional variance.

[8] Precipitation and wind forecasts are always positive, therefore we do not consider useful to include the cubic transformation.

Model 5 is:

$$\phi(L)(P_t - X_t - W_t) = \theta(L)\epsilon_t \tag{5.8}$$

$$\epsilon_t = \nu_t h_t^{1/2} \quad with \;\; h_t = \alpha_0 + \sum_{i=1}^{s} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{r} \beta_j h_{t-j} \tag{5.9}$$

5.4.6 Model 6: ARMAXW-GARCHW

Koopman *et al.* (2007) find that seasonal factors and other fixed effects in the variance equation are also important to estimate electricity prices. The fifth model extends model 4 by reformulating model 3 and 4 to incorporate Koopman *et al.* (2007) results. The conditional variance of the noise term in model 3 is assumed to be time-varying and modelled with a GARCH expression where some explanatory variables are added to the ARMA form of equation (5.7). The model is specified as:

$$\phi(L)(P_t - X_t - W_t) = \theta(L)\epsilon_t \tag{5.10}$$

$$\epsilon_t = \nu_t h_t^{1/2} \quad with \;\; h_t = \alpha_0 + \sum_{i=1}^{s} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{r} \beta_j h_{t-j} + \sum_{f=1}^{k+l} \varrho_f z_{f,t} \tag{5.11}$$

where $z_t = [x_t', w_t']'$, and where $\varrho = (\varrho_1, \varrho_2, ..., \varrho_{k+l})'$ is a $((k+l) \times 1)$ vector of coefficients. Koopman *et al.* (2007) assume autoregressive fractionally integrated moving average noises which we do not consider. And Koopman *et al.* (2007) include water reservoir, which as we explained we think less adequate to forecast local prices, and consumption, which we exclude to investigate the impact of weather on electricity prices both via the demand and the supply.

5.5 Empirical Results

We apply the models described in Section 5.4 to our dataset, and assess which model performs best in terms of forecasting accuracy. Before the out-of-sample forecast exercise, we estimate the set of models using the complete sample to have an ex-post predictability idea.

We describe some assumptions. We apply nonlinear ordinary least square (NLS) estimator (Davidson and MacKinnon (1993)) for ARMA type models and the quasi maximum likelihood (QML) estimator (Davidson and MacKinnon (1993) and Greene (1993)) for GARCH family models.

We restrict our ARMA type models to be ARMA(7,0), where only lags 1 and 7 are considered [9]. We do the same for the level equation of the GARCH models. Autocorrelation analysis and in-sample criteria would suggest more complex ARMA forms. However, the risk of over-parametrization and the evidence presented in previous studies, for example Lucia and Schwartz (2002) show that an ARMA(7,0) specification provides optimal forecasts on daily day-ahead electricity prices. Hence, this convinces us to restrict the models to the aforementioned specification. Following the same reasoning we choose a GARCH(1,1) specification for the variance equation of models 4, 5, and 6.

The inclusion of the weather variables follows from statistical evidence. We allow different transformations of the weather forecast variables on the three markets to incorporate the fact that the weather may affect only the supply of electricity, which is different in the three markets. Since the influence of the weather variables appears to be non-linear, the generic initially unrestricted model in all the three exercises includes the power transformation of the weather forecasts. We use selection criteria, as the the adjust R-square and Akaike information criteria, and parameter statistical significance to specify the model.

5.5.1 In-sample analysis: Oslo Case

The in-sample analysis is based on the overall sample, from December 24, 2003 to March 14, 2006. We start with the ARMAX model which is considered to be a very accurate forecasting model. The ARMAX model in Lucia and Schwartz (2002) is specified as:

$$P_t = X_t + \phi_1(P_{t-1} - X_{t-1}) + \phi_7(P_{t-7} - X_{t-7}) + \epsilon_t \tag{5.12}$$

where

$$X_t = c + d_1 D_{hol,t} + d_2 sin(2\pi t/365.25) + d_3 cos(2\pi t/365.25),$$

where P_t is the log of the price at day t, and where $D_{hol,t}$ is a dummy variable with value 0 if day t is a working day or 1 if day t is not a working day. Figure 5.8 shows that the errors of the ARMAX model have non-linear relations with the daily average temperature and the total precipitation, but it has a linear-like relation with the wind speed. Therefore, for temperature forecasts we use the level, the square and the cubic; for precipitation and wind forecasts we use the level and the square. The selection criteria result in the following reduced specific ARMAXW model for Oslo data:

$$P_t = X_t + W_t + \phi_1(P_{t-1} - X_{t-1} - W_{t-1}) + \phi_7(P_{t-7} - X_{t-7} - W_{t-7}) + \epsilon_t \tag{5.13}$$

[9] We also try an ARMA(7,0) with all the seven lags such as an ARMA(7,0) specified using all the data, therefore also information not available at time of forecasting, and results are almost identical to the chosen model.

where

$$X_t = c + d_1 D_{hol,t} + d_2 sin(2\pi t/365.25) + d_3 cos(2\pi t/365.25),$$

$$W_t = a_1 Temp_t + a_2 Temp_t^3 + b_1 Prec_t + b_2 Prec_t^2 + \gamma Wind_t,$$

where $Temp_t$, $Prec_t$ and $Wind_t$ are the forecasts on daily average temperature, total precipitation and wind speed, respectively, on day t. The estimation procedure indicates that the square of the temperature and the square of the wind can be excluded. The square of the temperature does not take into account the difference between very low and very high temperature, which is a serious limitation. The wind seems to have a direct linear relation with prices. The empirical findings are consistent with the graphical analysis.

Table 5.2 gives the results of the estimation of model (5.13) with the chosen W_t over the complete sample. We discuss the estimated coefficients for the temperature forecasts, a_1 and a_2. The temperature forecasts affect the day-ahead electricity price via the following function:

$$f(Temp_t) = a_1 Temp_t + a_2 Temp_t^3$$

Taking the first order derivative, we get

$$\frac{\mathrm{d}f(Temp_t)}{\mathrm{d}Temp_t} = a_1 + 3a_2 Temp_t^2.$$

By substituting in the previous equation a_1 and a_2 with their empirical estimates, and solving $\frac{\mathrm{d}f(Temp_t)}{\mathrm{d}Temp_t} = 0$, we find the roots as ± 15. From our data set, the minimum observed temperature is -15. So the only switch point is $Temp^* = 15$. When the temperature is lower than the switch point, it is negatively influenced, i.e. the lower the forecasted temperature, the higher the electricity price. Notice that when the temperature forecast is above the switch point, it is positively influenced, i.e. the higher the forecasted temperature, the higher the electricity price. Intuitively, it reflects the fact that when temperature forecast is relatively higher or lower than the switch point, the consumption of the electricity will arise. Meanwhile the difficulty of producing electricity is also increased when it is extremely hot or cold. This suggest that the weather forecasts can influence both the demand and supply side of the power. For further discussion on which side the weather forecasts affect the demand-supply curves, we refer to the next section.

Comparing to the ARMAX, the improvement of introducing the weather forecast variables is not impressive for the in-sample analysis as shown in Table 5.2. The inclusion of weather variables seems also appropriate in the GARCH specification. The parameters of the GARCHW equation are less persistent than the GARCH counterpart. Model 5 has the lowest Akaike information criteria.

Table 5.2: In-sample estimation: Oslo

Models	ARMAX	ARMAXW	ARMAX - GARCH	ARMAXW - GARCH	ARMAXW - GARCHW
ϕ_1	0.811	0.782	0.886	0.874	0.887
	[0.022]	[0.0224]	[0.018]	[0.019]	[0.019]
ϕ_7	0.174	0.2040	0.118	0.121	0.112
	[0.023]	[0.023]	[0.020]	[0.021]	[0.021]
c	3.4907	3.5410	2.871	2.927	6.055
	[0.122]	[0.127]	[0.970]	[0.973]	[3.205]
d_1	-0.0491	-0.0500	-0.030	-0.030	-0.031
	[0.004]	[0.004]	[0.002]	[0.002]	[0.002]
d_2	0.0412	0.011	-0.095	-0.129	-0.139
	[0.060]	[0.055]	[0.056]	[0.052]	[0.063]
d_3	-0.046	-0.098	-0.089	-0.105	-0.138
	[0.064]	[0.058]	[0.062]	[0.058]	[0.066]
a_1	-	-0.008	-	-0.004	-0.004
		[0.0013]		[0.001]	[0.001]
a_2	-	1.15E-05	-	3.06E-06	2.36E-06
		[3.84E-06]		[2.96E-06]	[2.80E-05]
b_1	-	-0.014	-	-0.007	-0.011
		[0.009]		[0.006]	[0.007]
b_2	-	0.006	-	0.002	0.004
		[0.005]		[0.003]	[0.004]
γ	-	-0.003	-	-0.001	-0.001
		[0.002]		[0.001]	[0.001]
α_0	-	-	8.71E-05	8.27E-05	5.28E-05
			[1.75E-05]	[1.45E-05]	[7.26E-05]
α_1	-	-	0.304	0.382	0.411
			[0.036]	[0.044]	[0.050]
β_1	-	-	0.696	0.651	0.584
			[0.026]	[0.027]	[0.033]
ϱ_1	-	-	-	-	2.35E-04
					[7.26E-05]
ϱ_2	-	-	-	-	9.38E-06
					[3.99E-05]
ϱ_3	-	-	-	-	2.92E-05
					[5.53E-05]
ϱ_4	-	-	-	-	4.08E-06
					[6.16E-06]
ϱ_5	-	-	-	-	1.48E-08
					[1.65E-08]
ϱ_6	-	-	-	-	-1.66E-04
					[1.67E-04]
ϱ_7	-	-	-	-	3.79E-04
					[1.69E-04]
ϱ_8	-	-	-	-	-2.25E-06
					[1.09E-05]
R-squared	0.909	0.917	0.907	0.909	0.909
Adj. R-squared	0.908	0.915	0.906	0.907	0.907
AIC	-3.286	-3.344	-3.661	-3.731	-3.731

Note: The table reports the coefficient estimates (and their standard errors between square brackets), and selection criteria tests of the models 2-6 with Oslo data.

As discussed in Section 5.4, we stand with the hypothesis of stationary for out-of-sample analysis.

5.5.2 Out-of-sample analysis: Oslo Case

The objective of the out-of-sample analysis is to forecast the electricity price from January 1, 2005 to March 14, 2006. We repeat the selection procedure in Section 5.5.1 over the initial in-sample period, from December 24, 2003 to December 31, 2004. The reduced specific model remains the same as in (5.13). In forecasting, the model is re-estimated to make any new forecast, but it is not re-specified. An expanding window is used, which means that, to forecast the price of one day, all the previous data is included.

Two criteria (typically used in the electricity forecasting literature, see e.g. Conejo *et al.* (2005), Knittel and Roberts (2005), Shahidehpur *et al.* (2002), Weron (2006)) are computed to compare the models. The first one is the Root Mean Square Prediction Error (RMSPE), defined as

$$RMSPE = \sqrt{\frac{1}{n}\Sigma_{s=1}^{n}(P_{T+s} - \hat{P}_{T+s})^2}$$

where P_{T+s} is the log price at time $T+s$, where $\hat{P}_{T+s}$ is the forecasted log price at time $T+s$, where $n = 438$ is the number of days being forecasted. The alternative criterion is the Mean Absolute Percentage Prediction Error (MAPE), see for example Misiorek *et al.* (2006). It is defined as

$$MAPE = \frac{1}{n}\Sigma_{s=1}^{n}\frac{|p_{T+s} - \widehat{p}_{T+s}|}{p_{T+s}}$$

We apply all five models in section 3 to forecast the daily price for Oslo data, and calculate the RMSPE and MAPE statistics. We also report results for the Random Walk (RW) model. The results are given in Table 5.3.

The statistic results for all models outperform the RW model, implying evidence of predictability in the electricity prices. It is also clear that the model 4, ARMAXW, is the best under both the criteria. When we compare the RMSPE statistics of model 3 (the best among the non-weather models) and model 4 with each other, we observe that an improvement of 3.8% in favour of model 4.

We test whether the difference between two forecasting methods is significant in order to show precisely how large is the improvement of the new weather forecast model. Taking the mean square prediction error (MSPE) as loss function we apply the Diebold-Mariano test (Diebold and Mariano (1995)). The null hypothesis is

$$H_0 : \text{The square of the forecast errors are equal.}$$

Table 5.3: Out-of-sample forecasting results

	Oslo		Eastern Denmark		The Netherlands	
Model	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE
RW	0.0602	0.0117	0.2160	0.0372	0.3054	0.0528
ARMA	0.0561	0.0116	0.1915	0.0333	0.2313	0.0400
ARMAX	0.0509	0.0105	0.1774	0.0309	0.2149	0.0377
ARMAXW	0.0490	0.0100	0.1705	0.0306	0.2120	0.0370
ARMAX-GARCH	0.0524	0.0105	0.1839	0.0313	0.2114	0.0368
ARMAXW-GARCH	0.0510	0.0101	0.1784	0.0307	0.2076	0.0358
ARMAXW-GARCHW	0.0496	0.0100	0.1764	0.0306	0.2110	0.0365

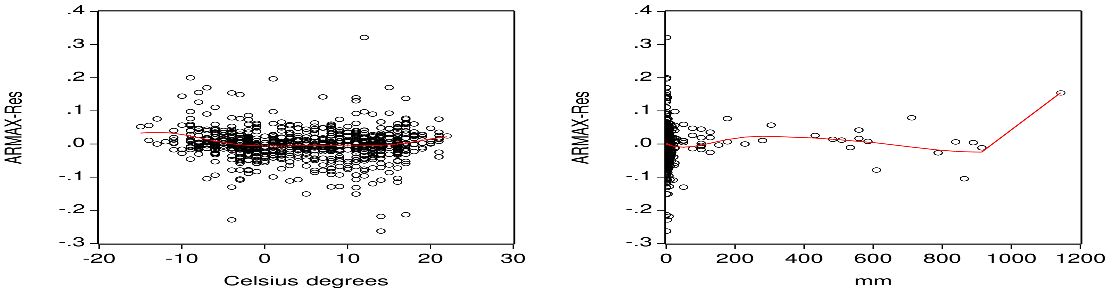
Note: The table reports forecasting statistics of the alternative models in the three electricity markets.

Statistics are in Table 5.4. The p-value of this test is $p = 0.0024$. We conclude that the ARMAXW model is significantly improved in terms of out-of-sample forecasting performance. Figure 5.9 shows the 60-day average RMSPE for the ARMAX model and the ARMAXW model. From the graph, we find that, when the error of model 3 is in a relatively lower level, the errors of two models are similar; but when there is a higher error from model 3 due to possible jumps, our weather forecast model often predicts better. Price jumps are mainly due to problems of inelasticity of the demand, and of non-storability of electricity with consequent shortage in the supply. These problems often arise when the weather conditions are adverse. Empirical results confirm the theoretical intuition that the weather forecasts help in predicting high prices or jumps, possible related to extreme adverse climate situations.

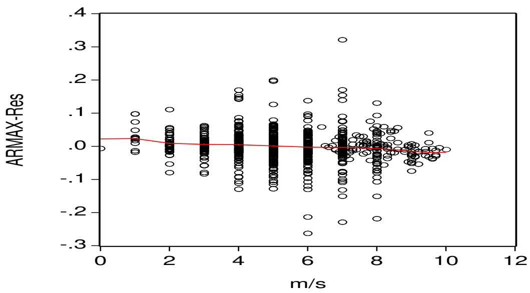
Adding weather forecast variables in a GARCH model is also very beneficial. Forecasts from the model 6, ARMAXW-GARCHW, give accurate forecasts that are quite similar to those of model 3. In contrast, model 4, ARMAX-GARCH, gives very poor forecasts, and extending the mean equation with weather variable, as model 5, ARMAXW-GARCH, is not enough. To sum up, a 'classical' GARCH specification is not adequate to predict electricity prices, but adding weather variables as shock indicators improve enormously the performance.

Although the weather forecast models show improvement on predicting the day-ahead prices, it is still mysterious whether this kind of influence is via the demand of the electricity or the supply. One way to verify this is to introduce the volume variable into a forecasting model. In principle, the volume indicates the demand of the electricity. Then, if the weather only influence the consumption of the power, introducing the volume at time $T+s$ to forecast the electricity price at time $T+s$ into the ARMAX model will lead to similar results as the ARMAXW model. We stress that the volume at time $T+s$ is not

Figure 5.8: Scatter plot: Oslo



(a)



(b)

Note: The graphs in this figure present in Panel a) the scatter plot of the errors of ARMAX model against the daily average temperature (in the left panel), and total precipitation (in the right panel); in Panel b) the scatter plot of the errors of ARMAX model against the wind speed in Oslo. The trend line is also provided in each figure.

known in advance, but previous literature form Engle *et al.* (1979) finds that it may be forecasted accurately, then we assume to know it. The model with volume (ARMAXV) is given as

$$P_t = X_t + V_t + \phi_1(P_{t-1} - X_{t-1} - V_{t-1}) + \phi_7(P_{t-7} - X_{t-7} - V_{t-7}) + \epsilon_t$$

$$X_t = c + d_1 D_{hol,t} + d_2 sin(2\pi t/365.25) + d_3 cos(2\pi t/365.25)$$

where V_t is the volume at time t. The calculated RMSPE is 0.0509, the improvement with respect to the ARMAX is 0.04%. We also apply the Diebold-Mariano test, the p-value is 0.8161. The comparison shows that introducing the volume on forecasting the day-ahead price is not comparable with the weather forecasts, even if the future (unknown in practice) volume is applied. In Oslo electricity market, the weather influence is not only via the demand of the electricity, but even more via the production of the electricity. This reflects to the producing method in Oslo, hydropower.

Table 5.4: Out-of-sample accuracy comparisons

	Oslo	Eastern Denmark	The Netherlands
RW	-4.001***	-5.184***	-6.522***
ARMA	-4.872***	-5.182***	-3.130***
ARMAXW	2.816***	4.063***	1.554
ARMAX-GARCH	-1.915*	-2.197**	4.152***
ARMAXW-GARCH	-0.171	-0.402	3.986***
ARMAXW-GARCHW	2.075**	0.499	1.928*

Note: The table reports Diebold-Mariano forecast accuracy comparison tests of the given models against those of the ARMAX model. The null hypothesis is that the two forecasts have the same mean square error. Positive values indicate superiority of the given models, one asterisk denotes significance relative to the asymptotic null hypothesis at 10%, two asterisks denote significance relative to the asymptotic null hypothesis at 5%, and three asterisks denote significance relative to the asymptotic null hypothesis at 1%.

5.5.3 Further Application: Eastern Denmark Case

From the in-sample estimation we find that the specified model for the Eastern Denmark data only depends on the temperature and wind speed as follows

$$P_t = X_t + W_t + \phi_1(P_{t-1} - X_{t-1} - W_{t-1}) + \phi_7(P_{t-7} - X_{t-7} - W_{t-7}) + \epsilon_t \tag{5.14}$$

$$X_t = c + d_1 D_{hol,t} + d_2 sin(2\pi t/365.25) + d_3 cos(2\pi t/365.25)$$

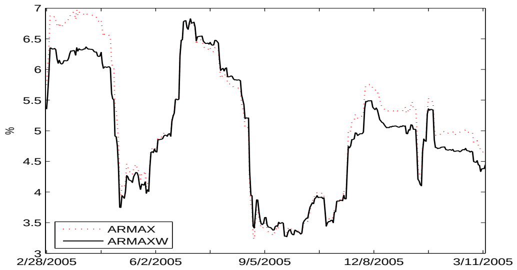
$$W_t = a_1 Temp_t + a_2 Temp_t^3 + \gamma Wind_t$$

The comparison of the six models is given in Table 5.3. Model 3, ARMAXW, remains the best model among the six models, along both criteria. In particular, for the RMSPE, the weather forecast model improves 3.89% from the ARMAX model. The difference is also statistically significant (the p-value of the Diebold-Mariano test is 0.00002).

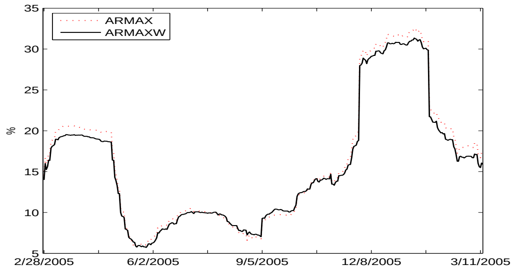
The reported evidence for Eastern Denmark market can be explained by how electricity is produced in that area. The wind power is a non-trivial part of the area's supply capacity. As in Oslo case, Figure 5.9 shows that the ARMAXW outperforms other models, when errors of other models are relatively higher and price jumps are observed.

The ARMAXW-GARCHW still provides reasonable accurate statistics, but they are always marginally higher than the statistics of the ARMAXW model and not statistically different. However, results of the ARMAXW-GARCHW are more accurate than results of the other two GARCH models, confirming again the role of weather forecasts in the GARCH specification.

Figure 5.9: 60 days average RMSPE



(a)



(b)

Note: The graphs in this figure present in Panel a) the 60 days moving average RMSPE given the ARMAX and ARMAXW models in forecasting Oslo log electricity prices; in Panel b) the 60 days moving average RMSPE given the ARMAX and ARMAXW models in forecasting Eastern Denmark log electricity prices.

5.5.4 A Different Story: The Netherlands Case

From our in-sample analysis we specify the weather variables W_t as for Eastern Denmark in equation (5.14). Therefore, forecasts on temperature, cubic of the temperature and wind speed are inserted in the regression model. This sounds realistic considering that the Netherlands is the country of windmills.

In this market, the ARMAXW does not provide the most accurate forecasts along any of the criteria. Compared to model 3, the improvement of the ARMAXW is only 1.32%, which is small and not statistically significant at 10% level.

Model 5, ARMAXW-GARCH, provides the lowest RMSPE. The ARMAX-GARCH and the ARMAXW-GARCHW also forecast accurately and their RMSPEs are similar to model 5. In terms of the Diebold-Mariano test the ARMAX-GARCH is the best model. The provided evidence suggests that the GARCH specification improves the forecast accuracy and that the contribution of the weather variables is only marginal. This can be explained by the nature of the power generation field in The Netherlands, which is predominantly thermal-based and therefore not directly related to weather. On the one hand, one could therefore argue that there is no ground for introducing weather forecasts as a price factor in such electricity market. On the other hand, this is also an evidence that when the production depends less on the weather, the weather forecasts play a minor role.

5.6 Conclusion

Electricity prices depend on several well-known temporal and regional price effects. Lucia and Schwartz (2002) have shown that including deterministic components that account for genuine regularities in the behavior of electricity prices give superior out-of-sample forecasts. However, more recent studies (see Bunn and Karakatsani (2003) and Knittel and Roberts (2005)) have found that the idiosyncratic price structure is not accurately described by Lucia and Schwartz (2002) model. In this Chapter we develop a set of models that add a new price factor in previous ARMA and ARMA-GARCH specifications: the weather forecasts.

Our empirical results suggest that the weather forecast variables play a central role in forecasting the day-ahead prices in different markets. In particular, the weather forecasts give relevant information to predict shocks in the prices. Intuitively, weather forecasts anticipate adverse weather conditions, which are often the cause of sharpen increase in the demand of electricity such as possible shortages in the supply. By studying this statement carefully, we find that the weather has high predictability power when the production plants of the market are related to the weather. This indicates that price jumps in those markets depend more on supply shortages or strategic supply decisions rather than demand increases.

The idea of weather forecast as new price factor also revaluates the GARCH class of models in forecasting electricity prices. Extending Koopman *et al.* (2007) we show that a GARCH process with weather forecasts predicts day-ahead prices successfully.

There are several topics for further research. Firstly, the set of weather forecasts might include other weather-related variables, such as water reservoir level, which to our knowledge are not modelled and forecasted at this time. Secondly, model 6 might be generalized by allowing seasonal variation in the parameters such as in periodic time series models and in periodic GARCH models, see Franses and Paap (2000). Weather forecasts might also be included in other nonlinear models, such as Markov regime-switching models, see Misiorek *et al.* (2006), or jump models. Finally, models based on weather forecasts might be used in derivative contracts. The reported evidence that weather has predictive power on the underlying day-ahead price process, could imply that this price factor might be reflected in the price of derivative instruments on day-ahead electricity contracts as well. Results might be extremely important since in most of the power derivative markets derivative contracts (e.g. callable options) are commonly traded.

Chapter 6

Summary

In this thesis we forecast financial time series. We focus on providing empirical rules to create optimal forecast, especially by applying model averaging. The analysis has been partitioned in four chapters. In the first two parts we concentrate on stock index data. In the next section we work on the term structure of interest rates. In the fourth part we focus on electricity prices.

In Chapter 2 we have reviewed time varying weight combination schemes. We have shown that time varying weights have features of adding individual models if information are heterogenous, and of coping with in-sample structural instability. In simulation exercises we have found that when data is subject to low predictability, strong heterogeneity of individual forecasts, and structural instability, they provide the best results comparing to other frequentist forecast combinations schemes and Bayesian model averaging methods. We have extended the analysis to forecast stock index returns, the S&P500 excess returns. As in the simulation exercise, stylized facts of stock index data are low predictability and possible structural instability. We have considered two forecasting models that represent different views on predicting stock index. The first one based on the assumption that a set of financial and macroeconomic variables have explanatory power, the second one based on the popular market saying “Sell in May and go away”, also known as the “Halloween indicator”. We show that firstly averaging strategies can give higher predictive gains than selecting the best models and believing on it. Secondly, the time varying weight schemes have higher statistical and economic values than other averaging methods, such as equal weights.

In Chapter 3 we have extended the previous averaging schemes to a new framework which models structural instability carefully. In particular the predictive specification that we have put forward allows for the treatment of three different sources of uncertainty, about the relevant predictor variables (model uncertainty), the values of the regression parameters (parameter uncertainty), and their stability (structural breaks).

Again, the implication of the three sources of uncertainty, and their relative importance, are investigated on predicting the S&P500 excess returns. Our empirical results suggest, first, that over the period 1966-2005 several structural breaks occurred in the relationship between US stock returns and predictor variables such as the dividend yield and interest rates. These changes appear to be caused by important events such as the oil crisis, changes in monetary policy, and the October 1987 stock market crash. Second, we find that allowing for model uncertainty and structural breaks jointly has considerable economic value. A typical investor would be willing to pay up to several hundreds of basis points annually to switch from a passive buy-and-hold strategy to an active strategy based on a return forecasting model that allows for model and parameter uncertainty as well as structural breaks in the regression parameters. The active strategy that incorporates all three sources of uncertainty performs considerably better than strategies based on more restricted return forecasting models.

In Chapter 4 we have moved to the analysis of forecasting the US term structure of interest rates. We have examined the forecast accuracy of a range of models with varying degrees levels of complexity. We have assessed the relevance of parameter uncertainty by examining the added value of using Bayesian inference compared to frequentist estimation techniques, and model uncertainty by averaging individual models. Following current literature we have also investigated the benefits of incorporating macroeconomic information in yield curve models. We have evaluated model forecasts over a ten-year out-of-sample period, using the entire period as well as several subperiod to show that the predictive ability of individual models varies over time considerably.

Our results show that adding macroeconomic factors is very beneficial for improving the out-of-sample forecasting performance of individual models. Models that incorporate macroeconomic variables seem more accurate in subperiods during which the uncertainty about the future path of interest rates is substantial. As an example we mention the period 2000-2003 when spreads were high. Models without macro information do particularly well in subperiods where the term structure has a more stable pattern such as in the early 1990s. Despite this, the predictive accuracy of models varies over time considerably, irrespective of using the Bayesian or frequentist approach. We show that mitigating model uncertainty by combining forecasts leads to substantial gains in forecasting performance, especially when applying a weighting method that is based on relative historical performance results and Bayesian inference on individual models, comparing to using individual models and the random walk benchmark.

In Chapter 5 we have studied how to improve forecasts of electricity prices. We have introduced the weather factor into well-known forecasting models to study its impact. In the literature the effects of weather on electricity sales are well-documented. However, studies that have investigated the impact of weather on electricity prices are still scarce, partly because the wholesale power markets have only recently been deregulated.

We find that weather has explanatory power for the day-ahead power spot price. Using weather forecasts improves the forecast accuracy, and in particular new models with power transformations of weather forecast variables are significantly better in term of out-of-sample statistics than popular mean reverting models. For different power markets, such as Norway, Eastern Denmark and the Netherlands, we build specific models. The dissimilarity among these models indicates that weather forecasts influence not only the demand of electricity but also the supply side according to different electricity producing methods.

To conclude, the research shows an increase of forecasting power of financial time series when parameter uncertainty, model uncertainty and optimal decision making are included. We highlight that, although the implementation of these techniques is not often straightforward and it depends on the exercise studied, the predictive gains are statistically and economically significant over different applications.

Nederlandse samenvatting (Summary in Dutch)

Deze dissertatie bundelt studies die nieuwe inzichten bieden in het voorspellen van financiële tijdreeksen. Verschillende optimale voorspellingstechnieken komen aan bod in vier studies die zich richten op het voorspellen van respectievelijk aandelenindices, de rente termijnstructuur, en elektriciteitsprijzen. Met name transformatie van verschillende voorspellingsmodellen naar een optimale gewogen voorspelling, ookwel bekend als "model averaging", passeert de revue.

Hoofdstuk 2 evalueert de voorspellende waarde van tijdsvariërende wegingen. We laten zien dat het gebruik van tijdsvariërende wegingen betere voorspellingen oplevert dan zowel traditionele "frequentist" voorspellingsmethoden als Bayesiaanse wegingstechnieken, met name wanneer de onderliggende data moeilijk voorspelbaar is en individuele voorspellingsmodellen heterogene en instabiele voorspellingen opleveren. Naast simulaties gebruiken we tevens voorspellingen van de rendementen op de S&P 500 index om de meerwaarde van tijdsvariërende wegingen aan te tonen. Het blijkt dat (i) gewogen voorspellingen succesvoller zijn dan het kiezen van n uniek model op basis van model selectie criteria, en (ii) tijdsvariërende wegingen leiden tot betere voorspellingen dan andere (statische) wegingsmethoden.

In Hoofdstuk 3 integreren we eerder besproken voorspellingsmethoden in een nieuw raamwerk dat drie vormen van voorspellingsonzekerheid aanpakt: onzekerheid over de optimale combinatie van voorspellende variabelen ("model uncertainty"), onzekerheid over de regressiecoefficienten behorende bij deze variabelen ("parameter uncertainty"), en parameter instabiliteit ("structural breaks"). Wederom worden de rendementen van de S&P 500 index gebruikt om de implicaties van deze onzekerheidsfactoren te benadrukken. Uit de empirische toetsen blijkt dat relaties tussen veelbesproken voorspellende factoren (zoals dividend yields en rentevoeten) en het S&P 500 rendement meerdere malen instabiel zijn geweest gedurende de periode 1966-2005. De innovaties in voorspellingen van modellen die omgaan met "model uncertainty" en parameter instabiliteit zijn economisch significant: een strategie gericht op het timen van de S&P 500 aan de hand van deze voorspellingen zal een belegger aanzienlijk hogere rendementen opleveren dan een simpele passieve (buy-and-hold) belegging in deze index.

In Hoofdstuk 4 staan voorspellingen van de Amerikaanse rentetermijn structuur centraal. Aan de hand van Bayesiaanse statistiek en gewogen voorspellingstechnieken wordt onzekerheid ten aanzien van de optimale voorspellingsfactoren en hun relatie met de rentetermijn structuur onderzocht. Het blijkt dat de voorspellende kracht van een gewogen gemiddelde van voorspellingen van individuele modellen aanzienlijk beter is dan de afzonderlijke (ongewogen) voorspellingen van deze modellen, waaronder die van het simpele random-walk model.

Tot slot concentreert Hoofdstuk 5 zich op het voorspellen van elektriciteitsprijzen. De impact van weerkarakteristieken op de spotprijs van elektriciteit staat centraal in dit onderzoek. We introduceren nieuwe modellen die weerkarakteristieken integreren in out-of-sample voorspellingen en tonen aan dat deze modellen beter presteren dan standaard mean-reverting modellen. Het blijkt dat de toegevoegde voorspellingswaarde van weerfactoren sterker is in landen waar het aanbod van elektriciteit relatief meer weersafhankelijk is.

Bibliography

Aiolfi, M. and A. Timmermann (2006), Persistence in Forecasting Performance and Conditional Combination Strategies., *Journal of Econometrics*, **135 (1)**, 31–53.

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

Ang, A. and G. Bekeart (2002), International Asset Allocation with Regime Shifts, *Review of Financial Studies*, **15**, 1137–1187.

Ang, A., S. Dong, and M. Piazzesi (2006a), No-Arbitrage Taylor Rules, *Working paper, Columbia University*.

Ang, A. and M. Piazzesi (2003), A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, *Journal of Monetary Economics*, **50**, 745–787.

Ang, A., M. Piazzesi, and M. Wei (2006b), What does the Yield Curve tell us about GDP Growth?, *Journal of Econometrics*, **131**, 359–403.

Avramov, D. (2002), Stock Return Predictability and Model Uncertainty, *Journal of Financial Economics*, **64**, 423–458.

Backus, D., S. Foresi, and C. Telmer (1998), Discrete-Time Models of Bond Pricing, *NBER Working paper*.

Barberis, N. (2000), Investing for the Long Run When Returns Are Predictable, *Journal of Finance*, **55**, 225–264.

Bates, J. M. and C. W. J. Granger (1969), Combination of Forecasts, *Operational Research Quarterly*, **20**, 451–468.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005), Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach, *Quarterly Journal of Economics*, **120**, 387–422.

Bernanke, B. S. and K. N. Kuttner (2005), What Explains the Stock Market's Reaction to Federal Reserve Policy?, *Jounral of Finance*, **60**, 1221–1257.

Bikbov, R. and M. Chernov (2005), No-Arbitrage Macroeconomic Determinants of the Yield Curve, *Working paper, Columbia University*.

Bliss, R. R. (1997), Testing Term Structure Estimation Methods, *Advances in Futures and Options Research*, **9**, 197–231.

Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

Bouman, S. and B. Jacobsen (2002), The Halloween Indicator, Sell in May and Go Away: Another Puzzle, *American Economic Review*, **92 (5)**, 1618–1635.

Bunn, D. W. and N. Karakatsani (2003), Forecasting Electricity Prices, *Working paper, London Business School*.

Campbell, J. Y. and R. J. Shiller (1988), Stock Prices, Earnings, and Expected Dividends, *Journal of Finance*, **43**, 661–676.

Campbell, J. Y. and S. B. Thompson (in press), Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, *Review of Financial Studies*.

Carter, C. and R. Kohn (1994), On Gibbs Sampling for State-Space Models, *Biometrika*, **81**, 541–553.

Chen, R. R. and L. Scott (1993), Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates, *Journal of Fixed Income*, **3**, 14–31.

Chib, S. (1995), Marginal Likelihood from the Gibbs Output, *Journal of American Statistical Association*, **90**, 972–985.

Christoffersen, P. F. and F. X. Diebold (1998), Cointegration and Long-Horizon Forecasting, *Journal of Business & Economic Statistics*, **16**, 450–458.

Clarida, R., J. Galí, and M. Gertler (2000), Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory, *Quarterly Journal of Economics*, **115**, 147–180.

Clemen, R. T. (1989), Combining Forecasts: A Review and Annotated Bibliography, *International Journal of Forecasting*, 559–583.

Cochrane, J. H. (2006), The Dog That Did Not Bark: A Defense of Return Predictability, Tech. rep., NBER.

Conejo, A. J., J. Contreras, R. Espinola, and M. A. Plazas (2005), Forecasting Electricity Prices foa a Day-Ahead Pool-based Electricity Energy Market, *International Journal of Forecasting*, **21(3)**, 435–462.

Corsi, F. (2004), Simple Long Memory Models of Realized Volatility, *Lugano Working paper*, 26.

Cox, J., J. E. Ingersoll, and S. A. Ross (1985), A Theory of the Term Structure of Interest Rates, *Econometrica*, **53**, 385–407.

Cremers, K. J. M. (2002), Stock Return Predictability: A Bayesian Model Selection Perspective, *Review of Financial Studies*, **15**, 1223–1249.

Dai, Q. and T. Philippon (2006), Fiscal Policy and the Term Structure of Interest Rates, *Working paper, New York University*.

Dai, Q. and K. J. Singleton (2000), Specification Analysis of Affine Term Structure Models, *Journal of Finance*, **55**, 1943–1978.

Davidson, R. and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

Dawid, A. P. (1984), Statistical Theory: The Prequential Approach, *Journal of the Royal Statistical Society Series A*, **147**, 278–292.

de Pooter, M., F. Ravazzolo, and D. van Dijk (2007), Predicting the Term Structure of Interest Rates, *Working paper, Tinbergen Institute*.

Deng, D. (2004), Modelling and investigating the relationship between electricity prices and water reservoir content at Nord Pool Power Exchange, *Working paper, Gothenburg University*.

DeWachter, H. and M. Lyrio (2006), Macro Factors and the Term Structure of Interest Rates, *Journal of Money, Credit and Banking*, **38**, 119–140.

Diebold, F. X. and C. Li (2006), Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics*, **130**, 337–364.

Diebold, F. X. and R. S. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business & Economic Statistics*, **13**, 253–263.

Diebold, F. X. and P. Pauly (1987), Structural Change and the Combination of Forecasts, *Journal of Forecasting*, **6**, 21–40.

Diebold, F. X., G. D. Rudebusch, and B. Aruoba (2006), The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach, *Journal of Econometrics*, **131**, 309–338.

Doppelhofer, G. and M. Weeks (2005), Jointness of Growth Determinants, Tech. rep., University of Cambridge.

Draper, D. (1995), Assessment and Propagation of Model Uncertainty, *Journal of the Royal Statistical Society Series B*, **56**, 45–98.

Duffee, G. (2002), Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance*, **5**, 405–443.

Duffie, D. and R. Kan (1996), A Yield-Factor Model of Interest Rates, *Mathematical Finance*, **6**, 379–406.

Eklund, J. and S. Karlsson (2007), Forecast Combination and Model Averaging using Predictive Measures, *Econometric Reviews*, **26**, 329–362.

Engle, R., C. W. J. Granger, R. Ramanathan, and A. Andersen (1979), Residential Load Curves and Time-Of-Day Pricing: An Econometric Analysis, *Journal of Econometrics*, **9**, 13–32.

Escribano, A., J. Pena, and P. Villaplana (2002), Modeling Electricity Prices: International Evidence, *Working paper, Universidad Carlos III de Madrid*.

Evans, C. L. and D. Marshall (1998), Monetary Policy and the Term Structure of Nominal Interest Rates: Evidence and Theory, *Carnegie-Rochester Conference Series on Public Policy*, **49**, 53–111.

Evans, C. L. and D. Marshall (2001), Ecoomic Determinants of the Nominal Treasury Yield Curve, *Federal Reserve Bank of Chicago Working paper series*, **16**.

Fama, E. F. and R. R. Bliss (1987), The Information in Long-Maturity Forward Rates, *American Economic Review*, **77**, 680–692.

Fama, E. F. and K. R. French (1988), Dividend Yields and Expected Stock Returns, *Journal of Financial Economics*, **22**, 3–25.

Fama, E. F. and K. R. French (1989), Business Conditions and Expected Returns on Stocks and Bonds, *Journal of Financial Economics*, **25**, 23–49.

Fernández, C., E. Ley, and M. F. J. Steel (2001), Model uncertainty in cross-country growth regressions, *Journal of Applied Econometrics*, **16**, 563–576.

Fleming, J., C. Kirby, and B. Ostdiek (2001), The Economic Value of Volatility Timing, *Journal of Finance*, **56**, 329–352.

Franses, P. H. and R. Paap (2000), Modeling Day-of-the-week Seasonality in the sp500 Index, *Applied Financial Economics*, **10**, 483–488.

French, K. R., G. W. Schwert, and R. F. Stambaugh (1987), Expected Stock Returns and Volatility, *Journal of Financial Economics*, **19**, 3–29.

Geman, H. (2005), *Commodities and Commodity Derivatives*, John Wiley & Sons.

Geman, S. and D. Geman (1984), Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

George, E. I. and R. E. McCulloch (1993), Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association*, **88**, 881–889.

Gerlach, R., C. Carter, and R. Kohn (2000), Efficient Bayesian Inference for Dynamic Mixture Models, *Journal of the American Statistical Association*, **95**, 819–828.

Geweke, J. and C. Whiteman (2006), Bayesian Forecasting, in G. Elliot, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, North-Holland.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005), There is a Risk-Return Trade-Off After All, *Journal of Financial Economics*, **76**, 509–548.

Giordani, P., R. Kohn, and D. van Dijk (2007), A Unified Approach to Nonlinearity, Outliers and Structural Breaks, *Journal of Econometrics*, **137**, 112–137.

Granger, C. W. J. and R. Ramanathan (1984), Improved Methods of Combining Forecasts, *Journal of Forecasting*, **3**, 197–204.

Greene, W. H. (1993), *Econometric Analysis, 4th. ed*, Prentice Hall.

Guidolin, M. and A. Timmermann (2005a), Economic Implications of Bull and Bear Regimes in UK Stock and Bond Returns, *Economic Journal*, **115**, 111–143.

Guidolin, M. and A. Timmermann (2005b), Asset Allocation under Multivariate Regime Switching, *Federal Reserve Bank of St. Louis*.

Guidolin, M. and A. Timmermann (2007), Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach, *forthcoming in Journal of Econometrics*.

Guo, H. and R. F. Whitelaw (2006), Uncovering the Risk-Return Relation in the Stock Market, *Journal of Finance*, **61**, 1433–1463.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Hansen, B. E. (2007), Least Squares Model Averaging, *Econometrica*, **75(4)**, 1175–1189.

Harvey, A. C. (1993), *Time Series Models*, Pearson Education.

Hendry, D. F. and M. P. Clements (2004), Pooling of Forecasts, *Econometric Reviews*, **122**, 47–79.

Hodges, J. (1987), Uncertainty, Policy Analysis and Statistics, *Statistical Science*, **2**, 259–291.

Hordahl, P., O. Tristani, and D. Vestin (2006), A Joint Econometric Model of Macroeconomic and Term-Structure Dynamics, *Journal of Econometrics*, **131**, 405–444.

Huurman, C., F. Ravazzolo, and Z. Chen (2007), The Power of Weather, *Working paper, Tinbergen Institute*.

Jagannathan, R. and T. Ma (2003), Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps, *Journal of Finance*, **58**, 1651–1683.

Jensen, G. R., J. M. Mercer, and R. R. Johnson (1996), Business Conditions, Monetary Policy, and Expected Security Returns, *Journal of Financial Economics*, **40**, 213–237.

Kandel, S. and R. Stambaugh (1996), On the Predictability of Stock Returns: An Asset Allocation Perspective, *Journal of Finance*, **51**, 385–424.

Keim, D. and R. Stambaugh (1986), Predicting Returns in the Stock and Bond Markets, *Journal of Financial Economics*, **17**, 357–424.

Kim, D. H. and J. H. Wright (2005), An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates, *Working paper, Federal Reserve Board Washington*, **33**.

Knittel, C. R. and M. R. Roberts (2005), An Empirical Examination of Restructred Electricity Prices, *Energy Economics*, **27**, 791–817.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley & Sons Ltd, West Sussex, England.

Koop, G. and S. M. Pooter (2004), Forecasting and Estimating Multiple Change-point Models with an Unknown Number of Change-points, *Working paper*.

Koopman, S. J., M. Ooms, and M. A. Carnero (2007), Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices, *Journal of American Statistical Association*.

Krippner, L. (2005), An Intertemporally-Consistent and Arbitrage-Free Version of the Nelson and Siegel Class of Yield Curve Models, *Working paper, University of Waikato*.

Kuo, L. and B. Mallick (1998), Variable Selection for Regression Models, *The Indian Journal of Statistics*, **60**, 65–81.

Leamer, E. (1978), *Specification Searches*, New York: Wiley.

Lettau, M. and S. van Nieuwerburgh (in press), Reconciling the Return Predictability Evidence, *Review of Financial Studies*.

Litterman, R. and J. Scheinkman (1991), Common Factors Affecting Bond Returns, *Journal of Fixed Income*, 54–61.

Lucia, J. J. and E. S. Schwartz (2002), Electricity Prices and Power Derivatives: Evidence from the Nordic Power Exchange, *Review of Dervatives Reserarch*, **5**, 5–50.

Madigan, D. M. and A. E. Raftery (1994), Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window, *Journal of the American Statistical Association*, **89**, 1335–1346.

Marcellino, M. (2004), Forecasting Pooling for Short Time Series of Macroeconomic Variables, *Oxford Bulletin of Economic and Statistics*, **66**, 91–112.

Marcellino, M., J. H. Stock, and M. W. Watson (2006), A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series, *Journal of Econometrics*, **127**, 499–526.

Marquering, W. and M. Verbeek (2004), The Economic Value of Predicting Stock Index Returns and Volatility, *Journal of Financial and Quantitative Analysis*, **39 (2)**, 407–429.

Min, C. and A. Zellner (1993), Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates, *Journal of Econometrics*, **56**, 89–118.

Misiorek, A., S. Trueck, and R. Weron (2006), Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models, *Studies in Nonlinear Dynamics & Econometrics*, **10**.

Mönch, E. (2006a), Term Structure Surprises: The Predictive Content of Curvature, Level, and Slope, *Working paper, Humboldt University Berlin*.

Mönch, E. (2006b), Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach, *Working paper, Humboldt University Berlin*.

Moral-Carcedo, J. and J. Vicns-Otero (2005), Modelling the non-linear response of Spanish electricity demand to temperature variations, *Energy Economics*, **27**, 477–494.

Nelson, C. R. and A. F. Siegel (1987), Parsimonious Modeling Of Yield Curves, *Journal of Business*, **60**, 473–489.

NordPool (2004), The Nordic Power Market: Electricity Power Exchange across National Borders, Tech. rep., Nord Pool.

Patelis, A. D. (1997), Stock Return Predictability and the Role of Monetary Policy, *Journal of Finance*, **52**, 1951–1972.

Paye, B. S. and A. Timmermann (2005), Instability of Return Prediction Models, *Journal of Empirical Finance*, **13**, 274–315.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006), Forecasting Time Series Subject to Multiple Structural Breaks, *Review of Economic Studies*, **73(4)**, 1057–1084.

Pesaran, M. H. and A. Timmermann (1995), Predictability of Stock Returns: Robustness and Economic Significance, *Journal of Finance*, **50**, 1201–1228.

Pesaran, M. H. and A. Timmermann (2002), Market Timing and Return Predictability Under Model Instability, *Journal of Empirical Finance*, **9**, 495–510.

Pettenuzzo, D. and A. Timmermann (2005), Predictability of Stock Returns and Asset Allocation under Structural Breaks, *Working paper*.

Piazzesi, M. (2003), Affine Term Structure Models, in Y. Ait-Sahalia and L. Hansen (eds.), *forthcoming in Handbook of Financial Econometrics*, Elsevier.

Pilipovic, D. (1997), *Energy Risk*, McGraw-Hill.

Politis, D. N. and J. P. Romano (1994), The Stationary Bootstrap, *Journal of the American Stastical Association*, **89**, 1303–1313.

Ravazzolo, F., R. Paap, D. van Dijk, and P. H. Franses (2007a), Bayesian Model Averaging in the Presence of Structural Breaks, in M. Wohar and D. Rapach (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Elsevier.

Ravazzolo, F., H. K. van Dijk, and M. Verbeek (2007b), Predictive Gains from Forecast Combination Using Time-varying Model Weight, *Econometric Institute Report 2007-26*.

Rudebusch, G. D. and T. Wu (2003), A Macro-Finance Model of the Term-Structure, Monetary Policy, and the Economy, *Working paper, Federal Reserve Bank of San Francisco*, **17**.

Sellin, P. (2001), Monetary Policy and the Stock Market: Theory and Empirical Evidence, *Journal of Economic Surveys*, **15**, 491–541.

Shahidehpur, M., H. Yamin, and Z. Li (2002), *Market Operation in Electricity Power Systems: Forecasting, Scheduling and Risk Management*, Wiley.

Stock, J. H. and M. W. Watson (2002a), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics*, **20**, 147–162.

Stock, J. H. and M. W. Watson (2002b), Forecasting Using Principal Components From a Large Number of Predictors, *Journal of the American Statistical Association*, **97**, 1167–1179.

Stock, J. H. and M. W. Watson (2004), Combination Forecasts of Output Growth in a Seven-country Data Set, *Journal of Forecasting*, **23**, 405–430.

Stock, J. H. and M. W. Watson (2005), Implications of Dynamic Factor Models for VAR Analysis, *Working paper*.

Strachan, R. and H. K. van Dijk (2007), Bayesian Model Averaging in Vector Autoregressive Processes with an Investigation of Stability of the US Great Ratios and Risk of a Liquidity Trap in the USA, UK and Japan, *Econometric Institute Report 2007-09*.

Tanner, M. A. and W. H. Wong (1987), The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, **82**, 528–550.

Terui, N. and H. K. van Dijk (2002), Predictability in the Shape of the Term Structure of Interest Rates, *International Journal of Forecasting*, **18**, 421–438.

Sala-i-Martin, X., G. Doppelhoffer, and R. Miller (2004), Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach, *American Economic Review*, **94**, 813–835.

Thornton, D. L. (2006), hen Did the FOMC Begin Targeting the Federal Funds Rate? What the Verbatim Transcripts Tell US, *Journal of Money, Credit and Banking*, **38**, 2039–2072.

Timmermann, A. (2006), Forecast Combinations, in G. Elliot, C. W. J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, North-Holland.

Vasicek, O. A. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, **5**, 177–188.

Weron, R. (2006), *Modelling and Forecasting Electricity Prices with Time Series Models*, John Wiley & Sons.

Weron, R. and A. Misiorek (2005), Forecasting Spot Electricity Prices with Time Series Models, *Proceedings of the European Electricity Market EEM-05 Conference*, 133–141.

White, H. (2000), A Reality Check for Data Snooping, *Econometrica*, **68**, 1097–1126.

Wilkinson, L. and J. Winsen (2002), What Can We Learn from a Statistical Analysis of Elctricity Prices in New South Wales, *The Electricity Journal*, **15**, 60–69.

Wu, T. (2005), Macro Factors and the Affine Term Structure of Interest Rates, *Working paper, Federal Reserve Bank of San Francisco*.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

365. S.M. BAJDECHI-RAITA, *The risk of investment in human capital*.
366. A.P.C. VAN DER PLOEG, *Stochastic volatility and the pricing of financial derivatives.*
367. R. VAN DER KRUK, *Hedonic valuation of Dutch Wetlands*.
368. P. WRASAI, *Agency problems in political decision making.*
369. B.K. BIERUT, *Essays on the making and implementation of monetary policy decisions*.
370. E. REUBEN, *Fairness in the lab: The effects of norm enforcement in economic decisions.*
371. G.J.M. LINDERS, *Intangible barriers to trade: The impact of institutions, culture, and distance on patterns of trade.*
372. A. HOPFENSITZ, *The role of affect in reciprocity and risk taking: Experimental studies of economic behavior.*
373. R.A. SPARROW, *Health, education and economic crisis: Protecting the poor in Indonesia.*
374. M.J. KOETSE, *Determinants of investment behaviour: Methods and applications of meta-analysis.*
375. G. MÜLLER, *On the role of personality traits and social skills in adult economic attainment*.
376. E.H.B. FEIJEN, *The influence of powerful firms on financial markets.*
377. J.W. GROSSER, *Voting in the laboratory.*
378. M.R.E. BRONS, M*eta-analytical studies in transport economics: Methodology and applications*.
379. L.F. HOOGERHEIDE, *Essays on neural network sampling methods and instrumental variables.*
380. M. DE GRAAF-ZIJL, *Economic and social consequences of temporary employment.*
381. O.A.C. VAN HEMERT, *Dynamic investor decisions.*
382. Z. ŠAŠOVOVÁ, *Liking and disliking: The dynamic effects of social networks during a large-scale information system implementation.*
383. P. RODENBURG, *The construction of instruments for measuring unemployment.*
384. M.J. VAN DER LEIJ, *The economics of networks: Theory and empirics.*
385. R. VAN DER NOLL, *Essays on internet and information economics.*
386. V. PANCHENKO; *Nonparametric methods in economics and finance: dependence, causality and prediction.*
387. C.A.S.P. SÁ, *Higher education choice in The Netherlands: The economics of where to go.*
388. J. DELFGAAUW, *Wonderful and woeful work: Incentives, selection, turnover, and workers' motivation.*

389. G. DEBREZION, *Railway impacts on real estate prices.*
390. A.V. HARDIYANTO, *Time series studies on Indonesian rupiah/USD rate 1995 – 2005.*
391. M.I.S.H. MUNANDAR, *Essays on economic integration.*
392. K.G. BERDEN, *On technology, uncertainty and economic growth.*
393. G. VAN DE KUILEN, *The economic measurement of psychological risk attitudes.*
394. E.A. MOOI, *Inter-organizational cooperation, conflict, and change.*
395. A. LLENA NOZAL, *On the dynamics of health, work and socioeconomic status.*
396. P.D.E. DINDO, *Bounded rationality and heterogeneity in economic dynamic models*.
397. D.F. SCHRAGER, *Essays on asset liability modeling*.
398. R. HUANG, *Three essays on the effects of banking regulations*.
399. C.M. VAN MOURIK, G*lobalisation and the role of financial accounting information in Japan.*
400. S.M.S.N. MAXIMIANO, *Essays in organizational economics.*
401. W. JANSSENS, *Social capital and cooperation: An impact evaluation of a women's empowerment programme in rural India.*
402. J. VAN DER SLUIS, *Successful entrepreneurship and human capital.*
403. S. DOMINGUEZ MARTINEZ, *Decision making with asymmetric information*.
404. H. SUNARTO, *Understanding the role of bank relationships, relationship marketing, and organizational learning in the performance of people's credit bank.*
405. M.Â. DOS REIS PORTELA, *Four essays on education, growth and labour economics.*
406. S.S. FICCO, *Essays on imperfect information-processing in economics.*
407. P.J.P.M. VERSIJP, *Advances in the use of stochastic dominance in asset pricing.*
408. M.R. WILDENBEEST, *Consumer search and oligopolistic pricing: A theoretical and empirical inquiry.*
409. E. GUSTAFSSON-WRIGHT, *Baring the threads: Social capital, vulnerability and the well-being of children in Guatemala*.
410. S. YERGOU-WORKU, *Marriage markets and fertility in South Africa with comparisons to Britain and Sweden.*
411. J.F. SLIJKERMAN, *Financial stability in the EU*.
412. W.A. VAN DEN BERG, *Private equity acquisitions*.
413. Y. CHENG, *Selected topics on nonparametric conditional quantiles and risk theory*.
414. M. DE POOTER, *Modeling and forecasting stock return volatility and the term structure of interest rates*.

In almost all cases a decision maker cannot identify ex ante the true process. This observation has led researchers to introduce several sources of uncertainty in forecasting exercises. In this context, the research reported in these pages finds an increase of forecasting power of financial time series when parameter uncertainty, model uncertainty and optimal decision making are included. The research contained herein evidences that although the implementation of these techniques is not often straightforward and it depends on the exercise studied, the predictive gains are statistically and economically significant over different applications, such as stock, bond and electricity markets.

Francesco Ravazzolo (1979) graduated in Economics and Business at Ca' Foscari University of Venice in 2002. In September 2003 he joined the Tinbergen Institute to carry out his doctoral research. Part of his research have been published and presented at international conferences. Currently, Francesco works at research department of Norges Bank as Senior Advisor.