Modeling and Forecasting Stock Return Volatility and the Term Structure of Interest Rates

ISBN: 978 90 5170 915 5

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

This book is no. 414 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.

Modeling and Forecasting Stock Return Volatility and the Term Structure of Interest Rates

Modelleren en voorspellen van de volatiliteit van aandelenrendementen
en de rentetermijnstructuur

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 27 september 2007 om 11.00 uur

door

Michiel David de Pooter
geboren te Terneuzen

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor:	Prof.dr. D.J.C. van Dijk
Overige leden:	Prof.dr. T.G. Andersen
	Prof.dr. P.H.B.F. Franses
	Dr.ir. M.P.E. Martens

Preface

The main question I asked myself before starting my PhD., now four and a half years ago, was whether this was really what I wanted to do and whether I was truly going to enjoy doing research. This question popped up in my mind on numerous occasions again during the past four years. Each time I thought it was still too soon to give a definite answer in order to try and decide whether I had taken the correct decision or not to start this journey. Now that I'm about to finish my PhD. it seems appropriate to try and finally answer it. I can now say that it has definitely been worth it. All in all, the last four and a half years have been a very interesting ride and it has been an experience I would not have wanted to miss. Four and a half years is, however, too long of a period to summarize in just one of two pages. Instead, I would just like to thank the many people who have contributed to this dissertation.

First and foremost I would like to thank my supervisor Dick van Dijk. His enthusiasm for doing research is infectious and has been a major influence on the type of researcher that I've become. During the past four years Dick was always willing and available to answer my questions, to share his experiences, to listen to my ideas and complaints, to guide me or to just let me do my own thing. Having a supervisor like Dick is an absolute privilege. I therefore hope that we can keep co-operating in the future.

I thank the members of the small committee, Philip Hans Franses, Torben Andersen and Martin Martens for evaluating an earlier draft of this dissertation and for providing useful comments. I also want to thank my other co-authors Francesco Ravazzolo, Rene Segers and Herman van Dijk. A special word of thanks goes out to Francesco, with whom I wrote down the initial idea for Chapter 6 on a beer mat once. Since then we've had countless discussions on research which in the future will no doubt lead to more joint projects.

I am grateful to NWO, Vereniging Trustfonds EUR, the Econometric Institute and Tinbergen Institute for their financial support which enabled me to attend conferences and to make several trips abroad and the Graduate School of Business at the University of Chicago for its hospitality. I especially want to thank Monika Piazzesi for hosting my visits to the GSB which convinced me that I should try and move to the U.S. more permanently. Furthermore, I want to thank Robin Lumsdaine for offering me the opportunity to temporarily work at Deutsche Bank in my second year as a PhD. student which in the end led to part B of this dissertation.

I want to thank friends and colleagues at the Tinbergen Institute and Econometric

Institute for a very pleasant working environment. Thanks go out also to the secretarial staff from both institutes for all their assistance. It was a pleasure to share offices with Grzegorz, Mariëlle and Matthijs during my time at the Tinbergen Institute.

Doing a PhD. can at times be difficult when progress is slow. Therefore it's great to know to have friends around, both on- and off-campus, who help you realize that there's more to life than just research. Knowing that I'm forgetting people, I want to mention: Chen, Francesco, Jelmer, Joop, Jeroen en Robin. Friends abroad, most of whom I probably never would've met if it hadn't been for my PhD(-travels) all had their share one way or another in the completion of this dissertation: Ali, Arthur, Ben, Birgit, Brian, Gary, LaLana, Laura, Robin, Sabine, Simon and Youjin. Last, but obviously not least, I want to thank my mother and sister who have always supported me throughout, although at times I must've seemed miles away to them, in a figurative as well as in a literal sense.

M

(Rotterdam, 070820)

Contents

List of Tables

Part A: Modeling and Forecasting Stock Return Volatility

Chapter 2

Chapter 3

Chapter 6

List of Figures

Part A: Modeling and Forecasting Stock Return Volatility

Chapter 2

Chapter 3

Part B: Modeling and Forecasting Interest Rates

Chapter 5

Chapter 6

Chapter 1

Introduction and Summary

1.1 Introduction and motivation

This dissertation consists of a collection of studies on two areas in quantitative finance: asset return volatility and the term structure of interest rates. Before describing the contents and contributions of the different chapters in more detail, a brief introduction of both topics is given.

Volatility is one of the crucial ingredients in many areas of finance. Examples of areas where volatility plays an important role are portfolio allocation, risk management and derivatives pricing. Volatility can be somewhat loosely defined as the variability of the random (or unpredictable) component of a time series variable. In finance these are typically asset returns, such as individual stocks or a stock index like the S&P 500 Index. The problem with volatility is that, unlike returns, it cannot be observed directly and therefore needs to be estimated. Hence, proxies need to be constructed in order to measure and subsequently model volatility.

A still popular approach is to use daily squared return innovations to approximate volatility. Although squared innovations are a noisy *ex-post* measure of the true volatility, they do illustrate its key property: volatility is not constant but varies over time. A second, related, property is that volatility is persistent which refers to the conditional variance displaying momentum meaning that past volatility explains current volatility. The degree of this persistence is important from an economic perspective as it signals whether shocks to volatility are permanent or transitory which affects, for example, risk premia. Persistence also suggests that unlike returns for which it is well established that these are very difficult to predict, volatility may be predictable to a much larger degree.

A workhorse model for capturing the persistence in conditional variance empirically has been the generalized autoregressive conditional heteroscedasticity (GARCH) model of Engle (1982) and Bollerslev (1986) which imposes a parametric autoregressive moving average (ARMA) structure on the squared return innovations. One drawback of the GARCH model is that it implies that shocks to volatility die out over time at an exponential rate. Research has shown, however, that the persistence of shocks may be better described by a so-called "long-memory" process under which shocks die out at a slower hyperbolic rate.

It is unclear whether volatility is truly long-memory. Although modelling this feature is certainly possible, finding a satisfactory explanation for why shocks should have such a long-lasting effect on volatility has proven to be difficult. A possible alternative explanation is given in for example Granger and Hyung (1999) and Diebold and Inoue (2001). These papers show that series that undergo breaks can mimic the properties of a long-memory series. As a result, breaks in volatility can potentially induce spurious long-memory features. It is therefore of importance to test and allow for structural breaks when modelling volatility.

Despite their popularity, it was long believed that models which are based on squared returns render poor volatility forecasts. Andersen and Bollerslev (1998a) show this is due to using an inaccurate ex-post volatility proxy. It was already shown by Merton (1980) that using intraperiod returns theoretically leads to error-free volatility measures. Andersen and Bollerslev (1998a) therefore suggest the use of a daily volatility measure that is based on the sum of squared intraday returns. They showed that judged from this so-called "realized volatility" measure the GARCH model forecasts well after all.

In addition to using it as an ex-post realization of the true observed volatility over a certain period, realized volatility can also be used to model volatility directly. Andersen, Bollerslev, Diebold, and Labys (2003a) show that volatility becomes virtually observable when realized volatility is based on an ever higher intraday frequency. Consequently, volatility can now be modelled directly using standard time-series techniques. Although the focus in the literature has primarily been on modelling the variance of individual asset returns, covariances *between* assets can similarly be modelled with high-frequency return data. Whereas traditional approaches of modelling the covariance matrix, e.g. the Multivariate GARCH model, have been seriously hampered by identification problems due to the large number of parameters, by using realized covariances modelling the entire covariance matrix becomes attainable.

Although the realized volatility literature has rapidly advanced over the last decade, various issues involving the use of realized measures are currently still under intense scrutiny.

One of the main questions arises when the theory is brought to the data. Although theory implies that higher sampling frequencies result in a more accurate measure of volatility, it is known that at higher sampling frequencies, intraday returns become increasingly contaminated with market microstructure noise due to for example bid-ask bounce and non-synchronous trading. A trade-off emerges between accuracy, which implies that the highest possible frequency should be used, and noise, which suggests that a lower frequency may be more suitable. An important question is therefore which sampling frequency should be used to construct realized variances and covariances.

The first part of this dissertation offers contributions to the literature on how to test for sudden changes in unconditional volatility, on modelling realized volatility and on the choice of optimal sampling frequencies for intraday returns.

The emphasis in the second part of this dissertation is on the term structure of interest rates. The term structure of interest rates, or yield curve, describes the relationship between interest rates and time to maturity. It determines the current value of future nominal payments and therefore guides economic decisions. The importance of studying the term structure becomes even clearer when realizing that long-term interest rates are risk-adjusted expectations of average future short rates and therefore contain information about future interest rates. Furthermore, the yield curve contains information about future economic activity in general. For example, the slope of the term structure (which is the difference between long and short yields) has successfully been used to forecast GDP growth and the occurrence of recessions. Macroeconomists are focusing more and more on trying to understand the relationship between interest rates, monetary policy and macroeconomic fundamentals.

The yield curve is therefore crucial for bond pricing, managing and hedging interest rate risk, public policy and monetary policy in particular as central banks often actively target the short end of the yield curve. Its importance explains why it is vital to first of all accurately estimate the current term structure, as the yield curve can often not be observed directly. Second of all, it is particularly important to be able to forecast the future yield curve. The focus in this dissertation is primarily on an out-of-sample forecasting perspective.

Forecasting interest rates based on past yield information only has proven to be difficult as a simple random walk model often outperforms complex models, see e.g. Duffee (2002). However, recent studies have shown that interest rates seem to be predictable after all, at least to a certain extent, mainly by incorporating macroeconomic information into models of the term structure. Identifying a single model with a consistently accurate forecasting performance is nevertheless still a difficult task.

The second part of this dissertation contributes to the literature by examining in greater detail one particular class of term structure models. Furthermore, the forecasting performance of several models of the term structure is examined for its stability and approaches for combining the informational contents of forecasts from different models are analyzed.

1.2 Summary and conclusions

This dissertation is divided into two parts. Part A consists of Chapters 2, 3 and 4 and focuses on volatility modelling. The main emphasis in the latter two chapters is on the use of high-frequency intraday return data for analyzing volatility from a univariate and multivariate perspective. Chapters 5 and 6 constitute Part B and discuss estimating and, in particular, forecasting the term structure of interest rates.

Part A: Modeling and forecasting stock return volatility

As stated in the introduction, the volatility of financial time-series has been found to undergo sudden, structural level shifts. Failing to account for these when modelling volatility can result in biased estimates of the volatility persistence. It is therefore important to identity the timing and size of these level shifts. Chapter 2, which is based on De Pooter and van Dijk (2004), considers tests for sudden changes in the unconditional volatility of conditionally heteroskedastic time series based on cumulative sums of squares (CUSUM). A prominent conclusion from the analysis in this chapter is that applying these tests to the raw time-series observations leads to severe size distortions, where the correct null hypothesis of no volatility change is rejected much too frequently. It is shown that it appears necessary, unless unrealistically large sample sizes are used, to filter the series in order to remove the heteroskedasticity prior to applying the CUSUM test. Extensive Monte Carlo simulations show that applying these tests to standardized residuals from an estimated GARCH(1,1) model results in good size and reasonable power properties when testing for a single break in the variance. The tests then also appear to be fairly robust to different types of misspecification. The chapter develops an iterative algorithm to test sequentially for the presence of multiple changes in volatility. The CUSUM tests appear to have difficulty, however, to detect multiple changes.

The application of the tests in this chapter to emerging markets stock returns confirms the properties of the different test statistics. Although there are concerns that the tests may be somewhat conservative, it is shown that the GARCH-filtered tests, compared to other studies that applied the original CUSUM statistics, lead to a considerably smaller, and much more realistic number of volatility changes.

Chapter 3 is based on Martens, van Dijk, and de Pooter (2004) and incorporates nonlinearities in models for realized volatility. In particular, a nonlinear autoregressive fractionally integrated model (ARFI) is proposed which accommodates level shifts, day-of-the-week and holiday effects, (pre-)announcement effects and leverage effects. The model incorporates long-memory as well as structural breaks, thus allowing for a analysis of both effects volatility persistence separately. The full model, as well as several restricted versions, are estimated for the S&P 500 stock index. The nonlinear model improves the in-sample description of the data with all individual nonlinearities being highly significant. The model also produces volatility forecasts that, for horizons up to 20 days, improve upon those obtained from a linear ARFI model and conventional squared return based time-series models that treat volatility as a latent variable. Adding the nonlinearities to simpler autoregressive models for realized volatility leads to similar improvements.

Many financial applications require estimates of not just variances but of the entire covariance matrix, making covariances or correlations between different assets just as important. Chapter 4 focuses on the merits of using high-frequency data for measuring and forecasting the daily covariance matrix. Whereas in Chapter 3 the popular five-minute frequency for the intraday returns is adopted for constructing realized variance, motivated as a trade-off between accuracy and potential market microstructure biases, this chapter addresses the issue of choosing an optimal sampling frequency directly. The optimal frequency is determined by forming and subsequently judging the performance of mean-variance efficient stock portfolios with daily rebalancing from the individual constituents of the S&P 100 index. Despite the fact that the S&P 100 stocks are relatively liquid, the optimal sampling frequency is shown to range between 30 and 65 minutes, considerably lower than the five-minute frequency. The results are found to be robust to the use of different bias-correction procedures, transaction costs and the portfolio rebalancing frequency. This third and final chapter of Part A is based on De Pooter, Martens, and van Dijk (2008).

Part B: Modeling and forecasting the term structure of interest rates

The second part of this dissertation is centered around modelling the term structure of interest rates. Chapter 5 is based on De Pooter (2007) and examines one particular class of term structure models in detail. The focus is on the class of Nelson-Siegel models which are gauged on their ability to fit the term structure at specific dates as well as on their performance in forecasting the future term structure. The chapter addresses various estimation techniques when using these models. The results demonstrate that it is worthwhile to extend the original Nelson and Siegel (1987) three-factor model with additional factors and by freeing up parameters in order to improve the accuracy with which to fit the term structure. Whereas this result is to be expected, it is further shown that a more flexible specification, in particular through the use of a four-factor model, produces also highly accurate forecasts. The four-factor model outperforms benchmark models across maturities and forecast horizons. Subsample analysis reveals that the outperformance is consistent over time.

Whereas the focus in Chapter 5 is on trying to identify a single model that forecasts future interest rates accurately and does so consistently over time, Chapter 6 takes a different approach and tries to forecast interest rates using a panel of models instead. This chapter is based on De Pooter, Ravazzolo, and van Dijk (2007) and examines the relevance of parameter uncertainty, model uncertainty, and macroeconomic information. The chapter examines the forecast performance of a range of models with varying degrees of complexity. This performance is assessed over a ten-year out-of-sample period and it is shown through subsample analysis that the predictive ability of individual models varies considerably over time. Trying to identify a best single model is therefore complicated which in turn provides strong support for combining the forecasts from multiple models to mitigate this model uncertainty. Several forecast combination techniques are therefore examined. It is demonstrated that combining forecasts leads to substantial gains in predictability, in particular with the use of a weighting method that is based on relative historical performance and combining models that include macroeconomic factors. The largest gains in forecasting performance are obtained for longer maturities in particular and these gains are shown to be consistent over time.

Part A

Modeling and Forecasting Stock Return Volatility

Chapter 2

Testing for Changes in Volatility in Heteroskedastic Time Series

A further examination

2.1 Introduction

Quite soon after the interest in modelling the conditional heteroskedasticity of financial time series variables developed in the early 1980s, the possibility was raised that these variables experience occasional large shifts in unconditional volatility, see Diebold (1986) and Lamoureux and Lastrapes (1990). While the issue of testing for changes in the unconditional variance of time series has received considerable attention in the literature, dating back to at least Wichern *et al.* (1976), most of the available testing procedures implicitly or explicitly assume constant conditional volatility, see Hsu (1977), Talwar and Gentle (1981), Sakata (1988), Inclán and Tiao (1994), and Chen and Gupta (1997), among others. Recently, however, Kim, Cho, and Lee (2000) and Kokoszka and Leipus (2000) have developed tests that use cumulative sums of squares (CUSUMs) to test for breaks in the unconditional variance of possibly heteroskedastic time series. The properties of the CUSUM statistic of Kokoszka and Leipus (2000) were examined by Andreou and Ghysels (2002, 2004), finding that the test has good power properties but also noting some problems, in that the test sometimes suffers from quite large size distortions. The purpose of this paper is to examine this deficiency of these CUSUM tests in more detail. In particular, we investigate whether we can fix the size properties by adopting the suggestion of Lee *et al.* (2003) to apply the CUSUM statistic to standardized residuals from an estimated GARCH model. An elaborate simulation analysis confirms that the tests have severe size distortions when applied to the original series, such that the correct null hypothesis of no change is rejected much too frequently, rendering the tests highly unreliable. However, when applied to standardized GARCH residuals, the tests are found to have only minor size distortions and reasonably good power in detection volatility changes. Furthermore, the tests appear to be quite robust to various types of misspecification. We apply the testing procedures to examine breaks in the unconditional volatility of a set of emerging stock market returns. Doing so allows to further assess the properties of the CUSUM tests and to compare the obtained results with earlier studies such as Aggarwal *et al.* (1999).

The outline of the paper is as follows. In Section 2.2, we discuss the three CUSUM tests that we consider in this study. In particular we demonstrate that all three tests emanate from the same basic setup. We also pay considerable attention to testing for multiple breaks and to the use of finite sample critical values. In Section 2.3, we use extensive Monte Carlo experiments to assess the size of the tests and their power for detecting both single as well as multiple breaks. We find that the tests, when applied to standardized returns, work reasonably well under different data generating processes and have quite good power properties. In Section 2.4 we apply the tests to daily emerging stock market returns. We find that the tests are certainly suitable for detecting variance changes in these series but that the results should be interpreted carefully, as two of the tests seem to have have a tendency to be conservative and potentially underestimate the number of actual breaks. Section 2.5 concludes.

2.2 CUSUM tests for changes in volatility

The issue that we want to address in this paper concerns testing for changes in the unconditional variance of a time series variable, in particular in the presence of conditional heteroscedasticity. Let $\{y_t\}_{t=1}^T$ denote the time series of interest with T being the available sample size, and assume (for simplicity, but without loss of generality) that y_t has a constant mean equal to 0. We consider the problem of testing the null hypothesis that the unconditional variance of y_t is constant, that is $\mathsf{H}_0 : \sigma_t^2 = \sigma^2$ for all $t = 1, \ldots, T$, against the alternative hypothesis of a single structural break, that is

$$\mathsf{H}_a : \sigma_t^2 = \begin{cases} \sigma_0^2 & \text{for } t = 1, \ldots, \kappa, \\ \sigma_1^2 & \text{for } t = \kappa + 1, \ldots, T, \end{cases} \tag{2.1}$$

where the change-point κ is assumed unknown. Many different approaches for tackling this testing problem have been developed, see the references in Section 2.1. Here we limit ourselves to test statistics based on cumulative sums of squares (CUSUMs), as first proposed by Inclán and Tiao (1994) and subsequently further developed by Kim *et al.* (2000), Kokoszka and Leipus (2000), and Lee and Park (2001). In this section we first discuss the design of the CUSUM statistics for the above single break testing problem. In particular, we demonstrate that all tests, which might appear to be quite different at first sight, nevertheless fit into a single framework. Next, we address the problem of testing for multiple breaks in volatility. We conclude this section with a discussion on the use of finite sample critical values.

2.2.1 Testing for a single structural change

Our starting point is the cumulative sum of squares (CUSUM) process $C_y(k) = \sum_{t=1}^{k} y_t^2$. The centered and normalized CUSUM process is then defined as

$$D_y(k) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{k} y_t^2 - \frac{k}{T\sqrt{T}} \sum_{t=1}^{T} y_t^2, \tag{2.2}$$

such that $D_y(0) = D_y(T) = 0$. When y_t satisfies the null hypothesis of constant unconditional variance, a plot of $D_y(k)$ against k will be a horizontal line oscillating around zero. However, under the alternative of a sudden change in the variance occurring at a certain point κ during the sample, the value of $D_y(k)$ will move away from zero in either the positive or negative direction for values of $k < \kappa$. Theoretically, the absolute value of $D_y(k)$ will achieve its maximum at $k = \kappa$, after which it will return towards zero. For this reason $|D_y(k)|$ provides a natural test for a volatility change, as well as an estimate of the change-point.

Suppose that the maximum of $|D_y(k)|$ is attained at $k = k^*$, that is

$$|D_y(k^*)| = \max_{1 \le k \le T} |D_y(k)|. \tag{2.3}$$

We then identify a breakpoint at k^* if $|D_y(k^*)|$ is larger then some predetermined critical value, which can be obtained from the asymptotic distribution of $D_y(k)$. It can be shown that under fairly mild regularity conditions, see Boswijk (2004) among others, that under the null $D_y(k)$ weakly converges to a (scaled) Brownian bridge, such that

$$\frac{1}{\gamma}|D_y(k^*)| \xrightarrow{d} \sup_{0 \le s \le 1} |B(s)|, \tag{2.4}$$

where γ^2 is the long-run variance of the squared series y_t^2, that is $\gamma^2 = \sum_{j=-\infty}^{\infty} \gamma_j$ with γ_j the j-th order autocovariance of y_t^2, and where $B(s)$ is a standard Brownian bridge, defined as $B(s) = W(s) - sW(1)$ with $W(\cdot)$ a standard Wiener process and $0 \le s = k/T \le 1$. It follows that an appropriate CUSUM test statistic is given by

$$U_y(k^*) = \frac{1}{\hat{\gamma}} \max_{1 \le k \le T} |D_y(k)|, \tag{2.5}$$

where $\hat{\gamma}^2$ is a consistent estimator for γ^2.

Obviously, the (assumptions concerning the) distributional properties of the time series y_t determine its long-run variance γ^2 and, furthermore, imply how it should be estimated. It is in this respect that the various CUSUM statistics that have been proposed differ. First, assuming that $\{y_t\}_{t=1}^T$ is a sequence of independent and identically distributed (iid) normal random variables, as in Inclán and Tiao (1994), the autocovariances of y_t^2 are all equal to zero, that is $\gamma_j = 0$, $\forall j \neq 0$, such that the long-run variance $\gamma^2 = \gamma_0$. Due to the normality assumption γ in fact reduces to $\sigma^2\sqrt{2}$, where σ^2 is the variance of y_t,[1] which can be consistently estimated by $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T y_t^2 = \frac{1}{T}C_y(T)$. It is then straightforward to show that the CUSUM statistic $U_y(k^*)$ as given in (2.5) is equivalent to

$$\sqrt{\frac{T}{2}} \max_{1 \le k \le T} \left| \frac{C_y(k)}{C_y(T)} - \frac{k}{T} \right|,$$

which is the form used in Inclán and Tiao (1994).

Second, assuming that y_t is iid but not necessarily normally distributed, the long-run variance γ^2 is still equal to γ_0, but now γ_0 should be estimated directly from the time series y_t as $\hat{\gamma}_0 = \frac{1}{T}\sum_{t=1}^T (y_t^2 - \sum_{t=1}^T y_t^2)^2 = \frac{1}{T}\sum_{t=1}^T y_t^4 - (\frac{1}{T}\sum_{t=1}^T y_t^2)^2$.

Third, one may relax the iid assumption and allow for various forms of dependence and heterogeneity in y_t. For example, Lee and Park (2001) allow for temporal dependence by assuming that y_t follows an MA(∞) process, that is $y_t = \sum_{j=1}^\infty \theta_j \varepsilon_{t-j} + \varepsilon_t$. Here we are mainly interested in cases where y_t displays conditional heteroskedasticity. In that respect, Kokoszka and Leipus (2000) assume that y_t follows an ARCH(∞) process,

$$\begin{aligned} y_t &= z_t\sqrt{h_t}, \\ h_t &= \omega + \sum_{j=1}^{\infty} \alpha_j y_{t-j}^2, \end{aligned} \tag{2.6}$$

with α_j being non-negative constants and $z_t \sim$ iid $N(0,1)$. Alternatively, Kim *et al.* (2000)

[1] If y_t is iid normal and assuming that $\mathsf{E}[y_t] = 0$, the kurtosis of y_t equals $3 = \mathsf{E}[y_t^4]/((\mathsf{E}[y_t^2])^2)$. Consequently, $3\sigma^4 = E[y_t^4]$. Given that $\gamma_0 = \mathsf{E}[(y_t^2 - \mathsf{E}[y_t^2])^2] = \mathsf{E}[y_t^4] - (\mathsf{E}[y_t^2])^2$, it follows that $\gamma_0 = 2\sigma^4$.

assume a GARCH(1,1) process for y_t,[2]

$$y_t = z_t\sqrt{h_t},$$
$$h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}, \tag{2.7}$$

with α, β positive constants such that $\alpha + \beta < 1$ and again $z_t \sim \text{iid } N(0,1)$.

In all these cases, the squared series y_t^2 has non-zero autocorrelations γ_j, $j \neq 0$, at all lags and, consequently, $\hat{\gamma}_0$ does not provide a consistent estimate of the long run variance γ^2. One possible solution to this problem is to derive an explicit expression for γ_j, and thereby for γ^2, based on the specific parametric structure of the process that is assumed for y_t, as is done in Kim *et al.* (2000) for the GARCH(1,1) case. However, one can imagine that this procedure is rather sensitive to model misspecification.[3] An alternative and more robust approach is to use a nonparametric or data-based estimator of γ^2, as advocated in both Kokoszka and Leipus (2000) and Lee and Park (2001). There are several possibilities in this case. Andreou and Ghysels (2002a), for example, use the autoregression heteroscedasticity and autocorrelation consistent (ARHAC) estimator of den Haan and Levin (1997). In our study, we use the popular Bartlett kernel estimator $\hat{\gamma}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{l} w_{j,l}\hat{\gamma}_j$ where $w_{j,l} = j/(l+1)$, with automatic selection of the truncation lag or bandwidth $l > 0$ using an AR(1) model, as suggested in Andrews (1991).

In sum, each of the three types of assumptions discussed above lead to test statistics based on the same CUSUM process $D_y(k)$. Hence, they share the same limiting distribution under the null hypothesis and under correctness of the underlying assumptions, namely that of a (scaled) Brownian bridge. The only difference between the tests is the use of a different scaling factor or estimate of γ. Specifically, (i) under iid normality: $\hat{\gamma}^2 = 2\hat{\sigma}^4$, (ii) under iid, but not necessarily normality: $\hat{\gamma}^2 = \hat{\gamma}_0$, and (iii) under general dependence and heterogeneity: $\hat{\gamma}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{l} w_{j,l}\hat{\gamma}_j$. In the Monte Carlo simulations reported below we consider all three statistics, which are denoted as $U_{y,\sigma}(k^*)$, $U_{y,\gamma_0}(k^*)$, and $U_{y,\gamma}(k^*)$, respectively.

It is shown in Section 2.3.1 that all tests, including $U_{y,\gamma}(k^*)$, suffer from severe size distortions in finite samples when y_t exhibits conditional heteroscedasticity, in particu-

[2] Alternative approaches to testing for parameter change in GARCH models have recently been developed by Chu (1995), Kokoszka and Teyssière (2002), and Lundbergh and Teräsvirta (2002), among others. A comparison of these tests with the CUSUM statistics considered here is of interest, but is left for future research.

[3] Lee *et al.* (2003) in fact observe that the parametric approach of Kim *et al.* (2000) does not work satisfactorily under all circumstances even when the DGP is a GARCH(1,1) process. In particular, the test suffers from size distortions and low power for certain parameterizations.

lar when y_t follows a GARCH(1,1) process as given in (2.7). Hence, it seems advisable to filter the series first, in order to remove the conditional heteroskedasticity. Interestingly, nonparametric or "model-free" approaches of standardizing the series y_t either with volatility estimates based on high-frequency data (such as quadratic variation) or with Riskmetrics' volatility estimates (obtained as $\hat{h}_t = (1-\lambda)y_{t-1}^2 + \lambda\hat{h}_{t-1}$ with $\lambda = 0.94$) do not work well. In particular, this renders severely undersized test statistics; this corresponds with the findings of Andreou and Ghysels (2003) for CUSUM-type tests in changes in co-movement of conditionally heteroskedastic time series.[4] Here we explore the suggestion of Lee *et al.* (2003) to apply the statistic in (2.5) based on the CUSUM process $D_{\hat{z}}(k) = \frac{1}{\sqrt{T}}\sum_{t=1}^{k}\hat{z}_t^2 - \frac{k}{T\sqrt{T}}\sum_{t=1}^{T}\hat{z}_t^2$ of standardized residuals $\hat{z}_t \equiv y_t/\sqrt{\hat{h}_t}$, where $\hat{h}_t$ is the estimated conditional volatility of y_t obtained from a GARCH(1,1) model estimated with (quasi-)maximum likelihood ((Q)ML) assuming a normal distribution for z_t. The properties of (squared) standardized (G)ARCH residuals have been studied quite intensively in recent years, see Horváth *et al.* (2001), Berkes and Horváth (2003), and Berkes *et al.* (2003), among others. Lee *et al.* (2003) prove that, given the correct conditional volatility specification, the (scaled) CUSUM process $D_{\hat{z}}(k)$ converges to a Brownian bridge, such that the limiting distribution result as given in (2.4) continues to hold. Indeed, in the simulations reported below we find that the associated $U_{\hat{z},\cdot}(k^*)$ statistics have satisfactory size and power properties. One may doubt the practical usefulness of this parametric approach, as the properties of $U_{\hat{z},\cdot}(k^*)$ might be very sensitive to misspecification of the conditional volatility process. We explore this issue in depth in Section 2.3, and find that the CUSUM statistics based on standardized GARCH(1,1)-residuals are in fact remarkably robust to various forms of misspecification.

2.2.2 Testing for multiple structural changes

In the above we focused on testing for a single change in the unconditional variance of y_t. However, there is no reason why the volatility of a time series might not experience multiple changes. Testing for multiple changes in volatility has been addressed in a number of recent articles, including Chen and Gupta (1997) and Lavielle and Moulines (2000). Both studies develop an information criterion based penalized least-squares estimation approach to test for (and date) multiple breaks simultaneously. Similar to the testing framework developed by Bai (1997, 1999) and Bai and Perron (1998, 2003), CUSUM statistics can be applied in

[4]Detailed simulation results demonstrating this result are available upon request.

a sequential manner to test for and identify multiple volatility changes. The basic idea is that first the entire sample is tested for the presence of a single break in volatility using the CUSUM statistics discussed in Section 2.2.1. If a significant change is detected, the sample is split into two segments with the split point being equal to the identified change-point. Next, each subsample is examined separately for a volatility break, again using a CUSUM test. This procedure continues until no more changes are detected in any of the subsamples or until the number of identified changes reaches a pre-specified maximum. Sometimes a final step is added in which all identified breaks are re-evaluated and/or breakpoints re-estimated. In this context, Inclán and Tiao (1994) develop the Iterated Cumulative Sums of Squares (ICSS) algorithm which repeatedly applies their $U_{y,\sigma}(k^*)$ statistic.

We adopt a sequential approach here as well, based on the basic set-up discussed above. Our procedure works as follows. Suppose that at some point in the algorithm N volatility changes have been detected, for $N < M$ with M being the maximum allowed number of breaks. Consequently, the sample for y_t can be split into $N+1$ segments, according to the associated change-point estimates $1 = k_0^* < k_1^* < \ldots < k_N^* < k_{N+1}^* = T$. To test whether any of the segments contains an additional volatility change, we compute one of the CUSUM statistics $U_{\hat{z},\sigma}(k^*)$, $U_{\hat{z},\gamma_0}(k^*)$ or $U_{\hat{z},\gamma}(k^*)$ for each subsample separately,[5] and select the segment for which the test statistic is largest. Suppose this occurs in the i-th segment for $1 \leq i \leq N+1$. If the value of the corresponding CUSUM statistic exceeds an appropriate critical value (see Section 2.2.3), we identify the $(N+1)$-th break in segment i. We repeat this procedure until either N equals M or the maximum of the test statistics across all segments is no longer significant. We control the overall significance level of the sequential procedure by using a significance level of $a/(N+1)$ when testing for the $(N+1)$-th change in volatility. Finally, we re-estimate all change-points, where the location of the i-th volatility change is re-estimated based on the segment determined by the adjacent breakpoints k_{i-1}^* and k_{i+1}^*.[6] This corresponds with the "repartitioning" step in the Bai and Perron (1998) procedure.

Apart from a maximum allowed number of breaks, a second restriction that we impose in the algorithm is that adjacent change-points have to be at least δ observations apart. The latter restriction is to prevent breaks from being identified unrealistically close together. Although the precise value of δ clearly is a subjective decision, we feel that for daily data

[5]Note that this includes estimating separate GARCH(1,1) models for all segments.

[6]For the first volatility change we use the sample from the first observation k_0^* up to and including the second change-point k_2^*. For the last volatility change we use the sample from observation k_{N-1}^* until the last observation k_{N+1}^*.

$\delta = 63$ or 126 business days (three and six months, respectively) seems appropriate. We impose the minimum distance restriction by calculating the maximum absolute value of the CUSUM test statistic in the i-th segment only using the permitted values of k, i.e. $k^*_{i-1} + \delta \leq k \leq k^*_i - \delta$, determined so far in the algorithm. Note that in the final step of the algorithm in which we re-estimate each change-point, we can actually not control the minimum distance between adjacent volatility changes. This would require treating the two adjacent change-points as fixed, whereas these can still be re-estimated at a different location. Hence, it may occur in practice that final breakpoint estimates are less than δ observations apart.

Our procedure as outlined above differs in a number of respects from the ICSS algorithm of Inclán and Tiao (1994). First, after detecting a first volatility change at $k = k^*_1$, the ICSS algorithm examines the first subsample $y_t, t = 1, \ldots, k^*_1$ exhaustively to identify the leftmost significant breakpoint, k^*_l, after which the same is done for the second subsample $y_t, t = k^*_1 + 1, \ldots, T$ to identify the rightmost significant break point k^*_r. If the leftmost break point differs from the rightmost breakpoint, that is $k^*_l < k^*_r$, then the procedure is repeated for the subsamples $t = k^*_l + 1, \ldots, k^*_r$ until $k^*_l = k^*_r$. In our procedure, we consider all N segments when testing for a $(N + 1)$-th break. Second, in the ICSS algorithm the same significance level is applied to each subsample, irrespective of how many breaks have already been found. Third, in the ICSS algorithm breaks can be arbitrarily close to each other, as no minimum distance restriction is imposed. Fourth, in the final step of the ICSS algorithm change-points are not only re-estimated, but the significance of all volatility breaks is determined again, using only the observations from the relevant segment. Earlier detected breaks are removed if they are no longer significant. Finally, and perhaps most important, the ICSS algorithm is based on the $U_{y,\sigma}(k^*)$ statistics, which does not account for possible non-normality and conditional heteroskedasticity.

2.2.3 Finite sample critical values

One issue we have not touched upon so far is the use of critical values. Especially when testing for multiple breaks, the length of the subsamples can quickly become quite small, which renders the use of asymptotic critical values questionable at the least. Therefore, we choose to use finite sample critical values. These are estimated through simulation using the response surface approach described in MacKinnon (2000).[7]

[7] An alternative would be to consider bootstrap procedures for computing critical values or p-values, as in Kokoszka and Teyssière (2002). Given the extent of the Monte Carlo simulations conducted in the next

Suppose that we need the quantile of the distribution of the CUSUM test under the null hypothesis corresponding to a certain significance level a and for a specific finite sample length T, and denote this quantile by $q^a(T)$. This can be obtained by simulating a large number, R, of series of length T from the data-generating process under the null hypothesis and calculating the test statistic for each series. The simulated test statistics can be used to construct the appropriate finite sample distribution and the relevant quantile. Repeating this experiment a total of E times for this specific sample length results in E observations for $q^a(T)$. By repeating this process for different values of T we can then estimate the following type of response surface regressions

$$q_e^a(T) = \theta_\infty^a + \theta_1^a T^{-0.5} + \theta_2^a T^{-1} + \varepsilon_e, \tag{2.8}$$

where $q_e^a(T)$ denotes the quantile estimate obtained in the e-th experiment for sample size T. Subsequently, the estimated response surface regression can be used to determine the appropriate finite sample critical value (quantile) for *any* sample size T. Also note that $\hat{\theta}_\infty^a$ is an estimate of the asymptotic critical value $q^a(\infty)$. The parameter estimates θ_1^a and θ_2^a in our case typically are negative, such that finite sample quantiles are smaller than their asymptotic counterparts. Hence, if asymptotic critical values were used, the tests would appear to be undersized.

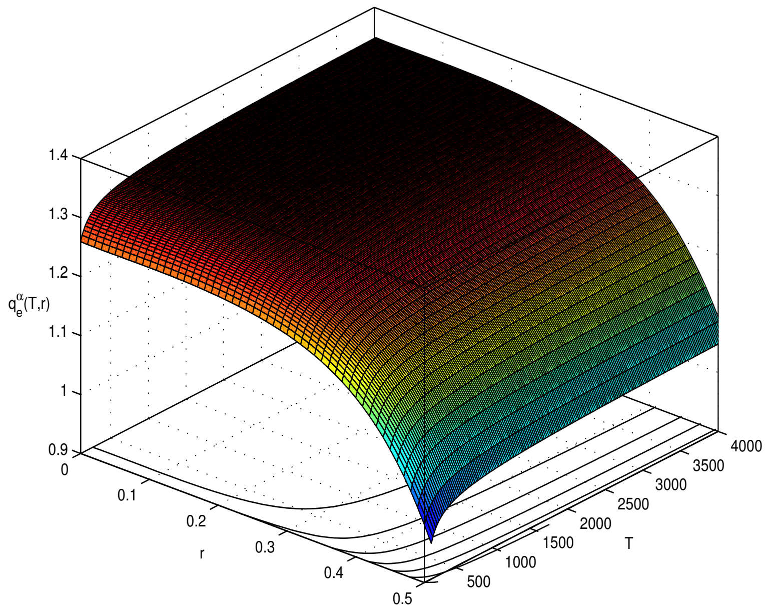
As discussed in the previous section we impose the restriction that two adjacent change-points should be at least δ observations apart, reducing the effective sample size. To account for this we modify the response surface specification by including powers of r, with r being the fraction of observations not considered at either side of the sample when calculating the test statistic, that is $r = \delta/T$. Specifically, we estimate response surface regressions of the form

$$q_e^a(T, r) = \theta_\infty^a + \theta_1^a T^{-0.5} + \theta_2^a T^{-1} + \phi_1^a r + \phi_2^a r^2 + \phi_3^a r^3 + \phi_4^a r^4 + \phi_5^a r^5 + \varepsilon_e. \tag{2.9}$$

To implement the response surface regression, we perform $E = 40$ experiments with $R = 50000$ iid $N(0,1)$ replications each for sample sizes $T_i \in \{50, 60, \ldots, 100, 125, \ldots, 250, 300, 350, \ldots, 500, 600, \ldots, 1000, 1500, \ldots, 3000, 4000, 5000\}$, trimming percentages $r \in \{0.025, 0.05, \ldots, 0.425, 0.45\}$, and for quantiles corresponding to significance levels $a =$ 10%, 5% and 1%. The fit of the response surface (2.9) generally is very good, with R^2 values never being below 97%. To illustrate Figure 2.1 shows the finite sample response surface for the $U_{\hat{z},\sigma}(k^*)$ test for a significance level of 5%. It is seen from the graph that

section, however, the response surface technique is more convenient.

Figure 2.1: Response surface



Note: Response surface for the 95% quantile $q_e^a(T,r)$ of the distribution of $U_{y,\gamma}(k^*)$ test when applied to a sample of length T, discarding a fraction of r observations at both ends of the sample when computing the test statistic.

finite sample critical values differ substantially from their asymptotic counterparts when either T is small or when r becomes close to 0.5.

2.3 Simulation design and results

In this section we report and discuss results from an extensive set of Monte Carlo simulations experiments, designed to examine the small-sample properties of the CUSUM tests and to assess their robustness to various types of misspecification. The (limited amount of) simulation results available in the literature typically only consider the properties of CUSUM tests for data-generating processes (DGPs) that match the assumptions under which a particular test was developed. For example, Inclán and Tiao (1994) evaluate the size and power properties of their test for iid normal series, while Kim *et al.* (2000) and Lee *et al.* (2003) perform simulations using a GARCH(1,1) process with normal shocks z_t as DGP. To some extent, Andreou and Ghysels (2002, 2004) constitute an exception, as they do consider alternative DGPs, namely GARCH(1,1) processes with possibly non-normal errors and with different degrees of volatility persistence, for the $U_{y,\sigma}(k^*)$ and $U_{y,\gamma}(k^*)$ tests. They find that both tests suffer from positive size distortions and their ability to correctly identify breaks varies. Although this provides a reasonable indication of the small sample properties of these tests, we feel that there is scope for broadening these results. In particular, our simulation study has two purposes. First, we assess the small sample properties of the CUSUM tests when applied to standardized GARCH(1,1) residuals $\hat{z}_t$.[8] Second, we examine the robustness of the CUSUM tests to various forms of misspecification, including alternative error distributions as well as misspecification in the conditional variance dynamics.

2.3.1 Size properties

We start our analysis by gauging possible size distortions for the three CUSUM tests. For each DGP discussed below, we generate 10000 replications of length $T = 500$, 1000, 2000 and 4000. We examine rejection frequencies at nominal significance levels $a = 10\%$, 5% and 1%, using finite sample critical values obtained from the response surface regression in (2.9), where we set $r = 0$ throughout. We consider a number of different DGPs where we focus mainly on those that relate to different types of potential misspecification.

First, we consider four DGPs under which the variance of y_t does not have conditional dependence. To be precise, we generated iid series from (i) a standard normal distribution, (ii) a Student-$t(\nu)$ distribution with the degrees of freedom parameter ν ranging from 4 to 8, (iii) a skewed-normal(λ) distribution, see Azzalini (1985), with the skewness parameter λ ranging from -5 (severe negative skewness) to -1 (moderate negative skewness) and finally (iv) normal-with-jumps. Under the latter DGP we add a jump component to the series in such a way that y_t jumps at random points in the sample, but with a fixed and predetermined jump size and jump intensity. Detailed results of these experiments are not shown here to save space but are available upon request. The results can be summarized as follows. When applied to the raw series as well as the standardized residuals from a GARCH(1,1) model, the $U_{\cdot,\sigma}(k^*)$ test is severely oversized for all distributions except the standard normal. This is not surprising given that this test critically depends on the normality assumption for y_t. The $U_{\cdot,\gamma_0}(k^*)$ and $U_{\cdot,\gamma}(k^*)$ tests are almost always correctly

[8] Andreou and Ghysels (2002a) do show results for the $U_{z,\gamma}(k^*)$ test applied to the standardized series $z_t = y_t/\sqrt{h_t}$. However, when constructing the standardized series, the true simulated conditional variance series h_t is used. As a result, they do not take into account the effects of parameter estimation uncertainty and misspecification of the conditional volatility process.

sized, albeit rejection frequencies tend to be somewhat below the nominal significance levels for the Student-t distribution when the number of degrees of freedom is small ($\nu = 4, 5$) and the normal-with-jumps DGP when the jump size is substantial (jumps of 5 or 10 times the standard deviation of the regular component of y_t).

Second, we consider GARCH-type DGPs, such that the variance of y_t exhibits conditional dependence. We first employ a standard GARCH(1,1) model using various combinations of α and β and with different distributions for the errors z_t. We consider the same four distributions as above, albeit for the Student-t distribution we only use $\nu = 5$ and for the skewed normal only $\lambda = -5$. Table 2.1 shows the results of applying the tests to the raw series y_t and to standardized GARCH(1,1) residuals $\hat{z}_t$ for a GARCH(1,1) DGP with normal shocks z_t. The left panel shows that the $U_{y,\sigma}(k^*)$ test is severely oversized.[9] Again, this occurs because the iid normality assumption underlying the test is violated. The $U_{y,\gamma_0}(k^*)$ test is oversized as well, due to the fact that the nonzero (positive) autocorrelations of y_t^2 are not accounted for. What is surprising though is that the $U_{y,\gamma}(k^*)$ test also suffers from substantial positive size distortions, which become larger when conditional volatility is more persistent (see also Table 1 of Andreou and Ghysels (2002a)). Hence, although this CUSUM test theoretically is valid in the presence of heteroscedasticity, as shown in Kokoszka and Leipus (2000), it may require unrealistically large sample sizes for this asymptotic result to apply.[10]

Turning to the right panel, we observe that, when applied to GARCH(1,1) residuals, no substantial size distortions occur for all tests across all parameterizations. Given these results and to facilitate comparison with Andreou and Ghysels (2002a), we only report results for the $U_{\hat{z},\gamma}(k^*)$ test in the remainder of this section. Detailed results for other statistics are available upon request. Table 2.2 reports results for the other three distributions for z_t, showing that the $U_{\hat{z},\gamma}(k^*)$ test statistic is properly sized for each of these. This might be expected of course, given that the normal QML estimator of the parameters in the GARCH(1,1) model is consistent. The unreported results for the other statistics show

[9] At first sight, it may seem odd that size distortions occur despite the use of finite sample critical values from the response surface (2.9). However, the response surface was created assuming a homoskedastic DGP, such that empirical rejection frequencies for heteroskedastic series can still differ from the nominal significance levels. Using asymptotic critical values renders even worse size distortions, given that finite sample critical values are smaller than asymptotic ones.

[10] The size distortion of the $U_{y,\gamma}(k^*)$ does diminish as the sample size increases, such that the empirical rejection frequencies converge to the nominal significance levels, albeit very slowly. Note that for the $U_{y,\sigma}(k^*)$ and $U_{y,\gamma_0}(k^*)$, the empirical size actually becomes worse as T becomes larger.

Table 2.1: Empirical rejection frequencies of CUSUM tests for a single change in volatility when DGP is GARCH(1,1)-N

			Tests applied to "raw" series y_t												Tests applied to standardized GARCH(1,1) residuals											
			T=500			T=1000			T=2000			T=4000			T=500			T=1000			T=2000			T=4000		
α	β	a	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
			$U_{y,\sigma}(k^*)$												$U_{\hat{z},\sigma}(k^*)$											
0.10	0.50		0.265	0.173	0.065	0.278	0.184	0.067	0.290	0.184	0.069	0.296	0.192	0.070	0.070	0.030	0.004	0.083	0.035	0.005	0.089	0.044	0.009	0.094	0.047	0.009
0.10	0.60		0.324	0.219	0.092	0.339	0.236	0.098	0.363	0.246	0.101	0.368	0.254	0.105	0.072	0.029	0.004	0.082	0.036	0.005	0.090	0.043	0.009	0.095	0.046	0.008
0.10	0.70		0.429	0.313	0.158	0.463	0.343	0.174	0.500	0.371	0.182	0.510	0.385	0.192	0.076	0.030	0.003	0.084	0.038	0.006	0.091	0.043	0.009	0.094	0.047	0.008
0.10	0.80		0.655	0.541	0.347	0.724	0.614	0.415	0.781	0.676	0.471	0.816	0.712	0.504	0.081	0.030	0.003	0.085	0.035	0.005	0.090	0.042	0.008	0.094	0.045	0.008
0.20	0.50		0.566	0.445	0.255	0.613	0.493	0.291	0.658	0.538	0.327	0.683	0.557	0.341	0.082	0.034	0.005	0.090	0.040	0.005	0.093	0.045	0.009	0.096	0.047	0.009
0.20	0.60		0.703	0.598	0.399	0.769	0.664	0.467	0.823	0.728	0.531	0.856	0.764	0.566	0.084	0.034	0.003	0.088	0.042	0.006	0.092	0.044	0.009	0.097	0.048	0.008
0.20	0.70		0.877	0.817	0.672	0.934	0.888	0.770	0.971	0.940	0.854	0.985	0.969	0.902	0.085	0.033	0.003	0.088	0.040	0.006	0.091	0.044	0.009	0.095	0.046	0.008
			$U_{y,\gamma_0}(k^*)$												$U_{\hat{z},\gamma_0}(k^*)$											
0.10	0.50		0.245	0.155	0.055	0.256	0.164	0.055	0.266	0.167	0.057	0.270	0.168	0.060	0.074	0.031	0.004	0.084	0.037	0.006	0.092	0.045	0.009	0.095	0.046	0.009
0.10	0.60		0.297	0.194	0.076	0.310	0.210	0.080	0.328	0.217	0.082	0.334	0.224	0.090	0.073	0.029	0.004	0.085	0.038	0.005	0.090	0.044	0.009	0.095	0.046	0.008
0.10	0.70		0.395	0.279	0.127	0.423	0.301	0.145	0.451	0.322	0.150	0.461	0.334	0.156	0.079	0.029	0.003	0.086	0.039	0.005	0.092	0.043	0.008	0.094	0.046	0.007
0.10	0.80		0.611	0.490	0.292	0.668	0.548	0.339	0.722	0.604	0.388	0.750	0.635	0.413	0.084	0.031	0.003	0.086	0.035	0.004	0.091	0.044	0.007	0.096	0.045	0.007
0.20	0.50		0.468	0.344	0.167	0.499	0.374	0.192	0.534	0.404	0.209	0.547	0.416	0.218	0.085	0.034	0.005	0.091	0.042	0.007	0.093	0.046	0.010	0.097	0.047	0.009
0.20	0.60		0.602	0.475	0.276	0.649	0.519	0.315	0.695	0.574	0.351	0.716	0.592	0.370	0.087	0.036	0.004	0.090	0.041	0.007	0.092	0.045	0.009	0.096	0.047	0.007
0.20	0.70		0.800	0.705	0.512	0.857	0.776	0.584	0.902	0.833	0.662	0.932	0.870	0.702	0.088	0.034	0.003	0.090	0.040	0.006	0.093	0.046	0.009	0.096	0.045	0.008
			$U_{y,\gamma}(k^*)$												$U_{\hat{z},\gamma}(k^*)$											
0.10	0.50		0.157	0.088	0.024	0.154	0.084	0.020	0.143	0.078	0.019	0.134	0.073	0.016	0.075	0.032	0.005	0.084	0.037	0.006	0.092	0.046	0.009	0.096	0.047	0.008
0.10	0.60		0.193	0.111	0.035	0.186	0.107	0.029	0.175	0.099	0.026	0.158	0.091	0.023	0.076	0.030	0.005	0.086	0.039	0.006	0.090	0.044	0.009	0.095	0.047	0.008
0.10	0.70		0.264	0.164	0.060	0.251	0.161	0.051	0.240	0.146	0.047	0.219	0.132	0.040	0.081	0.031	0.003	0.087	0.039	0.006	0.093	0.043	0.009	0.094	0.046	0.007
0.10	0.80		0.449	0.327	0.153	0.447	0.324	0.152	0.439	0.310	0.142	0.405	0.284	0.123	0.087	0.032	0.003	0.088	0.035	0.005	0.091	0.044	0.008	0.096	0.045	0.007
0.20	0.50		0.186	0.107	0.030	0.172	0.096	0.024	0.158	0.085	0.021	0.144	0.079	0.018	0.086	0.036	0.006	0.092	0.042	0.007	0.094	0.046	0.010	0.097	0.047	0.008
0.20	0.60		0.260	0.161	0.054	0.231	0.141	0.043	0.211	0.124	0.036	0.184	0.105	0.030	0.089	0.037	0.004	0.090	0.041	0.007	0.091	0.045	0.009	0.096	0.047	0.007
0.20	0.70		0.428	0.306	0.135	0.397	0.270	0.115	0.362	0.239	0.091	0.310	0.200	0.071	0.090	0.035	0.003	0.091	0.040	0.006	0.093	0.046	0.009	0.096	0.045	0.007

Note: Table entries indicate fractions of rejection of the null hypothesis of constant volatility against a single structural change across 10000 replications at nominal significance level a, using finite sample critical values obtained from the response surface (2.9) with $r = 0$. Series of length T are generated from a GARCH(1,1)-N process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1},$$

$z_t \sim \text{iid}\, N(0,1)$, and $\omega = 1 - \alpha - \beta$. Test statistics are applied to the "raw" series y_t (left panel) and to standardized QML residuals $\hat{z}_t$ from a GARCH(1,1)-N model (right panel).

Table 2.2: Empirical rejection frequencies of the $U_{\hat{z},\gamma}(k^*)$ test for a single change in volatility when DGP is GARCH(1,1) with alternative error distributions

			T=500			T=1000			T=2000			T=4000		
α	β	a	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
							GARCH(1,1)-t(5)							
0.10	0.50		0.064	0.027	0.003	0.071	0.028	0.005	0.079	0.035	0.006	0.082	0.042	0.006
0.10	0.60		0.066	0.026	0.003	0.071	0.028	0.005	0.080	0.036	0.005	0.081	0.041	0.006
0.10	0.70		0.066	0.028	0.003	0.074	0.028	0.004	0.079	0.036	0.006	0.080	0.041	0.006
0.10	0.80		0.070	0.028	0.003	0.075	0.030	0.003	0.077	0.035	0.004	0.081	0.040	0.006
0.20	0.50		0.074	0.032	0.004	0.076	0.033	0.005	0.083	0.038	0.007	0.083	0.042	0.006
0.20	0.60		0.075	0.032	0.003	0.078	0.033	0.005	0.083	0.038	0.006	0.083	0.042	0.007
0.20	0.70		0.073	0.030	0.004	0.079	0.032	0.004	0.082	0.038	0.006	0.083	0.042	0.007
							GARCH(1,1)-SN(-5)							
0.10	0.50		0.069	0.027	0.004	0.082	0.038	0.005	0.089	0.043	0.006	0.096	0.046	0.010
0.10	0.60		0.073	0.027	0.004	0.080	0.037	0.005	0.089	0.042	0.006	0.095	0.045	0.010
0.10	0.70		0.075	0.026	0.003	0.083	0.037	0.005	0.090	0.042	0.005	0.093	0.047	0.009
0.10	0.80		0.080	0.029	0.002	0.085	0.034	0.004	0.092	0.040	0.004	0.094	0.046	0.009
0.20	0.50		0.082	0.034	0.005	0.088	0.042	0.006	0.093	0.044	0.006	0.097	0.046	0.010
0.20	0.60		0.082	0.033	0.004	0.088	0.041	0.005	0.093	0.044	0.006	0.095	0.047	0.010
0.20	0.70		0.083	0.030	0.002	0.088	0.038	0.005	0.093	0.043	0.004	0.095	0.046	0.009
							GARCH(1,1)-N with jumps							
0.10	0.50		0.077	0.034	0.005	0.085	0.040	0.007	0.093	0.044	0.007	0.100	0.047	0.008
0.10	0.60		0.080	0.032	0.005	0.088	0.041	0.007	0.093	0.043	0.007	0.100	0.047	0.008
0.10	0.70		0.083	0.034	0.004	0.091	0.043	0.006	0.095	0.044	0.006	0.099	0.048	0.008
0.10	0.80		0.087	0.033	0.004	0.092	0.042	0.005	0.094	0.043	0.005	0.097	0.047	0.007
0.20	0.50		0.090	0.041	0.006	0.094	0.045	0.009	0.096	0.045	0.008	0.101	0.049	0.009
0.20	0.60		0.089	0.041	0.005	0.095	0.045	0.008	0.096	0.046	0.007	0.100	0.049	0.009
0.20	0.70		0.090	0.038	0.003	0.093	0.044	0.006	0.096	0.044	0.006	0.098	0.048	0.008

Note: Table entries indicate fractions of rejection of the null hypothesis of constant volatility against a single structural change across 10000 replications at nominal significance level a, using finite sample critical values obtained from the response surface (2.9) with $r = 0$. Series of length T are generated from (i) a GARCH(1,1)-t(5) process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1},$$

$z_t \sim \mathsf{iid}\, t(5)$ (top panel), (ii) a GARCH(1,1)-$SN(-5)$ process where $z_t \sim \mathsf{iid}\, SN(-5)$ (middle panel) and (iii) a GARCH(1,1)-N-with-jumps process where $y_t = z_t\sqrt{h_t} + \delta D_t$ with $z_t \sim \mathsf{iid}\, N(0,1)$ and D_t is a dummy variable taking the values 1 or -1 (with equal probability) at random time points $t_1, t_2, \ldots, t_{\tau T}$, and 0 otherwise, where $\tau = 0.005$ and $\delta = 5$ (top panel). For all models $\omega = 1 - \alpha - \beta$. The $U_{\hat{z},\gamma}(k^*)$ statistic is applied to standardized QML residuals $\hat{z}$ from a GARCH(1,1)-N model.

that the same holds for the $U_{\hat{z},\gamma_0}(k^*)$, while the $U_{\hat{z},\sigma}(k^*)$ is again oversized due to the non-normality of the shocks z_t (and hence also the standardized residuals $\hat{z}_t$).

We now turn to different types of misspecification of the conditional variance process. We consider first of all the asymmetric GARCH(1,1) process put forward by Glosten *et al.*

(1993), where negative shocks have a different impact on conditional volatility than positive shocks of the same magnitude. The GJR-GARCH(1,1) model is specified as

$$y_t = z_t\sqrt{h_t}, \tag{2.10}$$

$$h_t = \omega + \alpha y_{t-1}^2 + \phi y_{t-1}^2 \mathtt{I}[y_{t-1} < 0] + \beta h_{t-1}, \tag{2.11}$$

where we set $\alpha = 0$ and $\omega = 1 - \gamma/2 - \beta$, such that the unconditional variance of y_t equals 1.

Second, we examine a long-memory fractionally integrated GARCH process (FI-GARCH(1,1)), see Baillie *et al.* (1996),

$$y_t = z_t\sqrt{h_t}, \tag{2.12}$$

$$h_t = \omega + (1 - \beta - (1 - L)^d)y_{t-1}^2 + \beta h_{t-1}, \tag{2.13}$$

where d is the long memory parameter.

Finally, we consider a stochastic volatility (SV-AR(1)) DGP, see Taylor (1986),

$$y_t = z_t \exp(h_t/2) \tag{2.14}$$

$$h_t = \phi_0 + \phi_1 h_{t-1} + \eta_t \tag{2.15}$$

where $\phi_0 = -(1 - \phi_1)/2$, $\eta_t \sim \mathsf{iid}\, N(0, \sigma_\eta^2)$ with $\sigma_\eta^2 = 1 - \phi_1^2$, and z_t and η_t are independent. In all three models above, $z_t \sim \mathsf{iid}\, N(0, 1)$.

Rejection frequencies of the $U_{\hat{z},\gamma}(k^*)$ statistic for the GJR-GARCH(1,1) DGP are quite close to the nominal significance levels used, especially for larger sample sizes $T \geq 2000$; see the upper panel of Table 2.3. By contrast, for the FI-GARCH(1,1) process the test suffers from, sometimes quite severe, positive size distortions, which worsen as the sample size T increases. Apparently the test gets confused when volatility undergoes longer lasting upswings and downswings which are mistakenly considered as structural breaks. Note that this is the reverse phenomenon of mistaking structural changes for long-memory, which has been discussed at considerably length in the literature, see Lamoureux and Lastrapes (1990), Liu (2000), Diebold and Inoue (2001), Franses *et al.* (2002) and Mikosch and Starica (2004), among others. Finally, the test is somewhat conservative for the SV-AR(1,1) DGP, in the sense that empirical rejection frequencies are a bit below the nominal significance levels. Nevertheless, overall the $U_{\hat{z},\gamma}(k^*)$ test appears to be quite robust to various forms of misspecification.

Table 2.3: Empirical rejection frequencies of the $U_{\hat{z},\gamma}(k^*)$ test for a single change in volatility when conditional volatility model is misspecified

			T=500			T=1000			T=2000			T=4000		
γ / d	β	a	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
							GJR-GARCH(1,1)-N							
0.10	0.50		0.066	0.024	0.005	0.072	0.032	0.006	0.089	0.040	0.008	0.091	0.045	0.008
0.10	0.60		0.066	0.025	0.005	0.072	0.032	0.005	0.089	0.041	0.007	0.092	0.043	0.008
0.10	0.70		0.070	0.026	0.004	0.075	0.035	0.006	0.089	0.042	0.008	0.092	0.044	0.007
0.10	0.80		0.078	0.031	0.005	0.083	0.036	0.006	0.092	0.041	0.007	0.089	0.043	0.007
0.20	0.50		0.073	0.029	0.004	0.082	0.039	0.006	0.097	0.045	0.010	0.094	0.048	0.009
0.20	0.60		0.076	0.029	0.004	0.086	0.039	0.007	0.098	0.046	0.010	0.094	0.045	0.008
0.20	0.70		0.082	0.031	0.004	0.089	0.042	0.007	0.099	0.046	0.010	0.094	0.044	0.008
							FI-GARCH(1,1)-N							
0.40	0.10		0.320	0.186	0.040	0.480	0.351	0.136	0.583	0.461	0.243	0.682	0.560	0.342
0.40	0.30		0.203	0.079	0.009	0.328	0.184	0.029	0.417	0.281	0.095	0.493	0.359	0.160
0.60	0.30		0.251	0.123	0.016	0.393	0.256	0.066	0.484	0.355	0.154	0.574	0.446	0.235
0.60	0.50		0.176	0.057	0.002	0.266	0.128	0.014	0.352	0.220	0.057	0.428	0.303	0.128
0.80	0.50		0.169	0.074	0.009	0.258	0.146	0.032	0.314	0.209	0.075	0.377	0.264	0.122
0.80	0.70		0.138	0.043	0.005	0.181	0.077	0.016	0.259	0.157	0.069	0.381	0.292	0.190
							SV-AR(1)							
0.75			0.064	0.027	0.003	0.070	0.030	0.004	0.073	0.030	0.004	0.072	0.033	0.006
0.80			0.062	0.025	0.003	0.068	0.029	0.004	0.067	0.028	0.003	0.070	0.032	0.005
0.85			0.059	0.022	0.002	0.066	0.026	0.003	0.064	0.026	0.004	0.067	0.029	0.005
0.90			0.055	0.020	0.002	0.058	0.023	0.003	0.057	0.023	0.003	0.060	0.025	0.005
0.95			0.047	0.015	0.001	0.048	0.017	0.001	0.046	0.017	0.001	0.048	0.019	0.002
0.975			0.056	0.015	0.002	0.051	0.019	0.004	0.056	0.026	0.009	0.061	0.034	0.010

Note: Table entries indicate fractions of rejection of the null hypothesis of constant volatility against a single structural change across 10000 replications at nominal significance level a, using finite sample critical values obtained from the response surface (2.9) with $r = 0$. Series of length T are generated from
(i) a GJR-GARCH(1,1)-N process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega + \alpha y_{t-1}^2 + \phi y_{t-1}^2 \mathtt{I}[y_{t-1} < 0] + \beta h_{t-1},$$

$\alpha = 0$, and $\omega = 1 - \phi/2 - \beta$ (top panel),
(ii) a FI-GARCH(1,1)-N process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega + (1 - \beta - (1-L)^d)y_{t-1}^2 + \beta h_{t-1},$$

$\omega = 0.10$ (middle panel) and
(iii) a SV-AR(1) process, $y_t = z_t \exp(h_t/2)$, where

$$h_t = \phi_0 + \phi_1 h_{t-1} + \eta_t,$$

$\phi_0 = -(1-\phi_1)/2$, $\eta_t \sim \mathsf{iid}\, N(0, \sigma_\eta^2)$ with $\sigma_\eta^2 = 1 - \phi_1^2$, and z_t and η_t are independent (bottom panel). For all models $z_t \sim \mathsf{iid}\, N(0,1)$. The $U_{\hat{z},\gamma}(k^*)$ statistic is applied to standardized residuals $\hat{z}$ from a GARCH(1,1)-N model.

2.3.2 Power properties

We now turn to the power properties of the $U_{\hat{z},\gamma}(k^*)$ CUSUM test.[11] We first consider the case of a single break in volatility when the DGP is a GARCH(1,1) process with $z_t \sim N(0, \sigma_z^2)$. As the unconditional variance is given by $\sigma_y^2 = \frac{\omega\sigma_z^2}{(1-\alpha-\beta)}$ four potential causes for a variance change can be identified: a break in either ω, σ_z^2, α or β.[12] As the second cause is observationally equivalent to the first, we only consider breaks in the parameters ω, α and β. We allow for three different timings of the parameter change at τT for $\tau = 0.25$, 0.50 and 0.75, again using sample sizes $T = 500$, 1000, 2000 and 4000.

Table 2.4 shows rejection frequencies across 1000 replications from the GARCH(1,1) DGP with a break occurring in ω, where we only consider breaks that occur in the middle of the sample. A number of conclusions can be drawn from this table. These generally also hold true for subsequent tables so we discuss them in somewhat more detail here. First, power increases both with the magnitude of the change in ω (and thus in unconditional volatility) and with the sample size T, except when $\beta = 0.50$ and volatility after the change in ω is very small ($\sigma_a^2 = 0.50$). Second, there appears to exist asymmetry in the test's capability of detecting volatility changes, with volatility increases being picked up better than decreases or vice versa. The direction of the asymmetry depends on the volatility persistence as measured by β. Power is generally higher for volatility decreases for $\beta = 0.80$, whereas for $\beta = 0.50$ it is easier to detect volatility increases. Third, for the smaller sample sizes $T = 500$ and 1000 power is higher for low volatility persistence ($\beta = 0.50$). For the larger sample sizes $T = 2000$ and 4000 this continues to hold for small volatility changes, while large breaks in volatility are easier to detect under high volatility persistence.

Results for a single break in β are shown in Table 2.5. In addition to the increase in power with the magnitude of the change in unconditional volatility and with the sample size T, we observe that decreases in volatility now are easier detected under low volatility persistence ($\beta_b = 0.50$) as well. Furthermore, power is largest for breaks that occur in the middle of the series. For decreases in volatility, early changes ($\tau = 0.25$) are easier to detect than late ones ($\tau = 0.75$) while the reverse holds for volatility increases.[13]

Table 2.6 shows rejection frequencies when a break occurs in α. It is seen that a

[11] Again, detailed results for the other test statistics are available upon request.

[12] It can of course also happen that several parameters change at the same time. We do not consider that possibility here.

[13] The latter is noted and discussed in Inclán and Tiao (1994) by considering the expected value of $D_T(k)$.

Table 2.4: Empirical rejection frequencies of the $U_{\hat{z},\gamma}(k^*)$ test for a single change in volatility when DGP is GARCH(1,1)-N with break in ω

			T=500			T=1000			T=2000			T=4000		
σ_a^2	β	a	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
0.50	0.50		0.659	0.473	0.272	0.654	0.514	0.397	0.630	0.481	0.380	0.542	0.417	0.333
0.60	0.50		0.683	0.497	0.225	0.837	0.741	0.595	0.852	0.793	0.729	0.836	0.793	0.747
0.70	0.50		0.521	0.345	0.118	0.840	0.725	0.450	0.962	0.940	0.883	0.988	0.985	0.980
0.80	0.50		0.265	0.158	0.038	0.519	0.389	0.161	0.813	0.722	0.458	0.976	0.968	0.874
0.90	0.50		0.125	0.050	0.011	0.191	0.109	0.019	0.318	0.231	0.079	0.525	0.392	0.201
1.10	0.50		0.110	0.064	0.007	0.143	0.077	0.021	0.284	0.190	0.070	0.471	0.334	0.156
1.20	0.50		0.219	0.117	0.028	0.390	0.246	0.071	0.651	0.528	0.296	0.921	0.868	0.680
1.30	0.50		0.351	0.224	0.055	0.622	0.471	0.226	0.897	0.832	0.616	0.999	0.995	0.964
1.40	0.50		0.504	0.320	0.107	0.783	0.667	0.398	0.950	0.921	0.849	0.993	0.991	0.984
1.50	0.50		0.598	0.425	0.169	0.861	0.778	0.528	0.931	0.897	0.853	0.966	0.949	0.935
0.50	0.80		0.505	0.252	0.053	0.874	0.716	0.293	0.996	0.987	0.916	1.000	1.000	1.000
0.60	0.80		0.426	0.212	0.041	0.776	0.608	0.196	0.981	0.956	0.794	1.000	1.000	0.998
0.70	0.80		0.271	0.117	0.013	0.540	0.351	0.086	0.835	0.724	0.443	0.988	0.977	0.891
0.80	0.80		0.166	0.064	0.010	0.283	0.164	0.022	0.495	0.352	0.150	0.779	0.673	0.407
0.90	0.80		0.098	0.034	0.003	0.129	0.053	0.004	0.207	0.122	0.033	0.296	0.202	0.061
1.10	0.80		0.095	0.044	0.007	0.094	0.049	0.006	0.187	0.111	0.024	0.258	0.168	0.052
1.20	0.80		0.140	0.055	0.010	0.206	0.090	0.015	0.382	0.248	0.085	0.627	0.511	0.249
1.30	0.80		0.208	0.099	0.016	0.345	0.202	0.031	0.601	0.467	0.213	0.891	0.810	0.594
1.40	0.80		0.263	0.129	0.017	0.484	0.314	0.063	0.796	0.680	0.380	0.976	0.960	0.859
1.50	0.80		0.342	0.161	0.026	0.617	0.432	0.114	0.908	0.844	0.550	0.999	0.996	0.970

Note: Table entries indicate fractions of rejection of the null hypothesis of constant volatility against a single structural change across 1000 replications at nominal significance level a, using finite sample critical values obtained from the response surface (2.9) with $r = 0$. Series of length T are generated from a GARCH(1,1)-N process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega_t + \alpha y_{t-1}^2 + \beta h_{t-1},$$

$z_t \sim \text{iid } N(0,1)$, $\alpha = 0.10$, $\omega_t = \omega_b = 1 - \alpha - \beta$ if $t \leq \tau T$ and $\omega_t = \sigma_a^2\omega_b$ if $t > \tau T$, with $\tau = 0.50$. The $U_{\hat{z},\gamma}(k^*)$ statistic is applied to standardized residuals $\hat{z}$ from a GARCH(1,1)-N model.

break in α is generally more difficult to detect than a break in either ω or β leading to the same change in unconditional volatility. Only for substantial changes and only when volatility is lower after the change does the power of the test seem reasonable. Again, when volatility increases, power is slightly better for breaks occurring late in the sample compared to early breaks while the reverse holds for decreases in volatility. Finally, the level of volatility persistence, reflected by β, is of influence with the rejection frequencies being higher for low persistence as compared to high persistence.

We also considered breaks in ω, α and β with different distributions for the shocks z_t, including Student-$t(\nu)$ and skewed-normal(λ). Typically, power goes down somewhat

Table 2.5: Empirical rejection frequencies of the $U_{\hat{z},\gamma}(k^*)$ test for a single change in volatility when DGP is GARCH(1,1)-N with break in β

				T=500			T=1000			T=2000			T=4000		
σ_a^2	β_b	τ	a	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
0.60	0.50	0.25		0.535	0.358	0.135	0.799	0.713	0.498	0.876	0.814	0.746	0.897	0.827	0.784
0.60	0.50	0.50		0.698	0.528	0.262	0.833	0.738	0.620	0.832	0.764	0.698	0.800	0.733	0.693
0.60	0.50	0.75		0.487	0.298	0.064	0.826	0.697	0.389	0.912	0.881	0.822	0.935	0.922	0.915
0.80	0.50	0.25		0.185	0.077	0.014	0.351	0.248	0.065	0.637	0.513	0.255	0.909	0.845	0.649
0.80	0.50	0.50		0.263	0.155	0.036	0.522	0.396	0.161	0.811	0.725	0.467	0.978	0.970	0.878
0.80	0.50	0.75		0.156	0.071	0.012	0.317	0.186	0.043	0.560	0.438	0.181	0.885	0.788	0.530
1.20	0.50	0.25		0.133	0.067	0.006	0.201	0.104	0.021	0.414	0.287	0.101	0.692	0.554	0.298
1.20	0.50	0.50		0.213	0.113	0.028	0.377	0.233	0.063	0.629	0.500	0.269	0.908	0.838	0.646
1.20	0.50	0.75		0.146	0.084	0.016	0.207	0.129	0.035	0.453	0.323	0.129	0.740	0.644	0.358
1.40	0.50	0.25		0.259	0.139	0.030	0.502	0.333	0.093	0.853	0.744	0.428	0.995	0.988	0.928
1.40	0.50	0.50		0.451	0.292	0.081	0.756	0.634	0.344	0.947	0.915	0.813	0.997	0.994	0.991
1.40	0.50	0.75		0.290	0.163	0.038	0.561	0.400	0.144	0.877	0.783	0.542	0.996	0.993	0.962
0.60	0.80	0.25		0.299	0.135	0.023	0.640	0.440	0.098	0.942	0.867	0.583	1.000	0.998	0.978
0.60	0.80	0.50		0.460	0.257	0.041	0.832	0.661	0.278	0.989	0.975	0.872	1.000	1.000	1.000
0.60	0.80	0.75		0.235	0.093	0.006	0.533	0.315	0.070	0.902	0.790	0.423	0.999	0.998	0.960
0.80	0.80	0.25		0.116	0.045	0.006	0.201	0.093	0.005	0.341	0.224	0.071	0.578	0.441	0.213
0.80	0.80	0.50		0.162	0.058	0.011	0.281	0.159	0.018	0.486	0.349	0.148	0.783	0.674	0.407
0.80	0.80	0.75		0.100	0.031	0.003	0.165	0.082	0.007	0.310	0.193	0.043	0.529	0.377	0.136
1.20	0.80	0.25		0.105	0.043	0.003	0.097	0.048	0.004	0.214	0.119	0.029	0.344	0.215	0.060
1.20	0.80	0.50		0.126	0.052	0.006	0.175	0.080	0.012	0.332	0.208	0.060	0.569	0.429	0.183
1.20	0.80	0.75		0.100	0.043	0.005	0.110	0.056	0.006	0.241	0.131	0.033	0.372	0.252	0.085
1.40	0.80	0.25		0.140	0.049	0.003	0.220	0.085	0.011	0.446	0.278	0.075	0.768	0.617	0.292
1.40	0.80	0.50		0.229	0.104	0.010	0.401	0.226	0.035	0.678	0.535	0.248	0.946	0.890	0.697
1.40	0.80	0.75		0.136	0.057	0.007	0.232	0.112	0.019	0.466	0.329	0.097	0.787	0.692	0.393

Note: Table entries indicate fractions of rejection of the null hypothesis of constant volatility against a single structural change across 1000 replications at nominal significance level a, using finite sample critical values obtained from the response surface (2.9) with $r = 0$. Series of length T are generated from a GARCH(1,1)-N process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega + \alpha y_{t-1}^2 + \beta_t h_{t-1},$$

$z_t \sim$ iid $N(0,1)$, $\beta_t = \beta_b$ if $t \le \tau T$ and $\beta_t = \beta_a$ if $t > \tau T$, where β_a is such that the unconditional volatility after the break is equal to $(\omega/(1-\alpha-\beta_a) =)\ \sigma_a^2$, $\alpha = 0.10$, and $\omega = 1-\alpha-\beta_b$ such that the unconditional volatility before the break is equal to 1. The $U_{\hat{z},\gamma}(k^*)$ statistic is applied to standardized residuals $\hat{z}$ from a GARCH(1,1)-N model.

compared to DGPs with z_t being normally distributed. The same occurs when the DGP is a GJR-GARCH(1,1) process with a break in either ω or β. Detailed results for these experiments are available upon request.

Given that we wish to apply the CUSUM tests in the sequential procedure for multiple volatility changes, as described in Section 2.2.2, we consider their power in detecting mul-

Table 2.6: Empirical rejection frequencies of the $U_{\hat{z},\gamma}(k^*)$ test for a single change in volatility when DGP is GARCH(1,1)-N with break in α

				T=500			T=1000			T=2000			T=4000		
σ_a^2	β	τ	a	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010	0.100	0.050	0.010
0.60	0.70	0.25		0.326	0.164	0.026	0.683	0.506	0.167	0.935	0.870	0.650	0.992	0.981	0.954
0.60	0.70	0.50		0.502	0.299	0.053	0.845	0.717	0.382	0.990	0.977	0.891	0.999	0.998	0.994
0.60	0.70	0.75		0.248	0.105	0.011	0.541	0.353	0.078	0.893	0.780	0.448	0.998	0.995	0.961
0.80	0.50	0.25		0.146	0.060	0.011	0.276	0.192	0.041	0.529	0.384	0.176	0.812	0.717	0.481
0.80	0.50	0.50		0.194	0.113	0.020	0.406	0.282	0.093	0.690	0.546	0.310	0.935	0.871	0.716
0.80	0.50	0.75		0.130	0.054	0.006	0.228	0.123	0.026	0.418	0.292	0.097	0.711	0.579	0.283
0.80	0.70	0.25		0.119	0.042	0.005	0.200	0.090	0.009	0.322	0.206	0.078	0.543	0.415	0.185
0.80	0.70	0.50		0.143	0.061	0.007	0.260	0.149	0.025	0.435	0.302	0.130	0.715	0.597	0.331
0.80	0.70	0.75		0.104	0.029	0.001	0.162	0.073	0.008	0.268	0.159	0.040	0.447	0.308	0.101
1.20	0.50	0.25		0.105	0.066	0.005	0.128	0.071	0.014	0.251	0.142	0.044	0.386	0.254	0.096
1.20	0.50	0.50		0.150	0.083	0.014	0.207	0.114	0.037	0.390	0.278	0.108	0.629	0.520	0.263
1.20	0.50	0.75		0.117	0.065	0.011	0.152	0.085	0.024	0.289	0.188	0.061	0.453	0.333	0.141
1.20	0.70	0.25		0.088	0.043	0.006	0.100	0.052	0.009	0.170	0.108	0.028	0.247	0.152	0.040
1.20	0.70	0.50		0.112	0.058	0.007	0.140	0.078	0.015	0.250	0.168	0.058	0.423	0.294	0.122
1.20	0.70	0.75		0.098	0.050	0.010	0.106	0.060	0.010	0.209	0.116	0.031	0.306	0.201	0.059
1.40	0.50	0.25		0.163	0.089	0.014	0.239	0.129	0.029	0.474	0.345	0.124	0.785	0.650	0.371
1.40	0.50	0.50		0.275	0.158	0.042	0.455	0.321	0.108	0.726	0.618	0.370	0.962	0.923	0.780
1.40	0.50	0.75		0.194	0.114	0.023	0.305	0.208	0.060	0.567	0.449	0.219	0.852	0.768	0.572
1.40	0.70	0.25		0.119	0.066	0.007	0.157	0.082	0.014	0.315	0.191	0.059	0.529	0.395	0.151
1.40	0.70	0.50		0.178	0.086	0.014	0.282	0.165	0.042	0.508	0.387	0.168	0.785	0.693	0.456
1.40	0.70	0.75		0.128	0.069	0.012	0.191	0.110	0.019	0.366	0.251	0.081	0.638	0.492	0.250

Note: Table entries indicate fractions of rejection of the null hypothesis of constant volatility against a single structural change across 1000 replications at nominal significance level a, using finite sample critical values obtained from the response surface (2.9) with $r = 0$. Series of length T are generated from a GARCH(1,1)-N process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega + \alpha_t y_{t-1}^2 + \beta h_{t-1},$$

$z_t \sim \text{iid}\, N(0,1)$, $\alpha_t = \alpha_b = 0.15$ if $t \leq \tau T$ and $\alpha_t = \alpha_a$ if $t > \tau T$, where α_a is such that the unconditional volatility after the break is equal to $(\omega/(1-\alpha_a-\beta) =)\ \sigma_a^2$, and $\omega = 1-\alpha_b-\beta$ such that the unconditional volatility before the break is equal to 1. The $U_{\hat{z},\gamma}(k^*)$ statistic is applied to standardized residuals $\hat{z}$ from a GARCH(1,1)-N model.

tiple breaks in the GARCH parameters. To conserve space we only discuss a GARCH(1,1) DGP with two breaks in either ω or β. Results are reported in Tables 2.7 and 2.8, respectively. First of all we observe that power is reasonable when volatility first goes down and then jumps up again (or the reverse), but in such a way that it does not return to its initial level. For low volatility persistence the test detects the two breaks quite well. However, for high persistence, power is considerably lower. When volatility does return to its original level after the second change, the test typically identifies no breaks at all. In the latter

Table 2.7: Number of identified change points for $U_{\hat{z},\gamma}(k^*)$ test for changes in volatility when the DGP is GARCH(1,1)-N with two breaks in ω

				T=1000				T=2000				T=4000			
σ^2_{a1}	σ^2_{a2}	β	l	0	1	2	≥ 3	0	1	2	≥ 3	0	1	2	≥ 3
0.70	0.40	0.50		0.627	0.282	0.088	0.003	0.701	0.086	0.200	0.013	0.746	0.004	0.180	0.070
0.80	0.60	0.50		0.245	0.725	0.030	0.000	0.101	0.727	0.163	0.009	0.063	0.327	0.551	0.059
0.90	0.80	0.50		0.724	0.260	0.016	0.000	0.482	0.493	0.025	0.000	0.160	0.790	0.049	0.001
0.80	1.00	0.50		0.915	0.070	0.015	0.000	0.782	0.099	0.115	0.004	0.528	0.064	0.385	0.023
0.80	1.20	0.50		0.639	0.266	0.095	0.000	0.256	0.330	0.403	0.011	0.022	0.111	0.824	0.043
1.20	0.80	0.50		0.570	0.318	0.112	0.000	0.167	0.458	0.365	0.010	0.010	0.290	0.670	0.030
1.20	1.00	0.50		0.938	0.048	0.014	0.000	0.850	0.081	0.068	0.001	0.722	0.084	0.182	0.012
1.10	1.20	0.50		0.835	0.155	0.010	0.000	0.616	0.371	0.013	0.000	0.320	0.650	0.028	0.002
1.20	1.40	0.50		0.494	0.489	0.017	0.000	0.150	0.820	0.029	0.001	0.009	0.836	0.144	0.011
1.30	1.60	0.50		0.277	0.700	0.023	0.000	0.079	0.806	0.111	0.004	0.042	0.474	0.439	0.045
0.70	0.40	0.80		0.377	0.567	0.055	0.001	0.036	0.703	0.252	0.009	0.000	0.218	0.715	0.067
0.80	0.60	0.80		0.556	0.424	0.020	0.000	0.166	0.783	0.049	0.002	0.006	0.839	0.138	0.017
0.90	0.80	0.80		0.881	0.106	0.012	0.001	0.742	0.243	0.014	0.001	0.502	0.470	0.028	0.000
0.80	1.00	0.80		0.959	0.034	0.007	0.000	0.910	0.060	0.029	0.001	0.819	0.083	0.093	0.005
0.80	1.20	0.80		0.886	0.096	0.018	0.000	0.639	0.244	0.113	0.004	0.264	0.293	0.415	0.028
1.20	0.80	0.80		0.834	0.146	0.020	0.000	0.581	0.315	0.098	0.006	0.199	0.495	0.302	0.004
1.20	1.00	0.80		0.964	0.032	0.004	0.000	0.919	0.068	0.013	0.000	0.886	0.073	0.040	0.001
1.10	1.20	0.80		0.933	0.065	0.002	0.000	0.812	0.181	0.007	0.000	0.638	0.334	0.025	0.003
1.20	1.40	0.80		0.782	0.211	0.007	0.000	0.509	0.481	0.010	0.000	0.170	0.780	0.046	0.004
1.30	1.60	0.80		0.605	0.378	0.017	0.000	0.220	0.756	0.021	0.003	0.021	0.862	0.110	0.007

Note: Table entries indicate fractions of replications for which l structural changes in volatility were found across 1000 replications using the sequential procedure described in Section 2.2.2. Series are generated from a GARCH(1,1)-N process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega + \alpha y^2_{t-1} + \beta_t h_{t-1},$$

$z_t \sim \text{iid } N(0,1)$, $\alpha = 0.1$, $\omega_t = \omega_b = 1 - \alpha - \beta$ if $t \leq \tau_1 T$, $\omega_t = \sigma^2_{a1}\omega_b$ if $\tau_1 T \leq t \leq \tau_2 T$ and $\omega_t = \sigma^2_{a2}\omega_b$ if $t \geq \tau_2 T$ with $\tau_1 = 0.33$ and $\tau_1 = 0.67$. The $U_{\hat{z},\gamma}(k^*)$ statistic is applied to standardized residuals $\hat{z}$ from a GARCH(1,1)-N model. Finite sample critical values are obtained from the response surface (2.9) with $r = 0.15$, for initial nominal significance level $a = 0.05$.

case, the test only starts to pick up the changes in volatility when the series is fairly long ($T = 4000$). For a stepwise decrease in volatility from 1 to 0.7 to 0.4 we see that the test has difficulty in picking up the correct number of breaks, with two breaks being detected in only 20% of the cases for moderate samples sizes. The level of persistence again matters. Focusing for ease of discussion on $T = 2000$, we see that for $\beta = 0.50$ the test typically identifies no breaks at all, whereas for $\beta = 0.80$ a single change is identified in 70% of the replications. For a smaller step size (1 to 0.9 to 0.8) the pictures changes. Now for low volatility persistence either 0 or 1 break is identified equally frequent, whereas for high

Table 2.8: Number of identified change points for the $U_{\hat{z},\gamma}(k^*)$ test for changes in volatility when the DGP is GARCH(1,1)-N with two breaks in β

				T=1000				T=2000				T=4000			
σ^2_{a1}	σ^2_{a2}	β	l	0	1	2	≥ 3	0	1	2	≥ 3	0	1	2	≥ 3
0.70	0.40	0.50		0.666	0.241	0.093	0.000	0.789	0.059	0.141	0.011	0.775	0.007	0.149	0.069
0.80	0.60	0.50		0.227	0.735	0.038	0.000	0.123	0.689	0.174	0.014	0.092	0.285	0.560	0.063
0.90	0.80	0.50		0.727	0.256	0.017	0.000	0.473	0.500	0.027	0.000	0.156	0.789	0.053	0.002
0.80	1.00	0.50		0.911	0.073	0.016	0.000	0.779	0.101	0.116	0.004	0.521	0.063	0.394	0.022
0.80	1.20	0.50		0.642	0.266	0.092	0.000	0.272	0.322	0.398	0.008	0.020	0.110	0.826	0.044
1.20	0.80	0.50		0.575	0.316	0.109	0.000	0.172	0.470	0.349	0.009	0.009	0.311	0.652	0.028
1.20	1.00	0.50		0.941	0.046	0.013	0.000	0.863	0.075	0.061	0.001	0.735	0.078	0.177	0.010
1.10	1.20	0.50		0.849	0.139	0.012	0.000	0.633	0.354	0.013	0.000	0.341	0.632	0.025	0.002
1.20	1.40	0.50		0.536	0.449	0.015	0.000	0.186	0.782	0.031	0.001	0.014	0.864	0.113	0.009
1.30	1.60	0.50		0.319	0.652	0.029	0.000	0.063	0.851	0.082	0.004	0.007	0.579	0.381	0.033
0.70	0.40	0.80		0.295	0.614	0.087	0.004	0.075	0.582	0.333	0.010	0.011	0.150	0.725	0.114
0.80	0.60	0.80		0.504	0.476	0.020	0.000	0.123	0.822	0.051	0.004	0.002	0.792	0.188	0.018
0.90	0.80	0.80		0.877	0.112	0.011	0.000	0.753	0.234	0.012	0.001	0.511	0.464	0.025	0.000
0.80	1.00	0.80		0.959	0.033	0.008	0.000	0.902	0.071	0.026	0.001	0.821	0.087	0.087	0.005
0.80	1.20	0.80		0.889	0.095	0.016	0.000	0.670	0.235	0.091	0.004	0.314	0.289	0.378	0.019
1.20	0.80	0.80		0.850	0.129	0.021	0.000	0.629	0.292	0.077	0.002	0.240	0.518	0.239	0.003
1.20	1.00	0.80		0.965	0.031	0.004	0.000	0.928	0.062	0.010	0.000	0.906	0.062	0.031	0.001
1.10	1.20	0.80		0.938	0.058	0.004	0.000	0.838	0.156	0.006	0.000	0.710	0.267	0.021	0.002
1.20	1.40	0.80		0.866	0.129	0.005	0.000	0.630	0.359	0.011	0.000	0.297	0.659	0.042	0.002
1.30	1.60	0.80		0.733	0.255	0.012	0.000	0.390	0.595	0.014	0.001	0.062	0.870	0.063	0.005

Note: Table entries indicate fractions of replications for which l structural changes in volatility were found across 1000 replications using the sequential procedure described in Section 2.2.2. Series are generated from a GARCH(1,1)-N process, $y_t = z_t\sqrt{h_t}$, where

$$h_t = \omega + \alpha y^2_{t-1} + \beta_t h_{t-1},$$

$z_t \sim$ iid $N(0,1)$, $\beta_t = \beta_b$ if $t \leq \tau_1 T$, $\beta_t = \beta_{a_1}$ if $\tau_1 T < t \leq \tau_2 T$ and $\beta_t = \beta_{a_2}$ if $\tau_2 T < t$, $\alpha = 0.10$, and $\omega = 1 - \alpha - \beta_b$ with $\tau_1 = 0.33$ and $\tau_1 = 0.67$. β_{a_1} and β_{a_2} are such that the unconditional volatility $\omega/(1 - \alpha - \beta_{a_i}) = \sigma^2_{a_i}, i = 1, 2$. The $U_{\hat{z},\gamma}(k^*)$ statistic is applied to standardized residuals $\hat{z}$ from a GARCH(1,1)-N model. Finite sample critical values are obtained from the response surface (2.9) with $r = 0.15$, for initial nominal significance level $a = 0.05$.

persistence zero breaks are identified most often (74%). The same result hold for small increases (1 to 1.1 to 1.2). For larger stepwise increases, the null hypothesis of no breaks is increasingly often rejected, but the ability to find two breaks is still quite low. Similar to the single break DGPs, power is higher for decreases in volatility compared to increases in volatility.

2.4 Volatility changes in emerging stock markets

Research in emerging market finance has been rapidly expanding over the past two decades, see Bekaert and Harvey (2002, 2003) for comprehensive surveys of the past, present and future of this area. Among others, this growing interest stems from the fact that emerging market assets have become increasingly important in international investment portfolios. This has been made possible thanks to the capital market liberalizations many of these countries experienced in the late 1980s and early 1990s. Typically, the liberalization measures that were implemented included substantial reduction or even complete removal of barriers-to-entry for foreign investors. The financial and macroeconomic effects of these liberalizations have been studied intensively, see Bekaert and Harvey (2003) for an overview. The consequences of liberalization and the resulting (or at least hoped-for) integration with developed markets for stock market volatility are not clear *a priori*. A common perception is that the growing influence of highly mobile foreign capital (so-called "hot money") might lead to higher volatility in liberalized markets. However, empirical studies have found little support for this hypothesis, but instead document either no significant changes or declines in stock market volatility following liberalizations, see Richards (1996), Bekaert and Harvey (1997, 2000a), De Santis and Imrohoroglu (1997), Aggarwal *et al.* (1999) and Kim and Singal (2000), and the references contained therein. The finding of lower post-liberalization volatility typically is attributed to increased market efficiency and diversification effects.

Analyzing the effect of liberalizations on stock market volatility is not without complications, however. First, in most countries liberalization has been a gradual process, with different measures taken at different points in time. Second, emerging markets volatility may change for a host of reasons other than financial liberalization, including (both local and global) social, political or economic events. Consequently, multiple sudden and substantial changes in volatility may be observed in these markets. In this section we aim to identify volatility changes in emerging stock market index returns by means of the CUSUM tests. Our empirical study resembles that of Aggarwal *et al.* (1999), although they analyze a considerably smaller set of countries over a shorter sample period and only use the original Inclán-Tiao CUSUM-statistic $U_{y,\sigma}(k^*)$. The latter difference may be most crucial as the $U_{y,\sigma}(k^*)$ statistic does not account for possible non-normality and conditional heteroskedasticity, which are relevant characteristics of emerging stock market returns, see Bekaert and Harvey (1997) and Bekaert *et al.* (1998), among others.

We examine daily returns on MSCI indexes for a total of 27 emerging stock markets.

We select countries from each of the three emerging market clusters identified by MSCI: China, India, Indonesia, Korea, Malaysia, Pakistan, Philippines, Sri Lanka, Taiwan and Thailand (Asia), Argentina, Brazil, Chile, Colombia, Mexico, Peru and Venezuela (Latin America) and Czech Republic, Egypt, Hungary, Israel, Jordan, Morocco, Poland, Russia, South Africa and Turkey (Europe, Middle East and Africa).[14] The sample period runs from January 1, 1988 to December 31, 2003, resulting in a total of 4173 daily return observations, although not all series start on January 1, 1988. The second column of Table 2.9 shows the starting date of the returns series for each country. The countries with the shortest samples (Czech Republic, Egypt, Hungary, Morocco and Russia) still have over 2000 observations. Following Aggarwal *et al.* (1999), we consider returns measured in U.S. dollars as well as in local currency. Unreported summary statistics confirm the importance of non-normality (in the form of significant skewness, excess kurtosis and infrequent large jumps, both positive and negative) and conditional heteroskedasticity for these stock return series.

We start with the original Inclán-Tiao ICSS algorithm for detecting and dating multiple breaks in the unconditional volatility of demeaned returns.[15] Columns four and 11 of Table 2.9 show the number of breaks thus identified by the $U_{y,\sigma}(k^*)$ test, indicating an unrealistically large number of volatility changes. Furthermore, sometimes the identified change-points are only a few weeks or even days apart. It is therefore hardly justifiable to classify these as genuine shifts in the level of volatility.

We proceed with our sequential testing algorithm as described in Section 2.2.2, allowing for a maximum number of 10 breaks, which each have to be at least 126 (trading) days apart[16]. Appropriate finite sample critical values are obtained from (2.9), using an initial nominal significance level $a = 0.05$. We implement the algorithm with each of the three CUSUM statistics $U_{y,\sigma}(k^*)$, $U_{y,\gamma_0}(k^*)$ and $U_{y,\gamma}(k^*)$ to the demeaned returns. The number of detected volatility changes is drastically reduced, as shown by the results in columns 5-7 for US dollar returns and columns 12-14 for local currency returns. The $U_{y,\sigma}(k^*)$ test still often identifies the maximum number of 10 breaks, which is due to the fact that it cannot account for the non-normality of the return series. The $U_{y,\gamma_0}(k^*)$ test, which scales down

[14]In the remainder of the analysis we do not need to differentiate between countries from different regions. Consequently, countries are given in alphabetical order in the table and graphs below.

[15]We used the Schwarz Information Criterion (SIC) to determine the optimal lag order p in an autoregressive (AR(p)) model for each return series. It turned out that in general a lag order of $p = 0$ was selected and consequently we did not use any autoregressive lags when demeaning the return series.

[16]We also considered a minimum distance of 63 business days which resulted in the same number of change points for all but a few countries.

Table 2.9: Number of identified break points for emerging stock market returns

	Start		US Dollar returns							local currency returns						
				y_t			$\hat{z}_t$				y_t			$\hat{z}_t$		
	date	T	ICSS	σ	γ_0	γ	σ	γ_0	γ	ICSS	σ	γ_0	γ	σ	γ_0	γ
Argentina	1/1/1988	4007	36	10	10	4	2	0	0	36	10	10	1	2	0	0
Brazil	1/1/1988	4006	16	10	6	4	0	0	0	19	9	7	5	1	0	0
Chile	1/1/1988	4035	25	10	10	1	4	1	1	27	10	3	1	3	1	1
China	1/1/1993	2835	19	9	4	3	3	0	0	19	9	4	3	3	0	0
Colombia	1/1/1993	2739	24	8	2	2	3	0	0	20	7	3	2	3	0	0
Czech Republic	1/2/1995	2325	9	7	7	5	1	0	0	16	7	7	5	1	0	0
Egypt	1/2/1995	1945	18	8	7	2	2	0	0	23	8	7	3	1	0	0
Hungary	1/2/1995	2326	14	6	3	3	4	0	0	23	8	6	4	5	0	0
India	1/1/1993	2790	13	9	6	4	2	0	0	14	10	6	4	2	0	0
Indonesia	1/1/1988	4017	38	10	4	3	8	3	3	48	10	0	0	6	4	4
Israel	1/1/1993	2851	19	10	9	4	4	0	0	21	10	7	6	4	0	0
Jordan	1/1/1988	2958	36	8	0	0	3	0	0	29	9	1	0	4	0	0
Korea	1/1/1988	4012	12	9	7	7	3	2	2	15	8	6	6	3	1	1
Malaysia	1/1/1988	4069	40	10	5	4	5	0	0	30	10	4	4	7	0	0
Mexico	1/1/1988	4078	24	10	6	3	7	0	0	22	10	6	0	4	0	0
Morocco	1/2/1995	2334	20	4	2	2	3	1	1	21	9	5	1	2	1	1
Pakistan	1/1/1993	2506	25	9	4	4	3	0	0	25	10	4	3	0	0	0
Peru	1/1/1993	2785	26	10	6	1	4	1	1	32	10	2	1	2	1	1
Philippines	1/1/1988	4043	32	10	7	4	6	0	0	32	10	8	6	6	0	0
Poland	1/1/1993	2822	9	9	8	5	3	0	0	10	8	8	6	3	0	0
Russia	1/2/1995	2288	21	10	7	4	2	2	2	21	10	7	4	2	2	2
South Africa	1/1/1993	2839	19	10	7	3	4	0	0	19	7	3	3	2	2	2
Sri Lanka	1/1/1993	2679	36	9	6	4	7	0	0	33	10	6	1	8	0	0
Taiwan	1/1/1988	4006	25	10	6	4	2	0	0	18	10	6	4	2	0	0
Thailand	1/1/1988	4068	36	10	10	3	4	3	3	40	10	10	3	4	3	3
Turkey	1/1/1988	4057	30	10	10	8	2	1	1	30	10	8	3	3	1	1
Venezuela	1/1/1993	2731	31	9	0	0	7	0	0	25	10	0	0	6	0	0

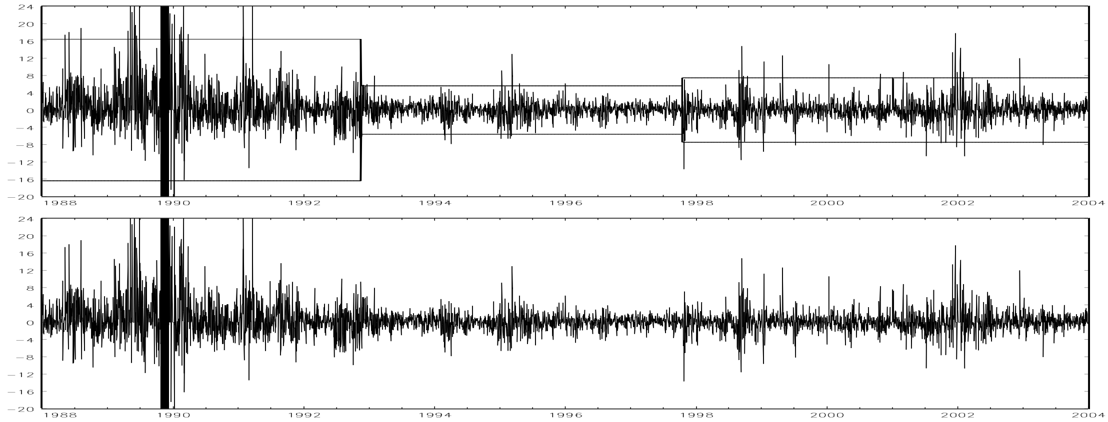
Note: Table entries represent the number of identified breaks in volatility of daily emerging stock market returns, both in U.S. dollars and local currency. Results are reported for the original ICSS algorithm of Inclán and Tiao (1994), and for the sequential procedure described in Section 2.2.2 using the $U_{\cdot,\sigma}(k^*)$, $U_{\cdot,\gamma_0}(k^*)$, and $U_{\cdot,\gamma}(k^*)$ statistics (in columns headed σ, γ_0 and γ, respectively), applied to demeaned returns y_t as well as GARCH(1,1)-standardized returns $\hat{z}_t$. The maximum allowed number of breaks is set to 10, while consecutive breaks have to be at least 126 trading days apart. Finite sample critical values obtained from (2.9), with an initial nominal significance level $a = 0.05$.

the centered sum of squares with the variance of the squared returns series, only hits the upper bound for four countries (Argentina, Chile, Thailand and Turkey) for the US dollar returns and for only two counties (Argentina and Thailand) for the local currency returns. The $U_{y,\gamma}(k^*)$ statistic on the other hand is never constrained by the imposed maximum number of breaks. The maximum number of identified breaks across all returns series is 8, for Turkey.

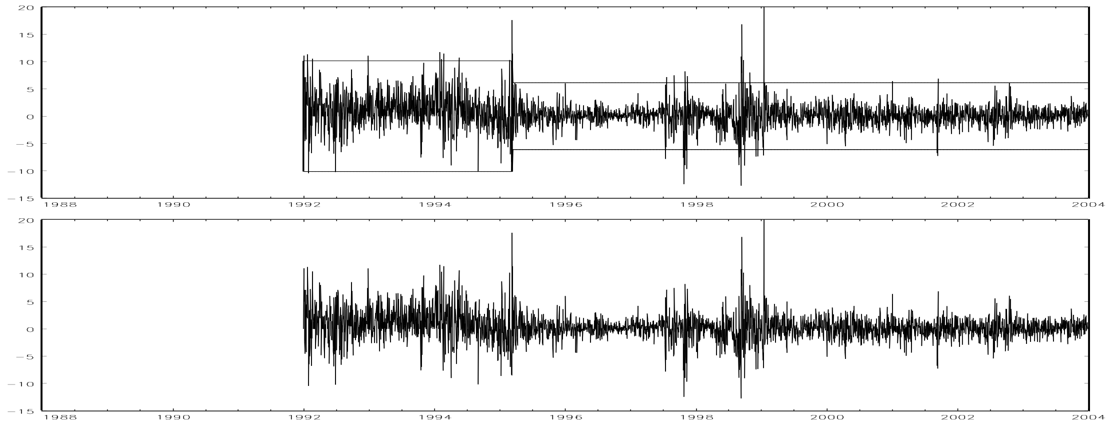
Although the number of variance changes based on the $U_{y,\gamma}(k^*)$ test appear to be quite reasonable, it seems only natural to apply the CUSUM tests to standardized returns in light of the size distortions documented in Section 2.3.1. Doing so using a GARCH(1,1) model yields the results shown in columns 8-10 and the final three columns of Table 2.9. Compared to the results for demeaned returns, the number of breaks further declines and to such an extent that for some countries no volatility changes are identified at all when using either the $U_{\hat{z},\gamma_0}(k^*)$ or $U_{\hat{z},\gamma}(k^*)$ statistics. Furthermore, these two tests now yield the exact same number of breaks for all countries. The $U_{\hat{z},\sigma}(k^*)$ statistic on the other hand still always identifies a positive number of breaks (except for U.S. dollar returns in Brazil and local currency returns in Pakistan), although considerably less than before and also less than the number of changes identified in Aggarwal *et al.* (1999).

The magnitude and timing of the identified volatility changes is examined graphically in Figure 2.4, which presents plots of the daily returns in local currency. The horizontal lines in these graphs indicate ± 3 times the unconditional standard deviation between consecutive change-points, as identified by the $U_{\hat{z},\sigma}(k^*)$ and $U_{\hat{z},\gamma}(k^*)$ statistics, in the upper and lower panels, respectively. Thick vertical lines correspond with the "official liberalization dates" as determined by Bekaert and Harvey (2000b) and Bekaert *et al.* (2003). It is seen that most identified breaks indeed correspond with marked shifts in volatility, many of which can be related to economic and political events such as, for example, the Asian and Russian financial crises in 1997 and 1998, respectively. Changes close to the official liberalization dates are found only for Chile and Indonesia, where in both cases volatility declined. What also becomes clear from Figure 2.4 is that whereas the $U_{\hat{z},\sigma}(k^*)$ test is inclined to find more, if not too many breaks, the $U_{\hat{z},\gamma_0}(k^*)$ and $U_{\hat{z},\gamma}(k^*)$ tests have a tendency to be conservative. For example, it is surprising to see that no volatility changes are identified for Indonesia, Korea and Malaysia around the middle of 1997 when the Asian crisis occurred. The time series plots suggest that the stock markets in these countries experienced a substantial and prolonged volatility increase around that time. This more or less confirms the reduced power for the $U_{\hat{z},\gamma}(k^*)$ statistic when testing for multiple breaks, reported in Section 2.3.2.

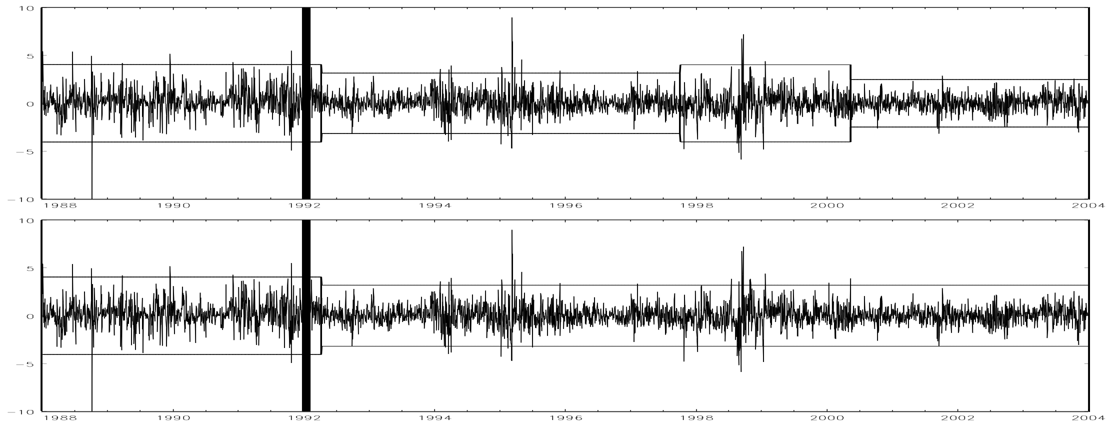
Figure 2.2: Break point locations



(i) Argentina



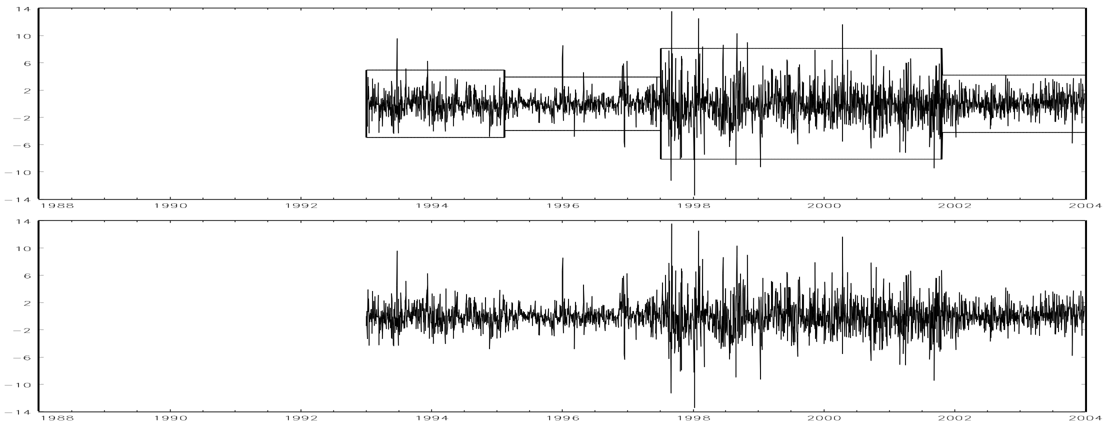
(ii) Brazil



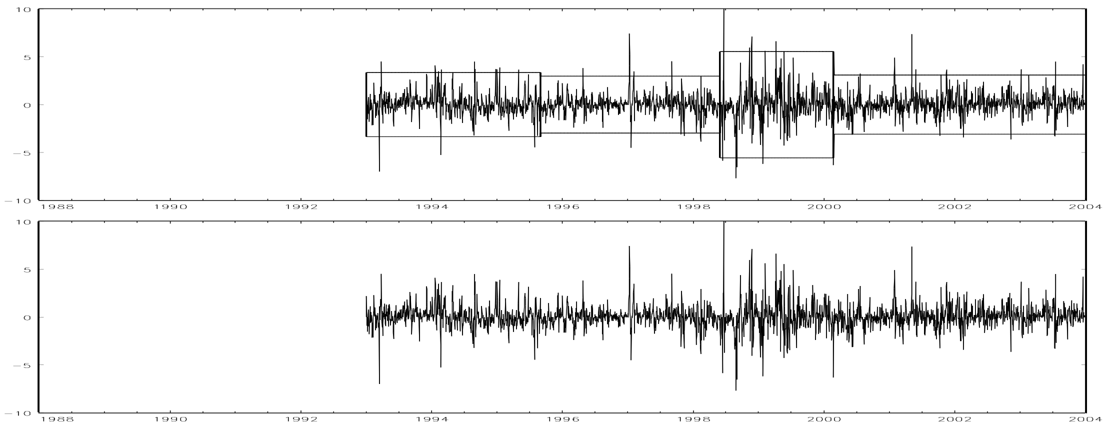
(iii) Chile

Note: The figure shows the daily emerging stock market returns in local currency. The full sample period is January 4, 1988 - December 31, 2003. The horizontal lines in these graphs indicate ±3 times the unconditional standard deviation between consecutive change-points, as identified by the $U_{\hat{z},\sigma}(k^*)$ and $U_{\hat{z},\gamma}(k^*)$ statistics, in the upper and lower panels, respectively. Thick vertical lines correspond with the "official liberalization dates" as determined by Bekaert and Harvey (2000b) and Bekaert *et al.* (2003).

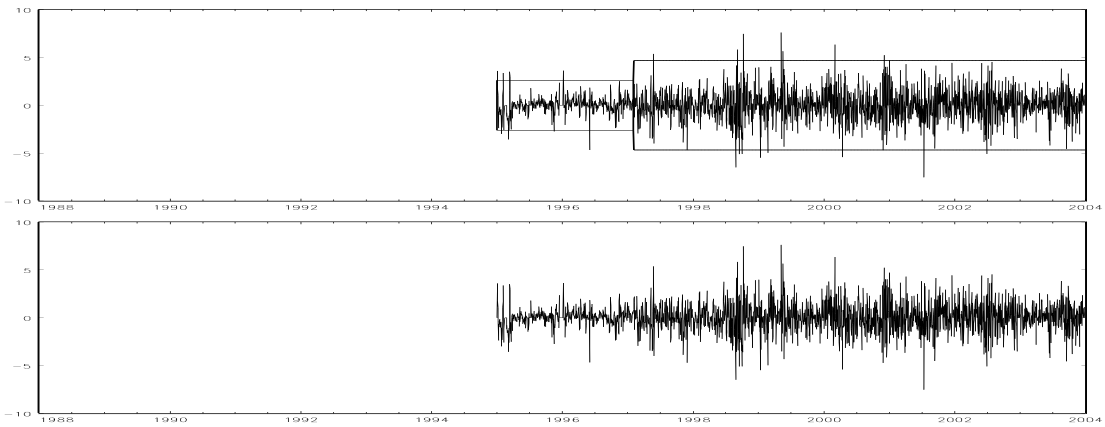
Figure 2 (continued)



(iv) China

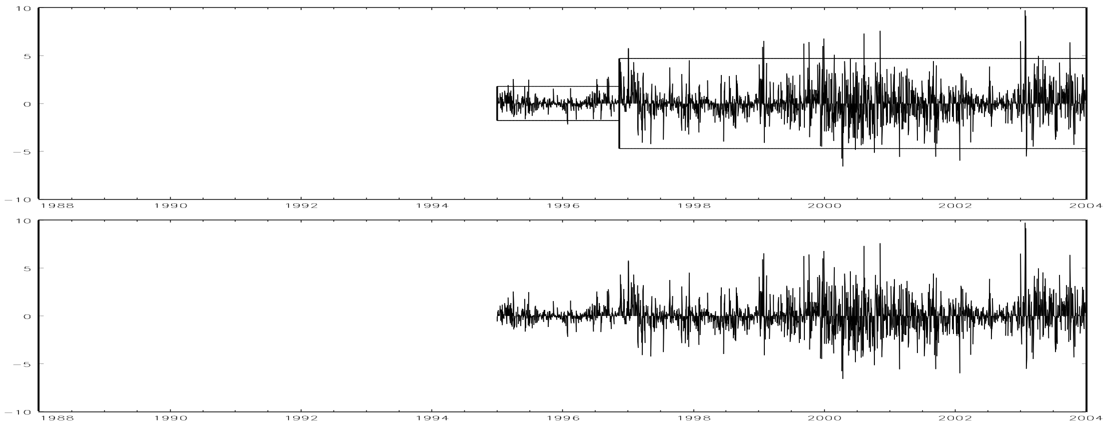


(v) Colombia

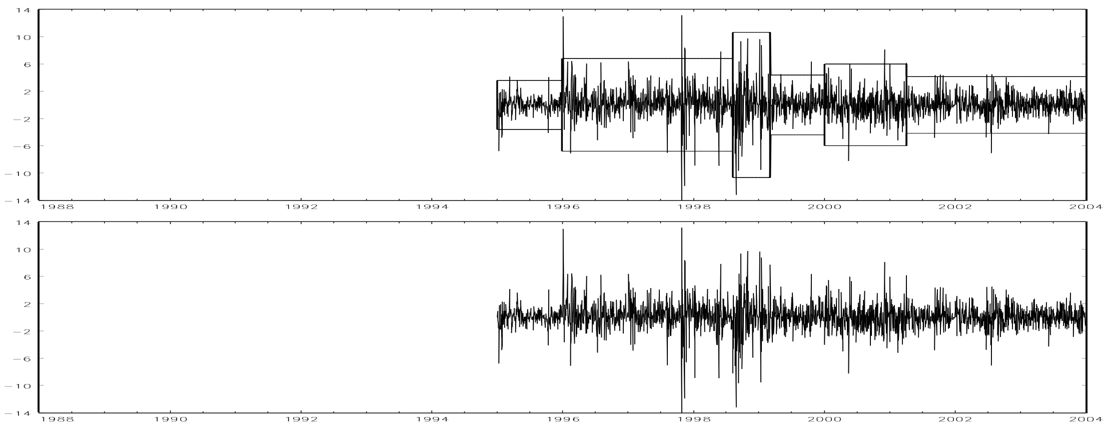


(vi) Czech Republic

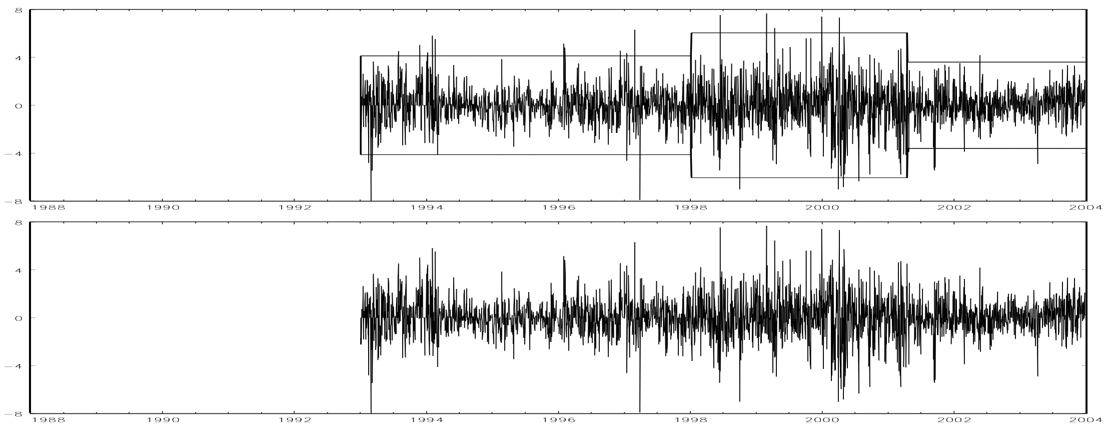
Figure 2 (continued)



(vii) Egypt

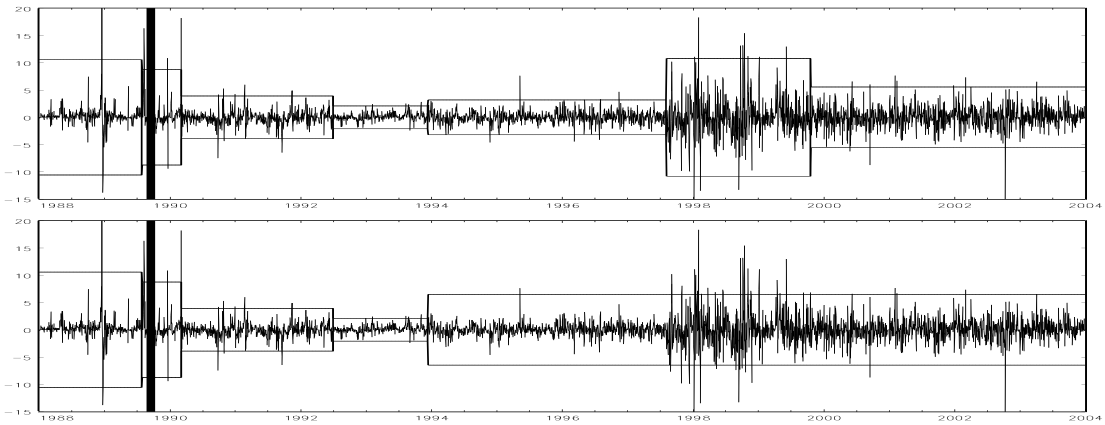


(viii) Hungary

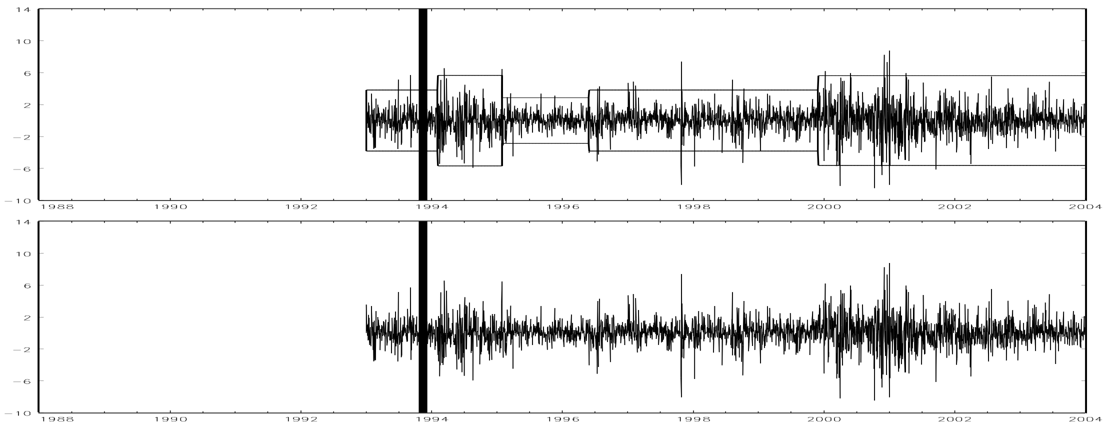


(ix) India

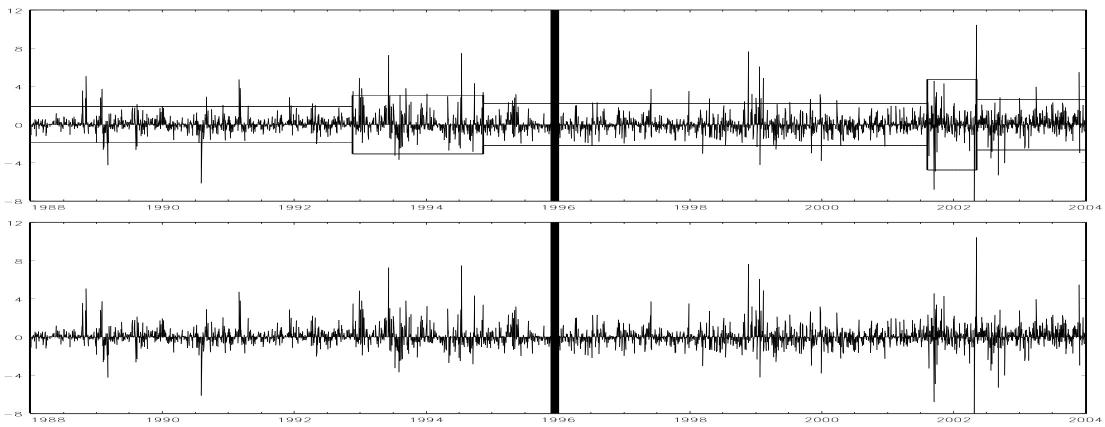
Figure 2 (continued)



(x) Indonesia

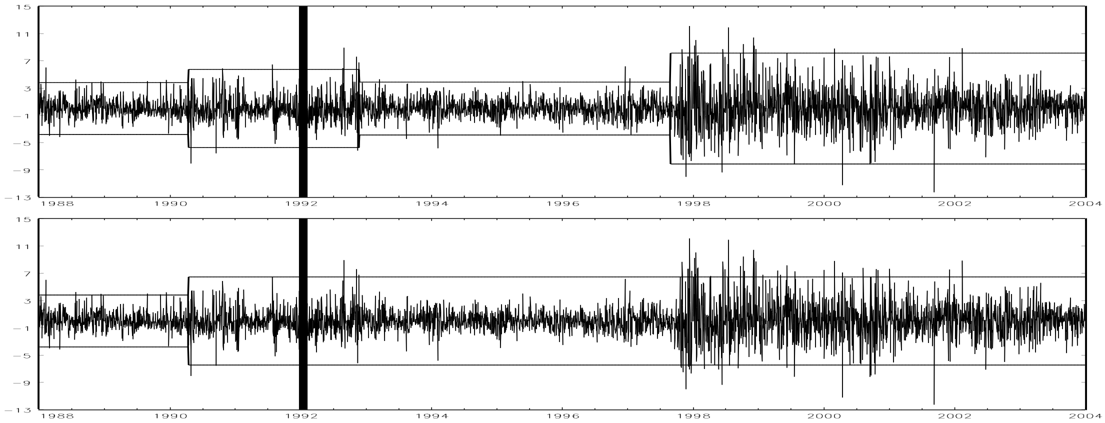


(xi) Israel

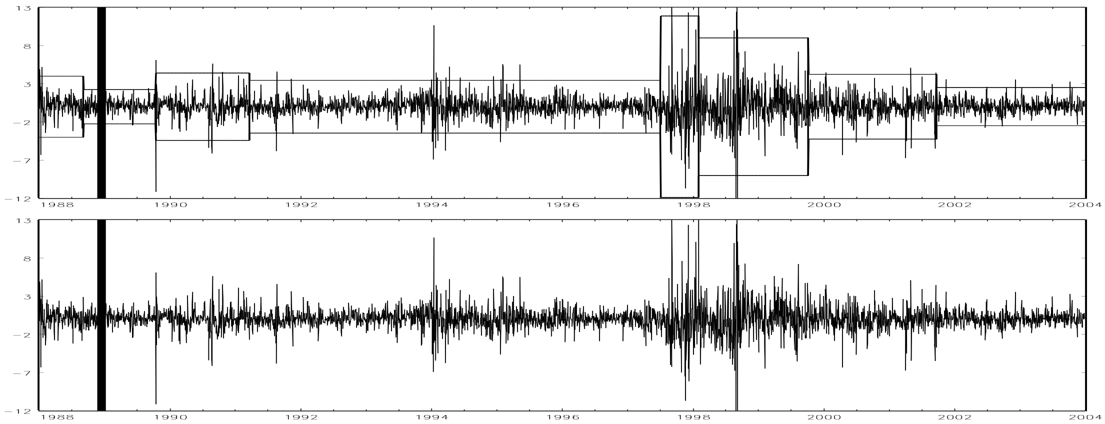


(xii) Jordan

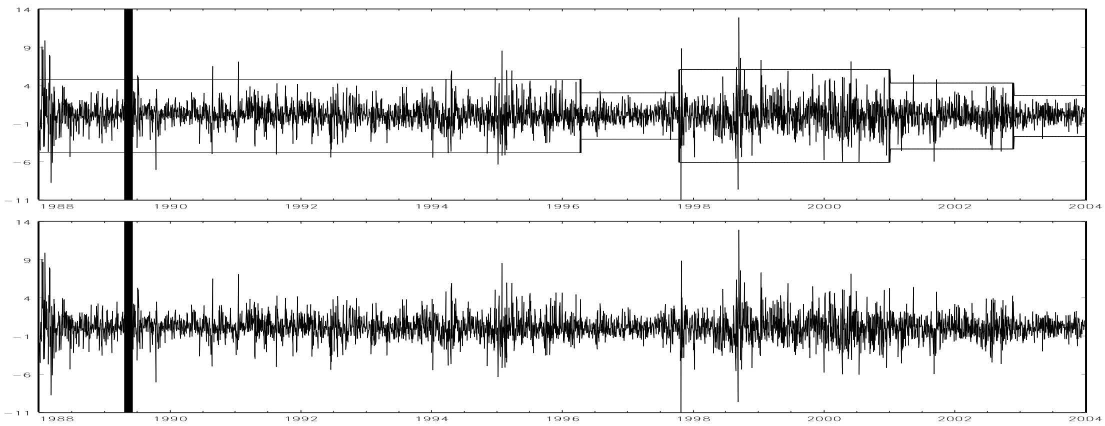
Figure 2 (continued)



(xiii) Korea

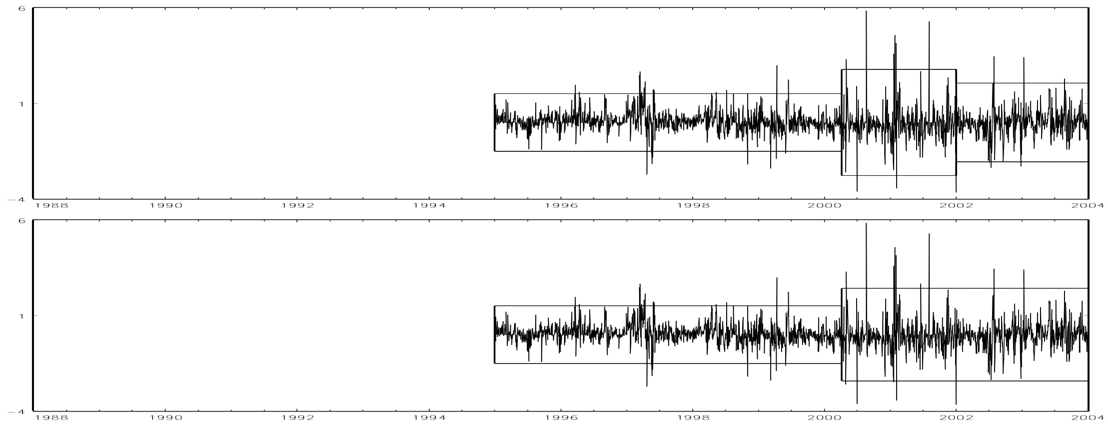


(xiv) Malaysia

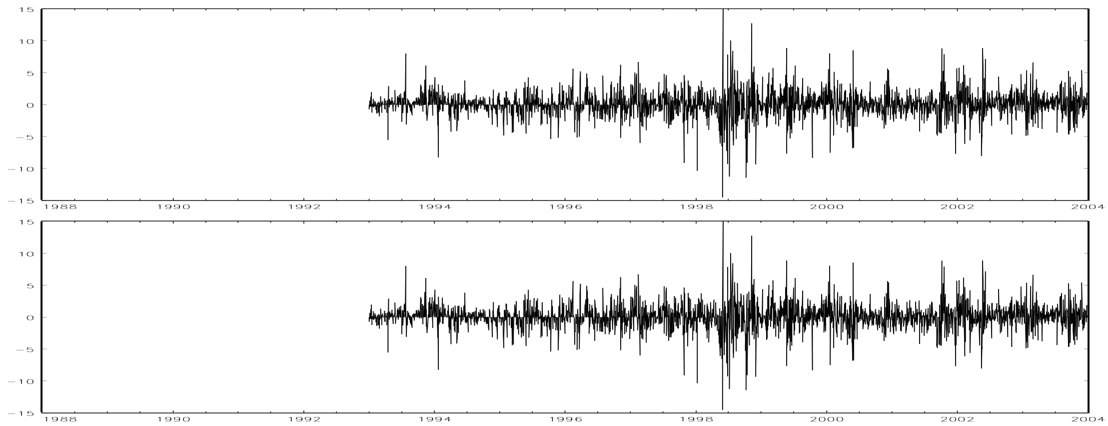


(xv) Mexico

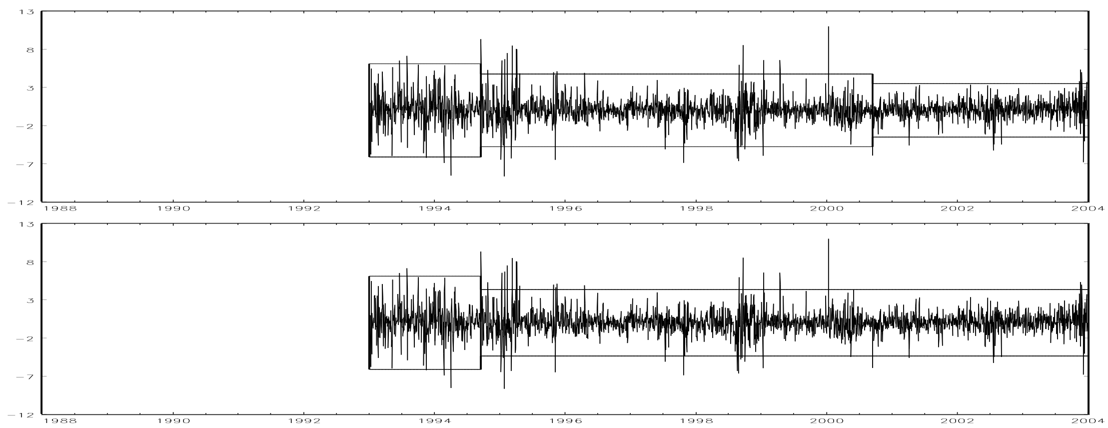
Figure 2 (continued)



(xvi) Morocco

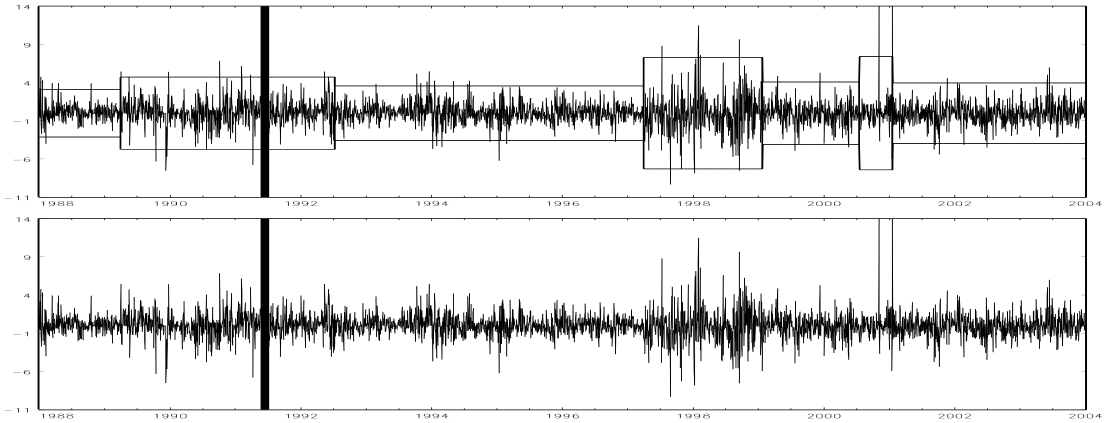


(xvii) Pakistan

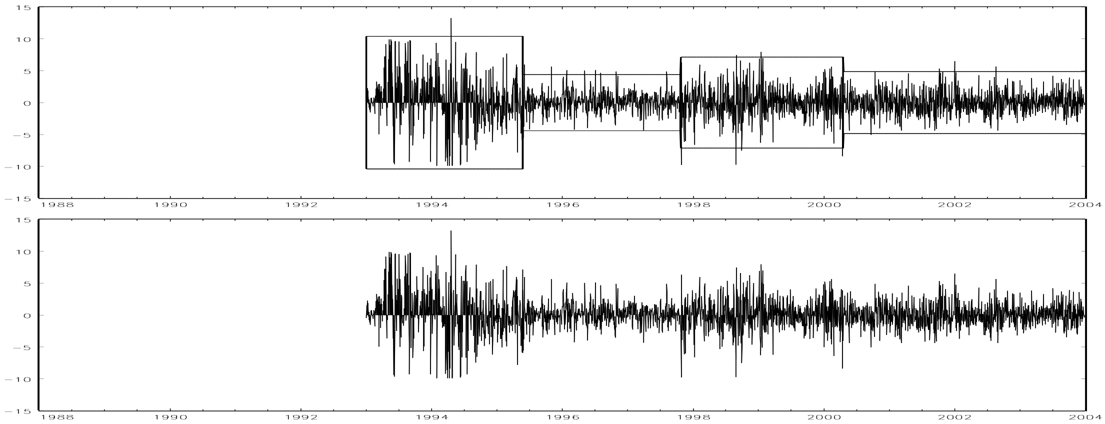


(xviii) Peru

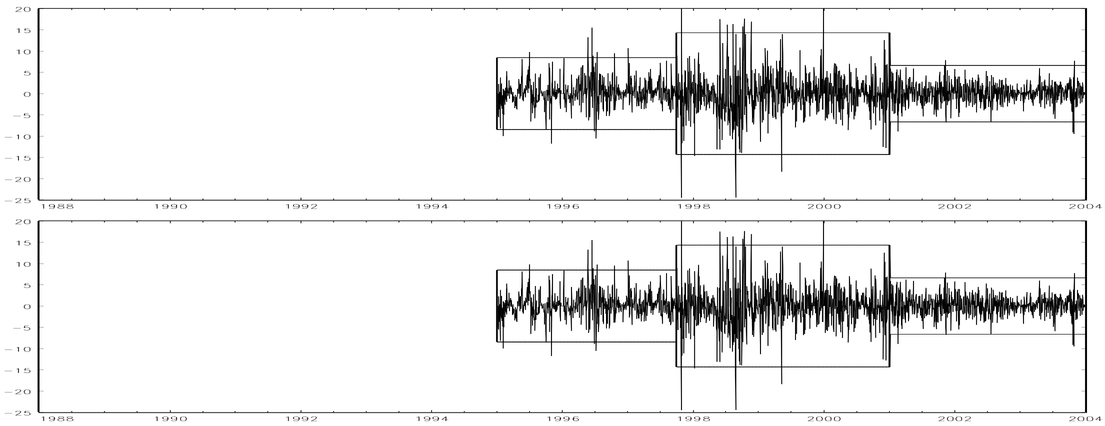
Figure 2 (continued)



(xix) Philippines

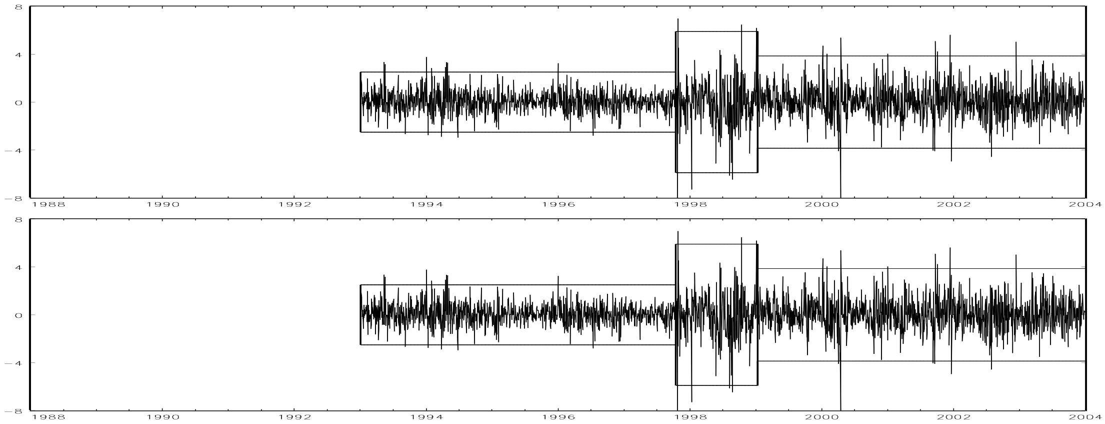


(xx) Poland

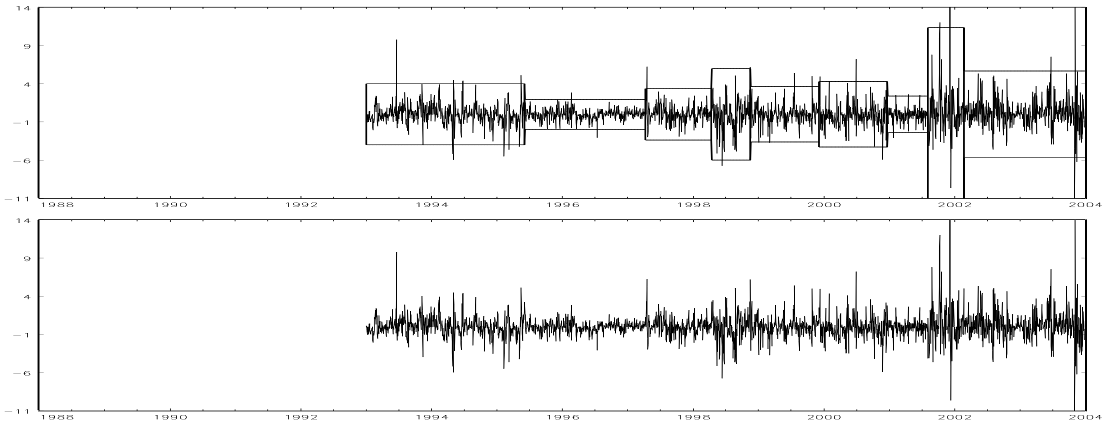


(xxi) Russia

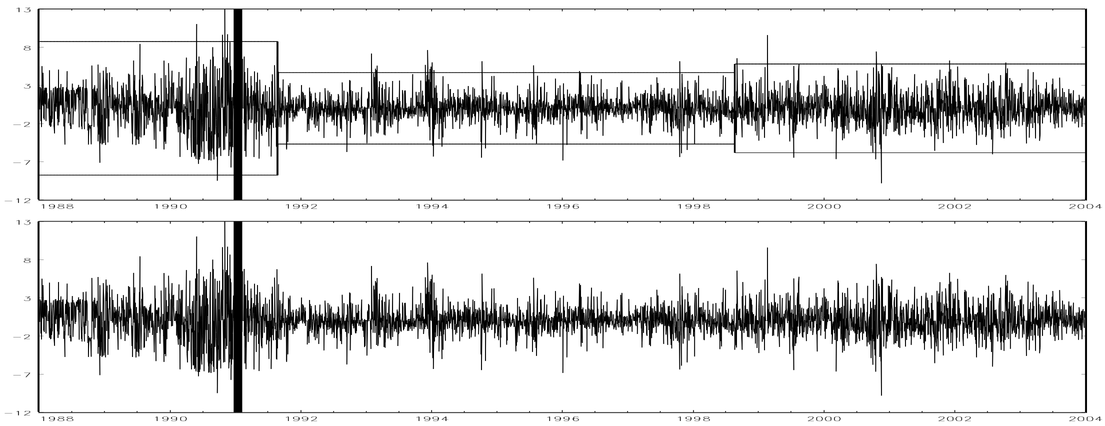
Figure 2 (continued)



(xxii) South Africa

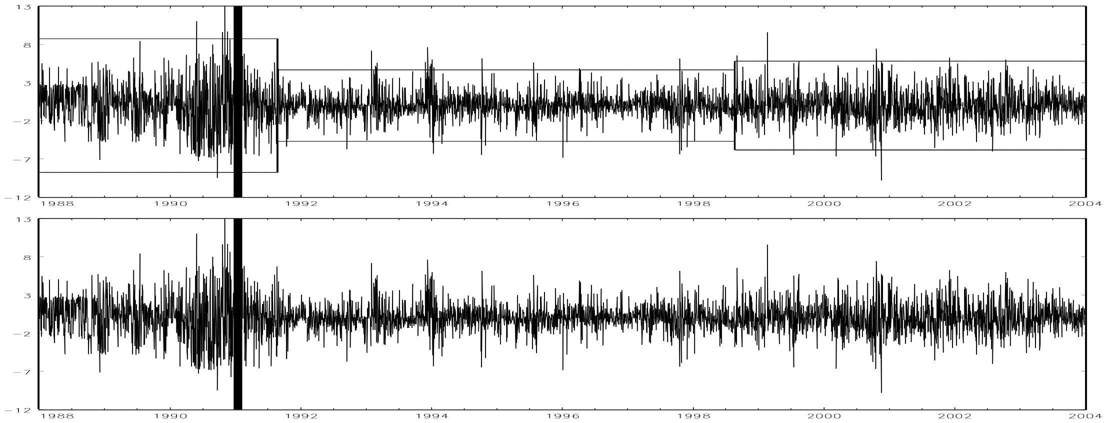


(xxiii) Sri Lanka

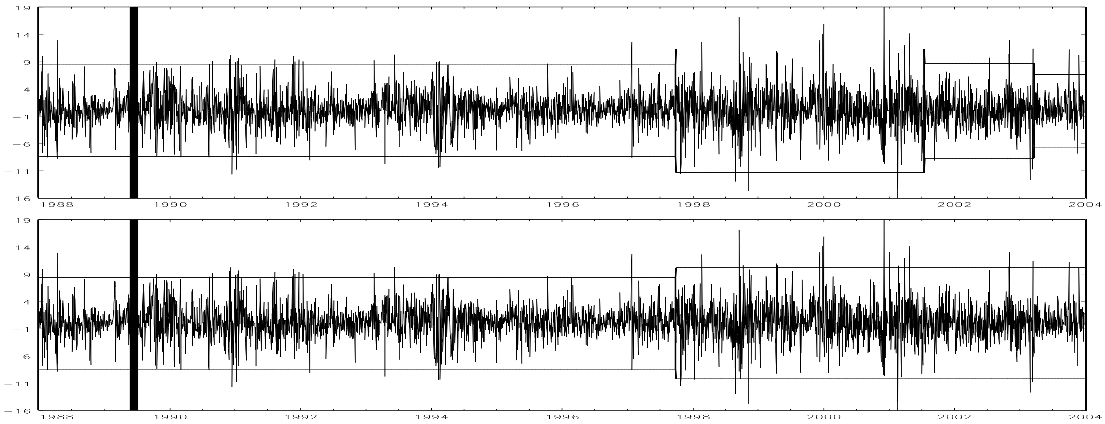


(xxiv) Taiwan

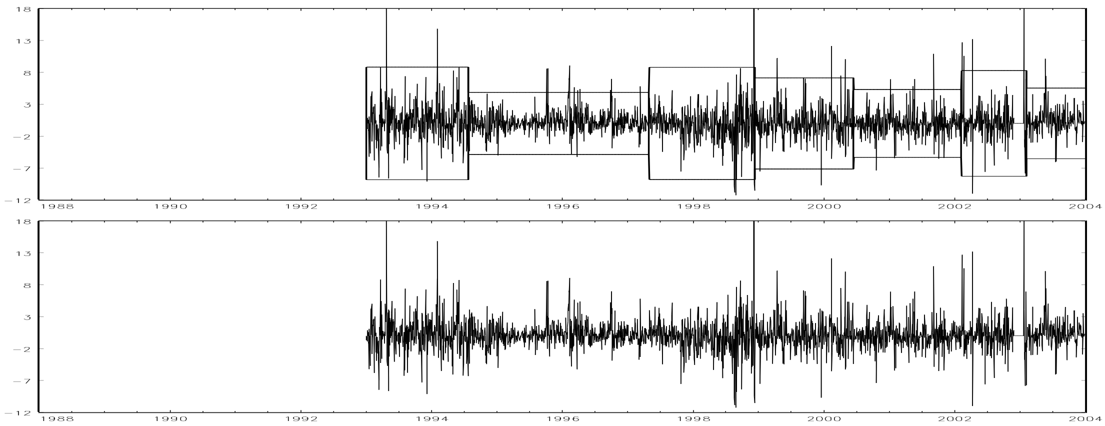
Figure 2 (continued)



(xxv) Thailand



(xxvi) Turkey



(xxvii) Venezuela

2.5 Concluding remarks

In this paper we have examined CUSUM-based tests for changes in the unconditional volatility of conditionally heteroskedastic time series. A prominent conclusion from our analysis is that application of these tests to the raw time series observations leads to severe size distortions, rendering the tests highly unreliable. Remarkably, this was also found to be the case for the CUSUM test of Kokoszka and Leipus (2000), which at least theoretically allows for the presence of heteroskedasticity. Our simulation results show that it may require unrealistically large sample sizes for this asymptotic result to apply. Consequently, it appears necessary to filter the series in order to remove the heteroskedasticity prior to applying the CUSUM test. As a practical way to accomplish this, we adopt the suggestion of Lee *et al.* (2003) to use a GARCH(1,1)-volatility filter. Put differently, we recommend to apply the CUSUM test to standardized residuals from an estimated GARCH(1,1) model. Extensive Monte Carlo simulations showed that this results in correctly sized tests with good power properties when testing for a single break. Furthermore, the tests were found to be reasonably robust against various forms of model misspecification. The CUSUM tests appear to have difficulty to detect multiple changes in volatility and, hence, developing a more powerful procedure for testing for multiple breaks is an interesting topic for future research. The general properties of the CUSUM tests were confirmed in an application to emerging stock market returns, where the GARCH-filtered tests led to a considerably smaller, and much more realistic number of volatility changes than the original CUSUM statistics.

Chapter 3

Modeling and Forecasting S&P 500 Volatility

Long memory, structural breaks, announcement effects and day-of-the-week seasonality

3.1 Introduction

Accurately measuring and forecasting financial volatility is of crucial importance for asset and derivative pricing, asset allocation and risk management. Merton (1980) already noted that the variance over a fixed period can be estimated arbitrarily accurately by the sum of squared intra-period returns, provided the data are available at a sufficiently high sampling frequency. With transaction prices becoming more widely available, Andersen and Bollerslev (1997, 1998a,b) kick-started a flurry of research on the use of high-frequency data for measuring and forecasting volatility. Andersen and Bollerslev (1998a) showed that ex-post daily exchange rate volatility is best measured by aggregating 288 squared five-minute returns. The five-minute frequency is a trade-off between accuracy, which is theoretically optimized using the highest possible frequency, and noise due to, for example, the bid-ask bounce.[1] Despite the remaining small measurement error the volatility essentially becomes "observable" ex-post.[2] As such, volatility can be modeled directly, rather than being treated as a latent variable as is the case in GARCH and stochastic volatility models. The main drawback of such models is the need to make specific assumptions regarding the distribution of shocks and the properties of the latent volatility factor. The

[1]See Zhang, Mykland, and Aït-Sahalia (2005), Aït-Sahalia, Mykland, and Zhang (2005) and Bandi and Russell (2005, 2006) among others for recent discussions involved in the choice of optimal sampling frequency in the presence of market microstructure noise.

[2]See Andersen, Bollerslev, and Diebold (2002) and Barndorff-Nielsen and Shephard (2002a,b, 2003, 2004b,c) for formal discussions of the theoretical properties of realized volatility and the related concept of power variation.

sum of intraday squared returns is also a much more accurate measure of daily realized volatility than the popular daily squared return.[3]

Several recent studies document the properties of realized volatilities constructed from high-frequency data for different financial assets, including exchange rates (Andersen, Bollerslev, Diebold, and Labys, 2001b), stock indexes and corresponding futures (Ebens, 1999; Areal and Taylor, 2002; Martens, 2002; Thomakos and Wang, 2003) and individual stocks (Andersen, Bollerslev, Diebold, and Ebens, 2001a). One of the stylized facts to come out of these studies is that realized volatilities are fractionally integrated of order d, where d typically is around 0.4. This property is used for modeling and forecasting volatilities at daily or longer horizons for both exchange rates (Andersen, Bollerslev, Diebold, and Labys, 2003a; Li, 2002; Pong, Shackleton, Taylor, and Xu, 2004 and stock indexes (Ebens, 1999; Koopman, Jungbacker, and Hol, 2005; Martens and Zein, 2004). These studies use Autoregressive Fractionally Integrated Moving Average (ARFIMA) models to capture both the long memory characteristic and any remaining short-term dynamics. The resulting forecasts generally outperform those obtained from ARCH models (including GARCH, EGARCH and FIGARCH), Riskmetrics' historical volatility with exponentially declining weights, and stochastic volatility models, and they can compete with implied volatility forecasts obtained from options. The latter is noteworthy as the current literature (e.g. Jorion (1995) and Christensen and Prabhala (1998)) suggests that implied volatility forecasts are superior to forecasts obtained from ARCH models, to the extent that ARCH forecasts do not contain any information not already subsumed by implied volatility forecasts.

Except for Ebens (1999) all the aforementioned studies use linear models, which ignore several empirically important aspects of financial volatility. First, linear models do not allow for the so-called leverage effect documented by, among others, Black (1976), Pagan and Schwert (1990) and Engle and Ng (1993). These studies show an asymmetric relation between news (as measured by lagged unexpected returns) and volatility using daily squared returns, in that a negative return tends to increase subsequent volatility by more than would a positive return of the same magnitude. Using realized volatility measures it is less clear whether a leverage effect will be present as a realized volatility measure will already contemporaneously capture negative return shocks. However, Bollerslev *et al.* (2006) show evidence of a prolonged leverage effect for S&P500 futures returns which can

[3] More accurate in the sense that it has smaller variance. Also, of relevance to this study, estimates of the degree of fractional integration are unbiased for daily volatility based on intraday returns, whereas they are severely downward biased when estimated from daily squared returns, see Bollerslev and Wright (2000).

last up to several days. Allowing for a leverage effect seems therefore worthwhile. Second, occasional structural breaks can spuriously suggest the presence of long memory, as shown by Diebold and Inoue (2001), among others. However, Andersen and Bollerslev (1997) show that long-memory is more likely to be an inherent feature of the return-generating process than an artefact due to structural streaks. As financial volatility has been found to experience irregular level shifts (see Lamoureux and Lastrapes, 1990, and Andreou and Ghysels, 2002a), which will at least influence the degree of long memory, it seems important to nevertheless consider this characteristic when modeling realized volatility. Third, Baillie and Bollerslev (1989) and Harvey and Huang (1991), among others, find that average volatility is not constant across the different days of the week but displays a rather pronounced U-shaped pattern with volatility being lowest on Wednesdays. Furthermore, volatility tends to be different on and around holidays. Fourth, Andersen and Bollerslev (1998b) explain the day-of-the-week effect in Deutsche Mark-Dollar volatility by the clustering of macroeconomic news announcements on specific weekdays (in particular on Fridays). Allowing for volatility to be different on days with news releases can disentangle calendar and announcement effects. Fifth, Bomfim (2003b) allows for pre-announcement effects as financial markets tend be relatively calm on days before important macro releases.

In this chapter, we propose a nonlinear model for realized volatilities that simultaneously captures long memory, leverage effects, structural breaks, day-of-the-week and macroeconomic news announcement effects. To the best of our knowledge, we are the first to develop such a comprehensive nonlinear model. The small number of previous studies that have considered nonlinearities in realized volatilities all are limited in one way or another. Ebens (1999), Oomen (2002) and Giot and Laurent (2004) incorporate leverage effects in a long memory model for various stock indexes. Only Giot and Laurent (2004) consider out-of-sample forecasting, but only at the one-day horizon. Maheu and McCurdy (2002) use a regime-switching model for the DM/$ exchange rate, but do not consider any other nonlinearities or long memory and only forecast one-day-ahead.

Our model is estimated and used to produce volatility forecasts at various horizons for S&P 500 index-futures. The results first of all show that level shifts in S&P 500 volatility can indeed not account for the long memory feature. The fractional integration parameter does decline when explicitly modeling structural breaks, but remains significantly different from zero. Second, the day-of-the-week dummies show that volatility is on average lower on Mondays and Tuesdays and higher on Fridays. This is an interesting contrast with the U-shaped pattern found in daily squared returns, which also attribute a higher volatility to Mondays and Tuesdays. Second, volatility tends to be substantially higher on announcement days and lower on pre-announcement days. Fourth, we find convincing evidence for the presence of a prolonged leverage effect in S&P volatility, in that negative returns significantly increase volatility on the following day whereas positive returns do not affect volatility at all. Incorporating these nonlinear features is important for out-of-sample forecasting as well. We find that 1-day-ahead volatility forecasts from the best nonlinear model improve upon those from a linear ARFIMA model on all evaluation criteria considered. For example, the R^2 from a regression of realized volatility on the volatility forecast increases from 52.1% to 55.4%. We find similar improvements when looking at simple AR models, i.e. dropping the long memory from the model. Also here adding the non-linearities improves the volatility forecasts. The simpler AR approach performs best for forecast horizons up to 20 days. For 10-day ahead forecasts for example. the R^2 is 59.2%, compared to 56.1% for the non-linear ARFIMA model, and 53.1% for the linear ARFIMA model.

The remainder of this chapter is structured as follows. First, we discuss the S&P 500 data in Section 3.2. The nonlinear long-memory model is developed in Section 3.3. We discuss the estimation of the model and discuss in-sample results for the S&P 500 in Section 3.4. In Section 3.5, we focus on out-of-sample forecasting. Section 3.6 concludes.

3.2 Data

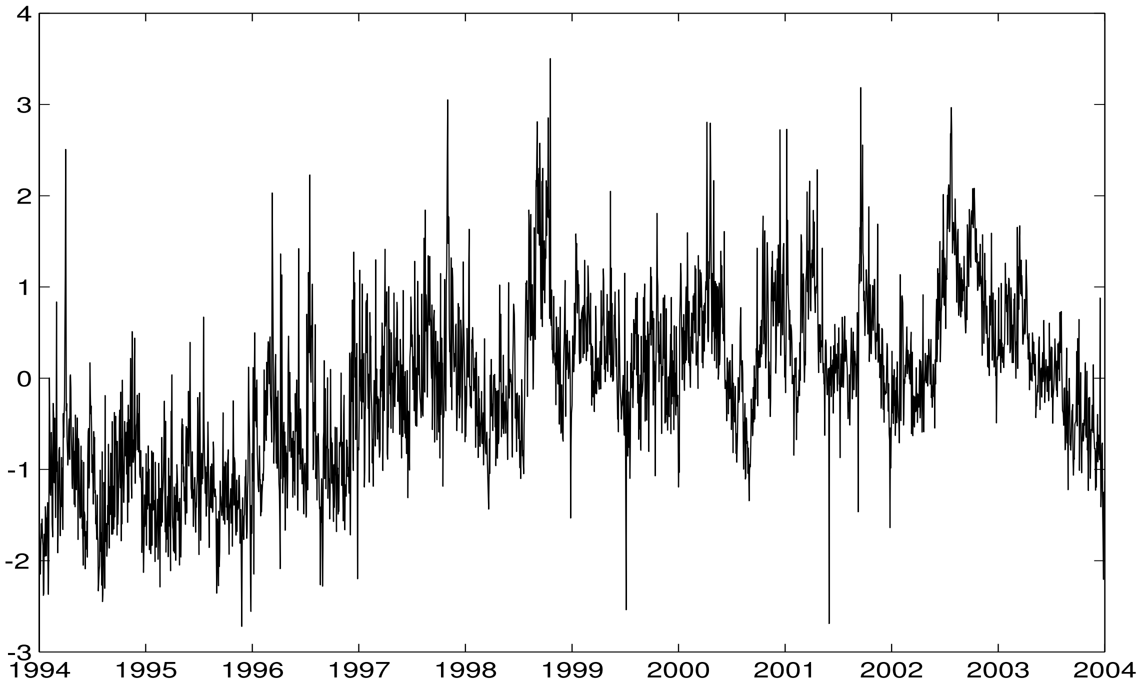
We construct our measure of daily realized volatility for the S&P 500 index using high-frequency futures data. S&P 500 index futures trade on the Chicago Mercantile Exchange (CME) on the trading floor from 8:30AM to 3:15PM (Eastern Standard Time minus 1 hour, EST-1). Since January 3, 1994, these contracts also trade overnight on GLOBEX, the electronic trading system of the CME, from 3:30PM to 8:00AM (8:15AM from February 26, 1996, onwards). As a result, S&P 500 futures trade almost round the clock, providing a similar opportunity to construct realized volatilities as for the 24-hour foreign exchange market. Martens (2002) tested various measures of S&P 500 realized volatility, finding that the sum of squared 30-minute intranight and 5-minute intraday returns is a more accurate measure of volatility than using only the intraday returns, or the sum of squared intraday returns and the squared close-to-open return, showing that it is useful to incorporate overnight trading prices. Hence, we will use the following measure of daily "realized volatility",

$$s_t^2 = \sum_{j=1}^{n_N} \left(r_{t,j}^N\right)^2 + \sum_{j=1}^{n_D} \left(r_{t,j}^D\right)^2, \tag{3.1}$$

where $r_{t,j}^N$ is the intranight (30-minute) return on day t in intranight period j ($j = 1, \ldots, n_N = 33$), and $r_{t,j}^D$ is the intraday (five-minute) return on day t for intraday period j ($j = 1, \ldots, n_D = 91$). Both $r_{t,j}^N$ and $r_{t,j}^D$ are continuously compounded returns.

Figure 3.1 shows a time series plot for the daily S&P 500 log realized volatility for the sample period from January 3, 1994, until December 29, 2003 (2521 daily observations). Table 3.1 contains descriptive statistics of the log realized volatility measure, as well as for daily returns $r_t = \sum_{j=1}^{n_N} r_{t,j}^N + \sum_{j=1}^{n_D} r_{t,j}^D$, for squared daily returns, and for daily returns standardized with the realized standard deviation, r_t/s_t. A number of interesting features emerge from this table, which closely correspond with the distributional characteristics for

Figure 3.1: Realized S&P 500 volatility



(a) Log realized variance

Notes: Daily log realized volatility for S&P 500 returns for the period from January 3, 1994, until December 29, 2003 (2521 observations).

Table 3.1: Descriptive statistics for daily S&P 500 return and realized volatility

	Mean	Med	Min	Max	Std.dev.	Skew	Kurt
Returns	0.022	0.058	−7.811	5.737	1.193	−0.164	6.435
Standardized returns	0.086	0.060	−2.725	3.105	0.999	0.073	2.730
Squared returns	1.423	0.400	0.000	61.005	3.314	7.788	100.921
Realized variance	1.421	0.927	0.066	33.220	1.907	6.006	62.907
Log realized variance	−0.107	−0.076	−2.721	3.503	0.934	0.133	3.026

Notes: The table contains summary statistics for daily S&P 500 return and realized volatility measures. The sample period covers January 3, 1994 until December 31, 2003 (2521 observations). Standardized returns are obtained by dividing the raw returns by the realized standard deviation.

realized exchange rate volatility documented in Andersen, Bollerslev, Diebold, and Labys (2001b). First, comparing the daily squared returns with the realized variance shows that these have an all but identical mean (1.423% and 1.421%, respectively). We would expect this to be the case, as both are unbiased measures of the true volatility. However, the standard deviation of the realized variance is at 1.907 much smaller than the standard deviation of the squared returns, which equals 3.314. It is precisely this characteristic that shows that realized variance is a less noisy estimate of true volatility than the daily squared return. Second, the realized variance is heavily skewed and exhibits excess kurtosis. By contrast, the logarithm of realized volatility, $\log(s_t^2)$, is much more symmetrically distributed and has much lower kurtosis. It is for this reason that we will consider time series models for the log realized volatility. Third, the daily S&P 500 returns are skewed and leptokurtic. Standardized returns r_t/s_t, however, exhibit much less skewness and excess kurtosis and are in fact very close to being normally distributed.

Fourth, as documented in other studies (and therefore not shown here explicitly), the sample autocorrelation functions of the realized volatility measures exhibit a slow hyperbolic decay, indicative for the presence of long memory. A further point is that the persistence in the autocorrelation functions is much stronger for the realized volatility measures than for the daily squared returns.

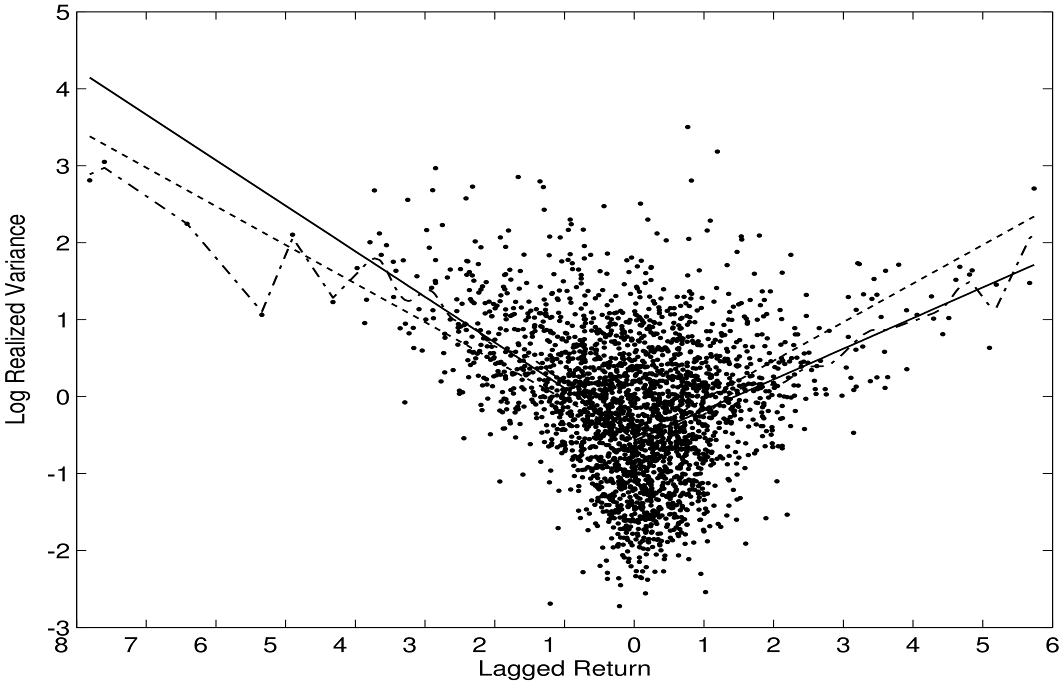
Fifth, returning to Figure 3.1, realized volatility appears to be higher on average in the second half of the sample period than during the first few years. It is difficult to pin down when exactly this level shift occurred, and it appears that it is most adequately characterized as a gradual increase of volatility during 1996-1997. After this transition period volatility seems to remain high until the last two years of the sample during which it declines sharply. An alternative possibility is that multiple structural breaks have occurred, as suggested by Andreou and Ghysels (2002a).

The scatter plot of $\log(s_t^2)$ against r_{t-1} in Figure 3.2 reveals a rather pronounced relationship between current volatility and lagged returns beyond that already captured in contemporaneous realized volatility. To examine the possible presence of a leverage effect, we estimate the "news impact curve" (Engle and Ng, 1993)

$$\log(s_t^2) = \beta_0 + \beta_1|r_{t-1}| + \beta_2\mathtt{I}[r_{t-1} < 0] + \beta_3|r_{t-1}|\mathtt{I}[r_{t-1} < 0], \tag{3.2}$$

where $\mathtt{I}[A]$ is an indicator function for the event A, being equal to 1 if A occurs, and 0 otherwise. The fit from this regression is included in Figure 3.2 as well, along with the fit

Figure 3.2: Leverage effects in realized volatility



Notes: Scatter plot of daily log realized variance and lagged returns, based on observations for the period from January 3, 1994, until December 29, 2003 (2521 observations). The solid line is the fit of the news impact curve (3.2), where log realized volatility is regressed on a constant, the lagged absolute return, a dummy for negative returns and an interaction term of this dummy with the lagged absolute return. The dashed line is the fit of a symmetric news impact curve, i.e. (3.2) with $\beta_2 = \beta_3 = 0$. The dot-dashed line is the fit from a nonparametric regression of log realized volatility on the lagged return.

from a symmetric version of this news impact curve, obtained by setting $\beta_2 = \beta_3 = 0$ in (3.2). It is clearly seen by the solid line in the graph that the impact of negative lagged returns is larger than the effect of positive returns of equal magnitude. Also, the parametric form in (3.2) appears to be quite reasonable, as can be seen by comparing the fit from this regression with a nonparametric regression of log realized volatility on the lagged return, also shown by the dash-dotted line in Figure 3.2.

Sixth, Table 3.2 shows the overall mean for all return and volatility measures on different types of days by distinguishing between regular days, holidays and days with macroeconomic news announcements. It is clear that returns and volatility are both higher after holidays and that volatility during the Christmas period is only roughly half of its level during regular days. Volatility is in particular substantially higher on announcement days, especially on days when decisions regarding the federal funds rate are released when realized variance is on average 2.379%, substantially above its non-announcement day mean of 1.352%.

Finally, Table 3.3 shows the overall mean and the mean on different days-of-the-week. The top panel in the table confirms the common finding based on daily returns that Mondays and Fridays exhibit higher volatility than other days by the S&P data. Interestingly,

Table 3.2: Descriptive statistics for daily S&P 500 return and realized volatility

	ALL	NONE	HOL	XMS	ANN	EMP	PPI	CPI	FF
Returns	0.022	−0.005	0.051	0.054	0.130	0.204	0.129	0.036	0.265
Standardized returns	0.086	0.061	0.157	0.113	0.187	0.250	0.092	0.235	0.243
Squared returns	1.423	1.325	2.694	0.761	1.817	2.042	1.710	1.650	2.166
Realized variance	1.421	1.352	1.721	0.602	1.816	1.927	1.800	1.523	2.379
Log realized variance	−0.107	−0.140	−0.058	−0.755	0.106	0.271	0.037	−0.031	0.156
Number of obs.	2521	1977	76	41	430	117	118	120	85

Notes: The table contains sample averages of daily S&P 500 returns and realized volatility measures. The sample period covers January 3, 1994 until December 31, 2003 (2521 observations). ALL indicates all days in the sample period; NONE indicates days without announcements, not following a holiday, and not in the Christmas period. HOL indicates days following a holiday. XMS denotes days during the Christmas period; ANN indicates all days with one or more macroeconomic announcements; EMP, PPI, CPI and FF indicate days with an announcement of employment, PPI, CPI, and the Federal Funds target rate, respectively. Standardized returns are obtained by dividing the raw returns by the realized standard deviation.

this pattern is quite different for the realized variance. Thursdays and Fridays exhibit the highest volatility but Mondays no longer have an above average volatility. In fact, for realized variance the mean is lowest on Mondays.

The observed patterns can to a large extent be explained by making the distinction between days with and without macro releases, similar as in Andersen and Bollerslev (1998b). The middle and bottom panels in Table 3.3 allow for a direct comparison. Squared returns and realized variance are both clearly higher on announcement days. From the

Table 3.3: Day-of-the-week effects in S&P 500 return and realized volatility

	Overall	MON	TUE	WED	THU	FRI
All days						
Returns	0.022	0.034	0.049	-0.008	0.000	0.038
Standardized returns	0.086	0.128	0.094	0.038	0.074	0.101
Squared returns	1.423	1.581	1.492	1.229	1.359	1.464
Realized variance	1.421	1.295	1.347	1.422	1.506	1.530
Non-announcement days						
Returns	0.000	0.052	0.001	-0.024	-0.022	-0.013
Standardized returns	0.065	0.137	0.039	0.031	0.060	0.049
Squared returns	1.342	1.514	1.597	1.183	1.281	1.038
Realized variance	1.339	1.247	1.361	1.354	1.436	1.292
Announcement days						
Returns	0.130	-2.967	0.261	0.093	0.172	0.113
Standardized returns	0.187	-1.323	0.332	0.087	0.182	0.176
Squared returns	1.817	12.276	1.035	1.534	1.946	2.089
Realized variance	1.816	8.975	1.284	1.867	2.036	1.880
Number of obs.	430	3	96	68	59	204

Notes: The table contains daily means for S&P 500 returns and realized variance. The sample period covers January 3, 1994 until December 31, 2003 (2521 observations). Standardized returns are obtained by dividing the raw returns by the realized standard deviation. The three panels distinguish between statistics computed using all days (top panel), days without any macro news announcements (middle panel) and days on which at least one macro figure is released (bottom panel). The final row in the table shows the total number of announcement days and how these are dispersed across the days of the week.

bottom panel it becomes clear that most macro releases occur on Fridays (204 out of the total of 430 announcements). Squared returns and realized variance are both higher on Friday than on other announcement days. This explains the on average higher values for realized variance on Fridays in the top panel.

After correcting for announcement effects, squared returns are still higher at the beginning of the week with averages of 1.514% and 1.597% for Mondays and Tuesdays respectively, something which is not the case for realized variance.

3.3 Nonlinear Long Memory models

Following previous studies, we employ Autoregressive Fractionally Integrated Moving Average (ARFIMA) models to describe the dynamic properties of logarithmic realized volatility $y_t = \log(s_t^2)$,

$$\phi(L)(1-L)^d(y_t - \mu_t) = \varepsilon_t, \tag{3.3}$$

where the order of integration d is allowed to take non-integer values, $\phi(L) = 1-\phi_1 L - \ldots - \phi_p L^p$ is a p-th order lag polynomial assumed to have all roots outside the unit circle and ε_t is a white noise process. It is common practice to set the mean μ_t equal to a constant, i.e. $\mu_t = c$. However, to capture the salient features of the S&P realized volatility discussed in the previous section, we extend the model to allow for gradual level shifts, day-of-the-week and holiday effects, announcement effects and nonlinear effects of lagged returns by setting

$$\begin{aligned}\mu_t &= c(t) + \beta_1|r_{t-1}| + \beta_2\mathtt{I}[r_{t-1}<0] + \beta_3|r_{t-1}|\mathtt{I}[r_{t-1}<0] \\ &+ \delta_1 D^*_{1,t} + \delta_2 D^*_{2,t} + \delta_4 D^*_{4,t} + \delta_5 D^*_{5,t} + \delta_6 D_{\mathrm{HOL},t} + \delta_7 D_{\mathrm{XMS},t} \\ &+ \delta_8 D_{\mathrm{EMP},t} + \delta_9 D_{\mathrm{PPI},t} + \delta_{10} D_{\mathrm{CPI},t} + \delta_{11} D_{\mathrm{FF},t} \\ &\qquad + \delta_{12} D_{\mathrm{EMP}_{t+1},t} + \delta_{13} D_{\mathrm{PPI}_{t+1},t} + \delta_{14} D_{\mathrm{CPI}_{t+1},t} + \delta_{15} D_{\mathrm{FF}_{t+1},t}\end{aligned} \tag{3.4}$$

where $D^*_{s,t} \equiv D_{s,t} - D_{3,t}\ D_{s,t}$, $s = 1, 2, 4, 5$ are "centered" daily dummy variables, with $D_{s,t} = 1$ when time t corresponds with day s (1=Monday, 2=Tuesday, etc.) and $D_{s,t} = 0$ otherwise. The dummies $D_{\mathrm{HOL},t}$ and $D_{\mathrm{XMS},t}$ are 1 for days following a holiday and days during the Christmas period respectively. In our choice of macro announcements we follow Ederington and Lee (1993) by including daily dummies for the release of the Employment report ($D_{\mathrm{EMP},t}$), Consumer Price Index ($D_{\mathrm{CPI},t}$) and Producer Price Index ($D_{\mathrm{PPI},t}$). In addition we include a daily dummy for FOMC meeting days on which decisions regarding the Federal Funds target rate are announced $(D_{\mathrm{FF},t})$[4]. Similar to Bomfim (2003b) we also include "calm-before-the-storm" dummies $(D_{\mathrm{EMP}_{t+1},t},\ D_{\mathrm{CPI}_{t+1},t},\ , D_{\mathrm{PPI}_{t+1},t}, D_{\mathrm{FF}_{t+1},t})$ to allow for volatility to be depressed on pre-announcement days.[5]

We capture gradual level shifts in volatility by specifying the parameter $c(t)$ as an arbitrary deterministic trend function on $t \in [1, ..., T]$ which is estimated semi-parametrically using kernel regression, see Beran and Ocker (2001) and Beran (1999) for further details.[6] The above ARFI model with Semi-Parametric mean (SP), holiday and day-of-the-week Dummies (D), macro news (pre-)Announcement dummies (A), lagged Return (R) and Leverage effects (L) will be denoted SPARFI-DARL.

We also estimate an alternative model, where (cf. Ebens, 1999) terms involving the lagged returns are not included in the conditional mean μ_t, but as "exogenous regressors" (X), leading to the model (denoted SPARFI-DAXRL)

$$\phi(L)(1-L)^d(y_t - \mu_t) = \tilde{\beta}_1|r_{t-1}| + \tilde{\beta}_2\mathtt{I}[r_{t-1} < 0] + \tilde{\beta}_3|r_{t-1}|\mathtt{I}[r_{t-1} < 0] + \varepsilon_t, \tag{3.5}$$

where now

$$\begin{aligned}\mu_t &= c(t) + \delta_1 D^*_{1,t} + \delta_2 D^*_{2,t} + \delta_4 D^*_{4,t} + \delta_5 D^*_{5,t} + \delta_6 D_{\mathrm{HOL},t} + \delta_7 D_{\mathrm{XMS},t} \\ &+ \delta_8 D_{\mathrm{EMP},t} + \delta_9 D_{\mathrm{PPI},t} + \delta_{10} D_{\mathrm{CPI},t} + \delta_{11} D_{\mathrm{FF},t} \\ &\qquad + \delta_{12} D_{\mathrm{EMP}_{t+1},t} + \delta_{13} D_{\mathrm{PPI}_{t+1},t} + \delta_{14} D_{\mathrm{CPI}_{t+1},t} + \delta_{15} D_{\mathrm{FF}_{t+1},t}\end{aligned} \tag{3.6}$$

Finally, for completeness, we examine a similar model where the (pre-)announcement dummies are now also included as exogenous regressors.

$$\begin{aligned}\phi(L)(1-L)^d(y_t - \mu_t) &= \tilde{\beta}_1|r_{t-1}| + \tilde{\beta}_2\mathtt{I}[r_{t-1} < 0] + \tilde{\beta}_3|r_{t-1}|\mathtt{I}[r_{t-1} < 0] \\ &+ \delta_8 D_{\mathrm{EMP},t} + \delta_9 D_{\mathrm{PPI},t} + \delta_{10} D_{\mathrm{CPI},t} + \delta_{11} D_{\mathrm{FF},t} \\ &\qquad + \delta_{12} D_{\mathrm{EMP}_{t+1},t} + \delta_{13} D_{\mathrm{PPI}_{t+1},t} + \delta_{14} D_{\mathrm{CPI}_{t+1},t} + \delta_{15} D_{\mathrm{FF}_{t+1},t} + \varepsilon_t,\end{aligned} \tag{3.7}$$

[4]Note that as our sample starts in 1994 we are not confronted with the issue that the Federal Reserve started to announce the decision about changing the target rate during regularly scheduled FOMC meetings only as of 1994, see Andersen *et al.* (2003b) and Bomfim (2003b). Note further that as of 1994 there are eight scheduled FOMC meetings each year and that during our sample period (1994-2003) there have been five announced meetings: April 18, 1994; October 15, 1998; January 3, 2001; April 18, 2001 and September 17, 2001. Also for these days we set $D_{\mathrm{FF},t} = 1$. Furthermore, for the FOMC meetings that last two days we set $D_{\mathrm{FF},t} = 1$ only for the second day.

[5]Bomfim (2003b) examines the news effect of monetary policy decisions on S&P 500 index returns and therefore only includes a calm-before-the-storm effect surrounding FOMC meetings. Here we allow volatility to be different on every pre-announcement day.

[6]More specifically we obtain the estimated function $\widehat{c}_t,\ t \in [1, ..., T]$ by using the Nadaraya-Watson kernel regression estimator with a Gaussian kernel and choosing the Silverman (1986) optimal bandwidth.

and where therefore

$$\mu_t = c(t) + \delta_1 D^*_{1,t} + \delta_2 D^*_{2,t} + \delta_4 D^*_{4,t} + \delta_5 D^*_{5,t} + \delta_6 D_{\mathrm{HOL},t} + \delta_7 D_{\mathrm{XMS},t} \quad (3.8)$$

To gauge the relative importance of the different nonlinear features of realized volatility, we also estimate several restricted versions of the full models. In particular, we consider (i) a model without the leverage effect but including the lagged absolute return ($\beta_2 = \beta_3 = 0$ in (3.4) or $\tilde{\beta}_2 = \tilde{\beta}_3 = 0$ in (3.5) and (3.7); SPARFI-DA(X)R), (ii) a model without any effect of lagged returns at all ($\beta_1 = \beta_2 = \beta_3 = 0$ in (3.4) or $\tilde{\beta}_1 = \tilde{\beta}_2 = \tilde{\beta}_3 = 0$ in (3.5) and (3.7); SPARFI-DA), (iii) a model without any effect of lagged returns and without (pre-)announcement effects (imposing in addition that $\delta_i = 0$, for $i = 8, ..., 15$ in (3.4), (3.6) or (3.8); SPARFI-D) and (iv) a model without any effect of lagged returns, (pre-)announcement effects and seasonal effects (thereby also imposing $\delta_1 = ... = \delta_7 = 0$ in (3.4), (3.6) or (3.8); SPARFI). Finally, we also estimate all models without allowing for structural changes (by imposing that $c(t)$ is equal to a constant ($c(t) = c$).

For estimation of the parameters in the ARFI models we use the Beran (1995) approximate maximum likelihood (AML) estimator for invertible and possibly non-stationary ARFIMA models (i.e. for $d > -0.5$), which amounts to minimizing the sum of squared residuals

$$Q_n(\theta) = \sum_{t=1}^{T} e_t^2(\theta), \quad (3.9)$$

where $\theta = (d, \gamma, \tau, c, \beta, \delta, \phi)$, T is the sample size and the residuals $e_t(\theta)$ are computed as

$$e_t(\theta) = (y_t - \mu_t) - \sum_{j=1}^{t+p-1} \pi_j (y_{t-j} - \mu_{t-j}) \quad (3.10)$$

where the π_j are the autoregressive coefficients in the infinite order AR representation of the ARFI models

$$\pi(L)(y_t - \mu_t) = \varepsilon_t, \quad (3.11)$$

that is $\pi(L) = 1 - \pi_1 L - \pi_2 L^2 - \ldots \equiv \phi(L)(1 - L)^d$, and the π_j can be computed by using the binomial expansion of the fractional differencing operator $(1 - L)^d$,

$$(1 - L)^d = 1 - dL + \frac{d(d-1)L^2}{2!} - \frac{d(d-1)(d-2)L^3}{3!} + \cdots . \quad (3.12)$$

The AML estimator is asymptotically efficient if the errors ε_t are normally distributed. Under less restrictive regularity conditions, it is $\sqrt{n}$ consistent and asymptotically normal.

We employ the Akaike Information Criterion (AIC) in the full nonlinear model to select the appropriate autoregressive order p. The selected lag order $p = 2$ is subsequently imposed in the nested models, to facilitate comparison of the parameter estimates.

3.4 Estimation results

All results discussed in this section are based on estimating models over the period from January 3, 1994 until December 31, 1997 (1011 observations). The remainder of the sample period will be used to evaluate the out-of-sample forecast performance of the various models. Detailed full-sample estimation results are available upon request. Table 3.4 contains estimation results for the different ARFI models in equations (3.4)-(3.8) which do not allow for structural change in μ_t ($c(t) = c$). Table 3.5 shows results for the corresponding models which do allow for such level shifts in realized volatility. Several conclusions can be drawn from these tables. First, the order of integration d ranges between 0.3 and 0.5, which is in line with estimates reported in previous studies. For some models the point estimate of d is very close to 0.5, suggesting that log realized volatility may be non-stationary. Note however, that in all models, the autoregressive parameters ϕ_1 and ϕ_2 are negative, such that the autoregressive coefficients in the AR(∞) representation are reduced as a result of which the model can still be considered stationary for practical purposes such as forecasting at relatively short horizons. Comparing the estimates of d for the models with constant $c(t)$ in Table 3.4 with those in Table 3.5 makes clear that allowing for structural change in μ_t lowers the order of integration, confirming that neglecting level shifts may spuriously suggest fractional integration, cf. Diebold and Inoue (2001). In the ARFI-DARL model, for example, d is estimated at 0.493, compared with 0.377 in the SPARFI-DARL model. Note, however, that the point estimates of d are still significantly different from zero in the models with structural change. Hence, as expected from the results in Andersen and Bollerslev (1997), the level shift cannot fully account for the long memory feature in realized volatility. It is also interesting to note that the order of integration is affected by the way the lagged returns are treated: if these are included as exogenous regressors, the estimate of d is substantially lower than if these are included in μ_t.

Table 3.4: Estimated ARFI models for daily S&P 500 realized volatility, January 1994-December 1997

	ARFI	ARFI-D	ARFI-DA	ARFI-DAR	ARFI-DARL	ARFI-DAXRL	ARFI-DXARL
$\hat{d}$	0.471	0.480	0.478	0.469	0.493	0.387	0.382
	(0.046)	(0.046)	(0.046)	(0.046)	(0.044)	(0.036)	(0.035)
$\hat{\alpha}$	−0.778	−0.778	−0.786	−0.868	−0.857	−1.517	−1.654
	(0.311)	(0.318)	(0.308)	(0.285)	(0.302)	(0.256)	(0.247)
$\widehat{\beta_1}$	–	–	–	0.137	−0.003	0.017	0.008
				(0.036)	(0.041)	(0.037)	(0.038)
$\widehat{\beta_2}$	–	–	–	–	−0.070	−0.042	−0.049
					(0.044)	(0.044)	(0.045)
$\widehat{\beta_3}$	–	–	–	–	0.308	0.346	0.364
					(0.051)	(0.048)	(0.049)
$\widehat{\delta}_{\mathrm{MON}}$	–	−0.139	−0.097	−0.101	−0.101	−0.103	−0.119
		(0.032)	(0.032)	(0.033)	(0.032)	(0.032)	(0.033)
$\widehat{\delta}_{\mathrm{TUE}}$	–	−0.078	−0.085	−0.089	−0.089	−0.093	−0.096
		(0.032)	(0.032)	(0.031)	(0.031)	(0.031)	(0.032)
$\widehat{\delta}_{\mathrm{THU}}$	–	0.003	0.069	0.086	0.086	0.092	0.102
		(0.030)	(0.032)	(0.032)	(0.031)	(0.031)	(0.031)
$\widehat{\delta}_{\mathrm{FRI}}$	–	0.176	0.064	0.062	0.048	0.049	0.064
		(0.034)	(0.035)	(0.035)	(0.034)	(0.035)	(0.034)
$\widehat{\delta}_{\mathrm{HOL}}$	–	0.304	0.340	0.335	0.299	0.302	0.286
		(0.169)	(0.158)	(0.163)	(0.167)	(0.170)	(0.173)
$\widehat{\delta}_{\mathrm{XMS}}$	–	−0.533	−0.485	−0.457	−0.434	−0.446	−0.455
		(0.184)	(0.179)	(0.163)	(0.149)	(0.144)	(0.143)
$\widehat{\delta}_{\mathrm{PPI}}$	–	–	0.182	0.169	0.134	0.118	0.080
			(0.087)	(0.089)	(0.088)	(0.088)	(0.092)
$\widehat{\delta}_{\mathrm{EMP}}$	–	–	0.539	0.564	0.568	0.568	0.587
			(0.094)	(0.096)	(0.095)	(0.095)	(0.103)
$\widehat{\delta}_{\mathrm{CPI}}$	–	–	0.089	0.100	0.090	0.084	0.064
			(0.074)	(0.070)	(0.065)	(0.063)	(0.066)
$\widehat{\delta}_{\mathrm{FF}}$	–	–	0.235	0.252	0.269	0.276	0.316
			(0.109)	(0.110)	(0.109)	(0.112)	(0.121)
$\widehat{\delta}_{\mathrm{PPI}(+1)}$	–	–	−0.022	−0.016	−0.018	−0.032	−0.063
			(0.075)	(0.076)	(0.073)	(0.073)	(0.077)
$\widehat{\delta}_{\mathrm{EMP}(+1)}$	–	–	−0.258	−0.265	−0.262	−0.269	−0.302
			(0.078)	(0.076)	(0.076)	(0.076)	(0.079)
$\widehat{\delta}_{\mathrm{CPI}(+1)}$	–	–	0.002	0.001	−0.002	0.003	0.018
			(0.093)	(0.096)	(0.097)	(0.098)	(0.105)
$\widehat{\delta}_{\mathrm{FF}(+1)}$	–	–	−0.115	−0.105	−0.112	−0.111	−0.098
			(0.090)	(0.089)	(0.089)	(0.090)	(0.095)
$\widehat{\phi}_1$	−0.104	−0.090	−0.044	−0.098	−0.193	−0.117	−0.139
	(0.053)	(0.053)	(0.052)	(0.055)	(0.056)	(0.046)	(0.045)
$\widehat{\phi}_2$	−0.086	−0.100	−0.094	−0.073	−0.074	−0.050	−0.043
	(0.041)	(0.041)	(0.041)	(0.042)	(0.043)	(0.039)	(0.038)
$\hat{\sigma}_\varepsilon$	0.581	0.563	0.543	0.537	0.525	0.516	0.518
AIC	−1.078	−1.129	−1.187	−1.205	−1.248	−1.280	−1.272
BIC	−1.059	−1.080	−1.099	−1.112	−1.146	−1.177	−1.169
$\mathrm{LM_{SC}}(1)$	0.135	0.289	0.026	0.334	0.434	11.140	0.948
	(0.71)	(0.59)	(0.87)	(0.56)	(0.51)	((8.7E−4))	(0.33)
$\mathrm{LM_{SC}}(5)$	0.776	0.730	0.824	0.599	0.605	3.302	0.926
	(0.57)	(0.60)	(0.53)	(0.70)	(0.70)	((5.8E−3))	(0.46)
$\mathrm{LM_{SC}}(10)$	0.875	0.489	0.837	0.635	0.514	2.223	0.999
	(0.56)	(0.90)	(0.59)	(0.79)	(0.88)	(0.01)	(0.44)

Notes: The table presents parameter estimates, diagnostic measures and misspecification tests for estimated ARFI models for daily S&P 500 realized volatility, over the period January 3, 1994-December 31, 1997 (1011 observations). The column ARFI-DA(X)RL contains estimates of the model including seasonal and (pre-)announcement dummies and leverage effects as part of μ_t (as exogenous variables), the column ARFI-DA(X)R contains estimates of the model including seasonal and (pre-)announcement dummies and symmetric effects of the lagged absolute return as part of μ_t (as exogenous variable), the column ARFI-DA contains estimates for the model with seasonal and (pre-)announcement dummies but without the lagged absolute return, the column ARFI-DA contains estimates for the model with only seasonal dummies but without (pre-)announcement dummies and the column ARFI shows estimates for a model without dummies and without lagged absolute returns. $\hat{\sigma}_\varepsilon$ is the residual standard deviation. $\mathrm{LM_{SC}}(q)$ denotes the (F variant of the) LM test of no serial correlation in the residuals up to and including order q. The numbers in parentheses below parameter estimates and test statistics are heteroskedasticity-consistent standard errors and p-values, respectively.

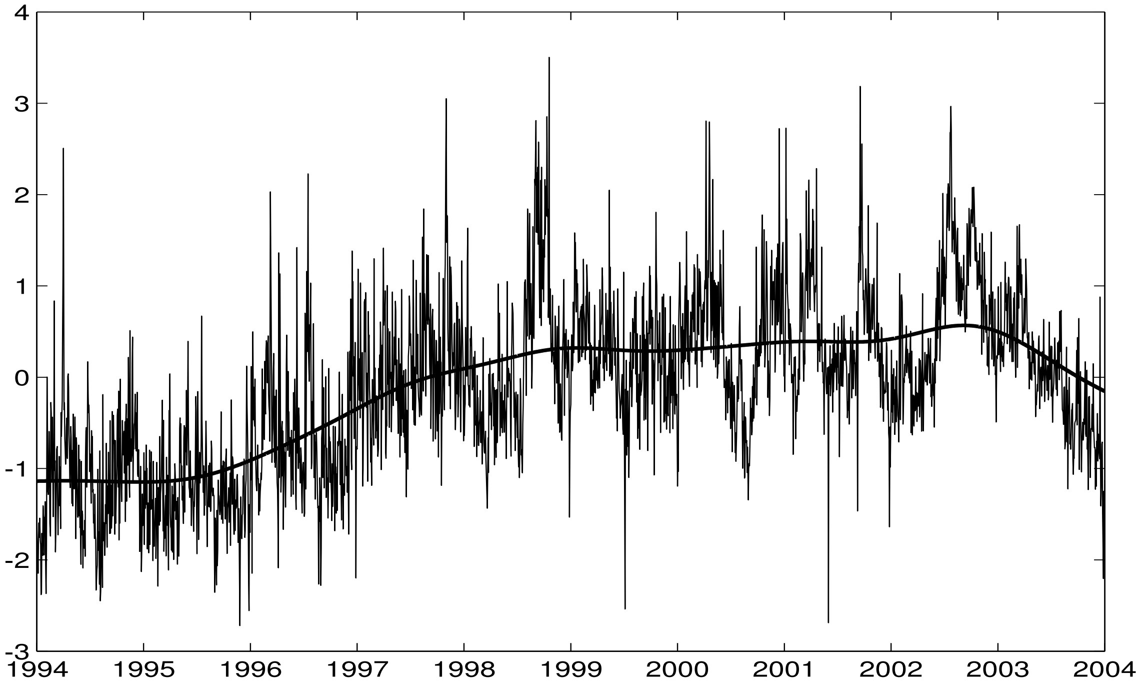
Table 3.5: Estimated SPARFI models for daily S&P 500 realized volatility, January 1994-December 1997

	SPARFI	SPARFI-D	SPARFI-DA	SPARFI-DAR	SPARFI-DARL	SPARFI-DAXRL	SPARFI-DXARL
$\hat{d}$	0.351	0.363	0.353	0.340	0.377	0.329	0.340
	(0.053)	(0.051)	(0.052)	(0.053)	(0.051)	(0.044)	(0.043)
$\widehat{\beta_1}$	–	–	–	0.141	0.001	−0.071	−0.090
				(0.037)	(0.041)	(0.025)	(0.027)
$\widehat{\beta_2}$	–	–	–		−0.069	−0.118	−0.135
					(0.044)	(0.032)	(0.034)
$\widehat{\beta_3}$	–	–	–		0.305	0.407	0.433
					(0.050)	(0.044)	(0.046)
$\widehat{\delta}_{\mathrm{MON}}$	–	−0.139	−0.098	−0.102	−0.102	−0.103	−0.126
		(0.032)	(0.033)	(0.033)	(0.032)	(0.032)	(0.032)
$\widehat{\delta}_{\mathrm{TUE}}$	–	−0.077	−0.085	−0.089	−0.090	−0.092	−0.096
		(0.032)	(0.032)	(0.032)	(0.031)	(0.032)	(0.032)
$\widehat{\delta}_{\mathrm{THU}}$	–	0.004	0.071	0.087	0.088	0.089	0.102
		(0.030)	(0.032)	(0.032)	(0.031)	(0.031)	(0.031)
$\widehat{\delta}_{\mathrm{FRI}}$	–	0.176	0.064	0.062	0.048	0.048	0.067
		(0.034)	(0.036)	(0.035)	(0.034)	(0.035)	(0.035)
$\widehat{\delta}_{\mathrm{HOL}}$	–	0.310	0.344	0.341	0.305	0.306	0.286
		(0.171)	(0.159)	(0.165)	(0.169)	(0.166)	(0.169)
$\widehat{\delta}_{\mathrm{XMS}}$	–	−0.550	−0.497	−0.472	−0.449	−0.460	−0.479
		(0.175)	(0.171)	(0.156)	(0.142)	(0.144)	(0.143)
$\widehat{\delta}_{\mathrm{PPI}}$	–	–	0.177	0.163	0.130	0.125	0.061
			(0.089)	(0.091)	(0.090)	(0.089)	(0.094)
$\widehat{\delta}_{\mathrm{EMP}}$	–	–	0.540	0.567	0.570	0.568	0.559
			(0.094)	(0.096)	(0.095)	(0.095)	(0.102)
$\widehat{\delta}_{\mathrm{CPI}}$	–	–	0.090	0.101	0.089	0.080	0.028
			(0.075)	(0.071)	(0.066)	(0.066)	(0.068)
$\widehat{\delta}_{\mathrm{FF}}$	–	–	0.241	0.259	0.274	0.270	0.291
			(0.112)	(0.113)	(0.112)	(0.113)	(0.122)
$\widehat{\delta}_{\mathrm{PPI(-1)}}$	–	–	−0.024	−0.018	−0.020	−0.025	−0.084
			(0.075)	(0.076)	(0.073)	(0.072)	(0.076)
$\widehat{\delta}_{\mathrm{EMP(-1)}}$	–	–	−0.261	−0.267	−0.264	−0.266	−0.326
			(0.079)	(0.076)	(0.077)	(0.078)	(0.080)
$\widehat{\delta}_{\mathrm{CPI(-1)}}$	–	–	0.006	0.006	0.001	−0.001	0.001
			(0.096)	(0.099)	(0.100)	(0.100)	(0.106)
$\widehat{\delta}_{\mathrm{FF(-1)}}$	–	–	−0.110	−0.100	−0.106	−0.118	−0.127
			(0.089)	(0.089)	(0.089)	(0.089)	(0.093)
$\widehat{\phi}_1$	0.001	0.011	0.065	0.014	−0.094	−0.043	−0.077
	(0.060)	(0.058)	(0.058)	(0.059)	(0.060)	(0.053)	(0.051)
$\widehat{\phi}_2$	−0.037	−0.054	−0.047	−0.020	−0.021	−0.020	−0.023
	(0.041)	(0.040)	(0.041)	(0.042)	(0.044)	(0.039)	(0.039)
$\hat{\sigma}_{\varepsilon}$	0.577	0.559	0.539	0.533	0.521	0.516	0.519
AIC	−1.093	−1.144	−1.202	−1.221	−1.264	−1.284	−1.272
BIC	−1.079	−1.100	−1.119	−1.133	−1.166	−1.186	−1.174
$\mathrm{LM_{SC}}(1)$	0.686	0.447	1.009	0.404	0.313	5.666	0.007
	(0.41)	(0.50)	(0.32)	(0.53)	(0.58)	(0.02)	(0.93)
$\mathrm{LM_{SC}}(5)$	1.572	1.275	1.487	1.056	1.239	2.271	0.692
	(0.17)	(0.27)	(0.19)	(0.38)	(0.29)	(0.05)	(0.63)
$\mathrm{LM_{SC}}(10)$	1.587	1.274	1.725	1.756	1.098	2.610	1.550
	(0.11)	(0.24)	(0.07)	(0.06)	(0.36)	(0.00)	(0.12)

Notes: The table presents parameter estimates, diagnostic measures and misspecification tests for estimated SPARFI models for daily S&P 500 realized volatility, over the period January 3, 1994-December 31, 1997 (1011 observations). The column SPARFI-DA(X)RL contains estimates of the model including seasonal and (pre-)announcement dummies and leverage effects as part of μ_t (as exogenous variables), the column SPARFI-DA(X)R contains estimates of the model including seasonal and (pre-)announcement dummies and symmetric effects of the lagged absolute return as part of μ_t (as exogenous variable), the column SPARFI-DA contains estimates for the model with seasonal and (pre-)announcement dummies but without the lagged absolute return, the column SPARFI-DA contains estimates for the model with only seasonal dummies but without (pre-)announcement dummies and the column SPARFI shows estimates for a model without dummies and without lagged absolute returns. $\hat{\sigma}_{\varepsilon}$ is the residual standard deviation. $\mathrm{LM_{SC}}(q)$ denotes the (F variant of the) LM test of no serial correlation in the residuals up to and including order q. The numbers in parentheses below parameter estimates and test statistics are heteroskedasticity-consistent standard errors and p-values, respectively.

Second, Figure 3.3 shows, for the full sample, that the semi-parametric $c(t)$ captures the structural break at the end of 1996 quite accurately as well as the downward trend in volatility at the end of the sample.

Figure 3.3: Modeling structural change



Notes: Plot of the daily log realized variance and the estimate of $c(t)$ in the SPARFI model (solid line) for the full sample from January 3, 1994 until December 29, 2003 (2521 observations).

Third, the estimates of the seasonal dummy parameters $\delta_1, \ldots, \delta_7$ confirm the descriptive statistics in Table 3.3, in that volatility is significantly higher after public holidays, significantly lower during the Christmas and on average realized volatility is also significantly lower on Mondays and Tuesdays and higher on Fridays. The latter is, however, for a large part due to not explicitly accounting for news announcement effects. The results for the models that include the (pre-)announcement dummy variables show that the day-of-week pattern to a large extent disappears. On announcement days volatility is substantially higher than on non-announcement days, especially on days when news about employment and the federal funds target rate is announced. The pre-announcement dummies show that there is also a calm-before-the-storm effect with volatility nearly always being lower on days before announcements. The effect is particularly strong on days before an employment record release.

Fourth, the models that include lagged returns indicate a significant relationship between $\log(s_t^2)$ and r_{t-1}. We also find convincing evidence for the presence of a leverage effect which corroborates the results of Bollerslev *et al.* (2006). The point estimates of β_1 and β_3 in ARFI-DA(X)RL models in fact suggest that only negative lagged returns affect current realized volatility, as $\hat{\beta}_1$ is not significantly different from zero.

Fifth, comparing the residual standard deviation, AIC and BIC across different columns shows that incorporating the different nonlinear features in the model enhances the in-sample fit. Allowing for (pre-)announcement dummies and a leverage effect appears to be most important in this respect, where the models which include the terms involving lagged returns as exogenous regressors (cf. Ebens, 1999) are preferred over models which include these terms in the conditional mean μ_t (cf. Oomen, 2002). Note that the AIC values for ARFI and SPARFI models do not differ substantially, suggesting that accounting for the level shift in realized volatility does not lead to much improvement of the model.

3.5 Forecasting volatility

The period from January 2, 1998 through December 29, 2003 (1510 observations) is used to judge the relevance of modeling the nonlinearities in S&P 500 realized volatility for out-of-sample forecasting purposes. All models are estimated recursively, using an expanding window of data. Volatility forecasts for 1- to 20-days ahead are constructed by means of Monte Carlo simulation, where we use the infinite order AR-representation of the ARFI models given in (3.11), truncated after 200 lags.[7] In addition to forecasts for logarithmic realized volatility, which is the dependent variable in the ARFI models, we also construct forecasts for the realized variance and realized standard deviation, where we ensure that these forecasts are unbiased.[8] In the forecast evaluation below, we concentrate on forecasts

[7] The truncated infinite order AR representation in (3.11) can be rewritten as $y_t = \hat{\mu}_t + \sum_{j=1}^{p^*} \pi_j (y_{t-j} - \hat{\mu}_{t-j}) + e_t$, where $p^* = 200$. The 1-step ahead forecast $y_{t+1|t}$ is obtained by sampling B independent random draws $z_{t+1}^{(i)}$, $i = 1, \ldots, B$ from a standard normal distribution, which are multiplied by the residual standard deviation $\hat{\sigma}_e$. The resulting shocks $e_{t+1}^{(i)} = z_{t+1}^{(i)} \hat{\sigma}_e$ are fed into the model to obtain a realization $y_{t+1|t}^{(i)} = \hat{\mu}_{t+1} + \sum_{j=1}^{p^*} \pi_j (y_{t+1-j} - \hat{\mu}_{t+1-j}) + e_{t+1}^{(i)}$. Finally, the 1-step ahead forecast $y_{t+1|t}$ is the mean across all B realizations, $y_{t+1|t} = \frac{1}{B} \sum_{i=1}^{B} y_{t+1|t}^{(i)}$. For multiple step ahead forecasts from models which include lagged returns in μ_t, these returns are simulated as well, by multiplying the standard deviation in the i-th path by another draw v_t from a standard normal distribution, e.g. $r_{t+1}^{(i)} = \sqrt{\exp(y_{t+1|t}^{(i)})} v_t$.

[8] This is achieved by applying the appropriate transformation to all simulated paths of log realized volatility individually, and then averaging. For example, the 1-step ahead forecast of the realized standard deviation is computed as $s_{t+1|t} = \frac{1}{B} \sum_{i=1}^{B} \sqrt{\exp(y_{t+1|t}^{(i)})}$.

for the standard deviation. Results for forecasts of the (logarithmic) variance are qualitatively similar and are available upon request. Furthermore, h-days ahead forecasts refer to realized standard deviations over the next h days, i.e. $\widehat{s}_{t+h|t} = \sqrt{\sum_{j=1}^{h} \widehat{s}^2_{t+j|t}}$ (instead of daily realized standard deviation h-days ahead).

3.5.1 Alternative forecasting models

For comparison purposes, forecasts were also produced for two popular models that only use daily returns and treat volatility as a latent variable. First, Riskmetrics uses historical volatility with exponentially declining weights,

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2, \tag{3.13}$$

with $\lambda = 0.94$.

Second, Glosten, Jagannathan, and Runkle (1993) (GJR) incorporated leverage effects into the popular Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. Also including day-of-week dummies, calm-before-the-storm dummies and announcement dummies, the GJR-GARCH(1,1)-DA model is specified as

$$\begin{aligned}
r_t &= c + u_t && (3.14)\\
u_t|\Psi_{t-1} &\sim N(0,\sigma_t^2) && (3.15)\\
\sigma_t^2 &= \omega + \delta_1 D^*_{1,t} + \delta_2 D^*_{2,t} + \delta_4 D^*_{4,t} + \delta_5 D^*_{5,t} + \delta_6 D_{\mathrm{HOL},t} + \delta_7 D_{\mathrm{XMS},t}\\
&\quad + \delta_8 D_{\mathrm{EMP},t} + \delta_9 D_{\mathrm{PPI},t} + \delta_{10} D_{\mathrm{CPI},t} + \delta_{11} D_{\mathrm{FF},t}\\
&\quad + \delta_{12} D_{\mathrm{EMP}_{t+1},t} + \delta_{13} D_{\mathrm{PPI}_{t+1},t} + \delta_{14} D_{\mathrm{CPI}_{t+1},t} + \delta_{15} D_{\mathrm{FF}_{t+1},t}\\
&\quad + \alpha u_{t-1}^2 + \gamma u_{t-1}^2 \mathtt{I}[u_{t-1} < 0] + \beta\sigma_{t-1}^2, && (3.16)
\end{aligned}$$

where Ψ_{t-1} is the information set containing all daily information up to day t, and the error terms u_t are assumed to follow a conditional normal distribution with mean zero and variance σ_t^2. Restricting $\gamma = 0$ results in the GARCH(1,1)-DA model, and further imposing $\delta_1 = \delta_2 = ... = \delta_{15} = 0$ renders the standard GARCH(1,1) model.[9]

We analyze two additional forecasting models. In both of these realized variance is modelled directly as in our ARFIMA model. However, standard time-series models are used, which, contrary to the ARFIMA models, do not incorporate the long-memory filter $(1-L)^d$. First, we consider an AR(p) model, given by

$$\phi(L)(y_t - \mu_t) = \varepsilon_t, \tag{3.17}$$

[9]The results for the standard GARCH models are not reported but are available upon request. See also Koopman *et al.* (2005) for more results on comparing daily GARCH models with ARFIMA models.

where $\phi(L)$ is again a $p-$th order lag polynomial. We use the following lag lengths; $p = 1,\ 5,\ 10,\ 20$. Using longer lag lengths allows the AR(p) model to mimic long-memory-type behaviour.

Second, we consider the Heterogeneous Autoregressive (HAR) realized volatility model proposed by Corsi (2004) which is given by

$$y_t = \mu_t + \phi_d y_{t-1}^{(1)} + \phi_w y_{t-1}^{(5)} + \phi_m y_{t-1}^{(20)} + \varepsilon_t \tag{3.18}$$

where $y_{t-1}^{(1)}$ is the previous-day realized volatility and where $y_{t-1}^{(5)}$ and $y_{t-1}^{(20)}$ are the (sum of) realized volatilities over the last week and month respectively. Note that the HAR-RV can be obtained from (3.17) by imposing restrictions on the autoregressive parameters in the lag polynomial $\phi(L)$. Both for the AR and HAR models μ_t can be modelled in the same way as for the ARFIMA models, i.e. including all or some of the non-linearities as well as specifying these as exogenous regressors. To save space and because the ARFI-DAXRL model was shown to provide the best description of the data, we will only show forecasting results for the versions that include all non-linearities. We use the same notation as before, e.g. AR10-DAXRL and HAR-DAXRL for the AR(10) and HAR models with seasonal dummies and (pre-)announcement dummies as part of μ_t and the terms involving lagged returns as exogenous variables.

3.5.2 Forecast evaluation

A number of out-of-sample forecast performance measures are used to evaluate and compare the various models. First, the quality of individual forecasts is assessed by regressing the observed h-day realized standard deviation on the corresponding forecast,

$$s_{t+h|t+1} = \sqrt{\sum_{j=1}^{h} s_{t+j}^2} = b_0 + b_1 \widehat{s}_{t+h|t} + \nu_t, \tag{3.19}$$

where b_0 and b_1 should be equal to 0 and 1, respectively, for the forecast to be unbiased and efficient.

Forecasts from two different models are compared directly by means of the encompassing regression

$$s_{t+h|t+1} = \sqrt{\sum_{j=1}^{h} s_{t+j}^2} = b_0 + b_1 \widehat{s}_{t+h|t,(1)} + b_2 \widehat{s}_{t+h|t,(2)} + \nu_t, \tag{3.20}$$

where $\widehat{s}_{t+h|t,(i)}$ is the forecast of model $i = 1, 2$ for the volatility from $t+1$ to $t+h$, made at the end of day t. In addition a number of popular error metrics are computed, namely the Mean Squared Prediction Error (MSPE; $MSPE = \frac{1}{R}\sum_{i=1}^{R}(s_{t+h|t+1} - \widehat{s}_{t+h|t})^2$ where R denotes the number of forecasts), the Mean Absolute Error (MAE; $MAE = \frac{1}{R}\sum_{i=1}^{R}|s_{t+h|t+1} - \widehat{s}_{t+h|t}|$), and the Heteroskedasticity-adjusted MSPE (HMSPE; $HMSPE = \frac{1}{R}\sum_{i=1}^{R}(1 - \frac{\widehat{s}_{t+h|t}}{s_{t+h|t+1}})^2$). In most cases we will focus the discussion on the MSPE results, with Patton (2006) showing this is the most reliable metric given that the forecast target is a proxy for volatility. Finally, we also report the Mean Error (ME; $ME = \frac{1}{R}\sum_{i=1}^{R} s_{t+h|t+1} - \widehat{s}_{t+h|t}$).[10]

3.5.3 Forecast comparison

To test for significant differences in forecast accuracy between competing models we apply the Model Confidence Set approach proposed by Hansen *et al.* (2005a). Given a set of forecasting models, $\mathcal{M}_0$, the goal of the Model Confidence Set (MCS) procedure is to identify the MCS set $\widehat{\mathcal{M}}^*_{1-\alpha} \subset \mathcal{M}_0$ which is the set of models constructed such that it will contain the "best" forecasting model, given a level of confidence α.

Starting with the full set of models $\mathcal{M} = \mathcal{M}_0$, the MCS procedure repeatedly tests the null hypothesis of equal forecasting accuracy

$$H_{0,\mathcal{M}}: \ E(d_{ij,t}) = 0, \ \text{for all } i, j \in \mathcal{M}$$

where $d_{ij,t}$ is the loss differential between models i and j in the set: $d_{ij,t} = L_{i,t} - L_{j,t}$ with L an appropriate loss function. The MCS procedure sequentially eliminates the worst performing model from $\mathcal{M}$ as long as the null is being rejected. This trimming of models is repeated until the null is accepted and the surviving set of models then form the Model Confidence Set, $\widehat{\mathcal{M}}^*_{1-\alpha}$.

Hansen *et al.* (2005a) discuss three different test statistics to test the null hypothesis $H_{0,\mathcal{M}}$. We follow Hansen *et al.* (2003) by applying the MCS procedure with two of these; the range statistic, T_R, and the semi-quadratic statistic, T_{SQ}. Both test statistics are based on the following t-statistics (using $\bar{d}_{ij} = \frac{1}{R}\sum_{i=1}^{R} d_{ij,t}$),

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\text{var}}(\bar{d}_{ij})}} \quad \text{for } i, j \in \mathcal{M} \tag{3.21}$$

[10] See also Andersen, Bollerslev, and Meddahi (2004, 2005) for recent discussions on issues involved in evaluating realized volatility forecasts.

and are given by

$$T_R \equiv \max_{i,j\in\mathcal{M}} |t_{ij}| \quad \text{and} \quad T_{SQ} \equiv \sum_{i,j\in\mathcal{M}} t_{ij}^2 \tag{3.22}$$

Following Hansen *et al.* (2005a) we implement the MCS procedure using the stationary block bootstrap of Politis and Romano (1994) with an average block length of 20 days and we choose MSPE for the loss function L.

The MCS procedure assigns p-values to each model in the initial set. For a given model $i \in \mathcal{M}_0$, the MCS p-value, $\widehat{p}_i$, is the threshold that determines whether the model belongs to the MCS set or not. It hold that $i \in \widehat{\mathcal{M}}^*_{1-\alpha}$ if and only if $\widehat{p}_i \geq \alpha$. We report results for confidence levels of 10% and 25%.

In addition to applying the MCS approach to identify the best performing models from a whole set of models, we also apply it pairwise to test for significant differences between just two competing models, like the Diebold and Mariano (1995) test of equal forecast accuracy.[11] Finally, we also use the MCS framework to test whether the forecasts from one model encompass those of another model, in the spirit of Harvey *et al.* (1998). In both cases the reported p-values are based on the T_{SQ} test.

Note that a number of issues need to be addressed when testing for significant differences in forecast error metrics between models. These include comparing forecast which are based on estimated parameters, the choice of estimation window and comparing forecasts from nested models, see West (2006) a.o. for a discussion. As noted by Hansen *et al.* (2005a), the MCS procedure is also potentially affected when forecasts are compared which are based on estimated parameters but it nevertheless seems informative to apply the MCS procedure here.

3.5.4 Forecast results

Results for the one-day ahead forecasts for all models are presented in Table 3.6. For the ARFI models that allow for structural change, as well as for the AR and HAR models, we only report results for the models that include all nonlinearities. Furthermore, for the AR

[11] The Diebold and Mariano (1995) test, however, relies on asymptotics whereas here we evaluate test statistics using a bootstrap methodology.

Table 3.6: Out-of-sample forecast evaluation, January 1998-December 2003 - standard deviation, one-day ahead

	ME	MSPE	MAE	HMSPE	b_0	b_1	R^2	$\tilde{R}^2$
Riskmetrics	−0.050	0.190	0.303	0.114	0.194	0.813	0.390	0.574
	(0.011)	[0.001]	[0.000]	[0.000]	(0.040)	(0.035)		[0.023]
GJR-G-DA(1,1)	−0.006	0.160	0.266	0.089	0.228	0.814	0.488	0.582
	(0.010)	[0.029]	[0.000]	[0.000]	(0.035)	(0.031)		[0.016]
ARFI	−0.003	0.143	0.246	0.073	−0.076	1.058	0.521	0.587
	(0.010)	[0.366]	[0.000]	[0.001]	(0.045)	(0.040)		[0.026]
ARFI-D	−0.005	0.142	0.243	0.069	−0.043	1.031	0.525	0.584
	(0.010)	[0.479]	[0.008]	[0.057]	(0.045)	(0.040)		[0.031]
ARFI-DA	−0.006	0.137	0.242	0.069	−0.017	1.009	0.538	0.585
	(0.010)	[0.891]	[0.000]	[0.017]	(0.046)	(0.041)		[0.033]
ARFI-DAR	−0.006	0.137	0.241	0.071	0.007	0.990	0.538	0.578
	(0.010)	[0.892]	[0.002]	[0.004]	(0.047)	(0.042)		[0.028]
ARFI-DARL	−0.009	0.136	0.236	0.068	0.074	0.935	0.546	0.562
	(0.009)	[0.930]	[0.033]	[0.008]	(0.050)	(0.044)		[0.137]
ARFI-DAXRL	−0.046	0.139	0.240	0.074	0.140	0.857	0.554	0.562
	(0.010)	[0.038]	[0.000]	[0.000]	(0.063)	(0.051)		[0.441]
ARFI-DXARL	−0.050	0.141	0.242	0.076	0.136	0.857	0.551	0.562
	(0.010)	[0.007]	[0.000]	[0.000]	(0.065)	(0.053)		[0.615]
SPARFI-DAXRL	−0.051	0.139	0.243	0.075	0.130	0.862	0.555	0.562
	(0.010)	[0.260]	[0.000]	[0.000]	(0.057)	(0.047)		[0.150]
HAR-DAXRL	0.010	0.136	0.229	0.062	0.184	0.861	0.558	0.562
	(0.009)	[0.800]	[0.330]	[0.072]	(0.056)	(0.049)		[0.031]
AR10-DAXRL	−0.005	0.136	0.230	0.063	0.196	0.840	0.562	
	(0.010)				(0.053)	(0.045)		

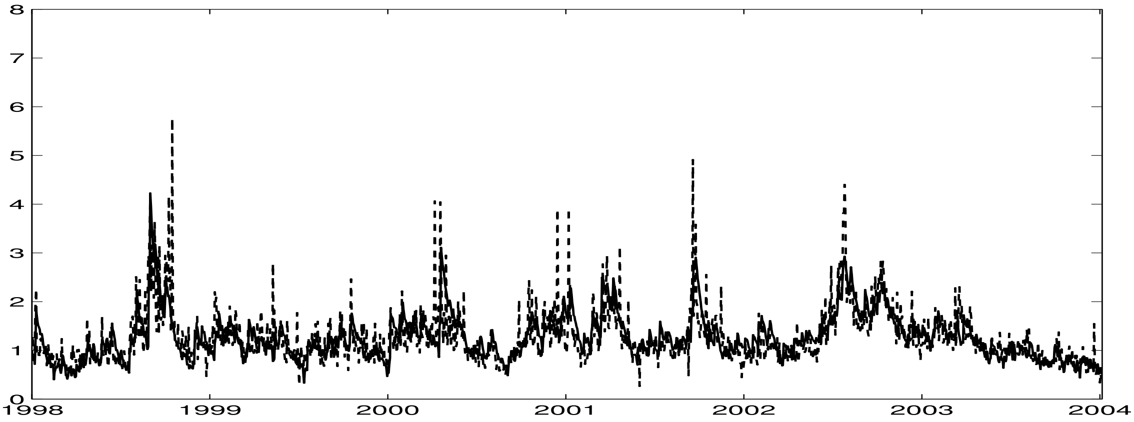
Notes: The table presents estimates of regressions of realized standard deviation for the S&P 500 on a constant and one-day ahead forecasts from different models. The regression is $s_{t+1} = b_0 + b_1\widehat{s}_{t+1|t,(1)}\ [+b_2\widehat{s}_{t+1|t,(2)}] + u_t$, where $\widehat{s}_{t+1|t,(i)}$ is the one-day ahead forecast of the realized standard deviation from model i. The forecast evaluation period covers January 2, 1998-December 31, 2003 ($R = 1510$). SPARFI-DA(X)RL refers to the full ARFI model including structural change, seasonal dummies, (pre-)announcement dummies and leverage effects as part of μ_t (as exogenous variables), ARFI-DA(X)RL refers to the ARFI model without structural change but including seasonal dummies, (pre-)announcement dummies and leverage effects as part of μ_t (as exogenous variables), ARFI-DAR refers to the model with seasonal dummies, (pre-)announcement dummies and only symmetric effects of the lagged absolute return as part of μ_t, ARFI-DA to the model with seasonal dummies, (pre-)announcement dummies but without the lagged absolute return, ARFI-D to the model with only seasonal dummies and ARFI to the model without any dummies or lagged absolute returns. HAR-DAXRL is the full HAR model and ARX10-DAXRL the full AR(10) model. Figures in brackets below b_j, $j = 0, 1, 2$ and ME are heteroskedasticity and autocorrelation-consistent standard errors. Figures in straight brackets below MSPE, MAE and HMSPE are MCS p-values of testing equal forecast accuracy using the T_{SQ} test, comparing the relevant model with the AR10-DAXRL model, where values below a certain confidence level (e.g. 5%) indicate that the AR10-DAXRL model is more accurate. Figures in straight brackets below the $\tilde{R}^2$ are MCS p-values of forecast encompassing tests in encompassing regressions, testing the null that the forecasts from the AR10-DAXRL model encompass the forecasts from the alternative model that is included in the regression.

models we only show the AR(10) model as it showed the best performance. We confirm the findings of earlier work (e.g. Koopman *et al.*, 2005), in that the ARFI models produce volatility forecasts that are superior to popular volatility models based on daily data. For example, the regression R^2 of the "basic" ARFI model is 52.1% compared to 39.0% for Riskmetrics. Including the leverage effect, seasonal dummies and (pre-)announcement dummies in the standard GARCH(1,1) model increases its regression R^2 to 48.8%, but this is still short of the R^2 of even the simplest ARFI model. The top panels in Figures 3.4 and 3.5 show the daily realized standard deviations for the out-of-sample period along with the one-day-ahead forecasts from the GJR-GARCH model and the ARFI-DAXRL model, respectively. It is seen that the GJR-GARCH forecasts track the general pattern, or low-frequency movements in volatility quite well. The main advantage of the long memory model is that, in addition, it captures a much larger part of the high-frequency movements in volatility.

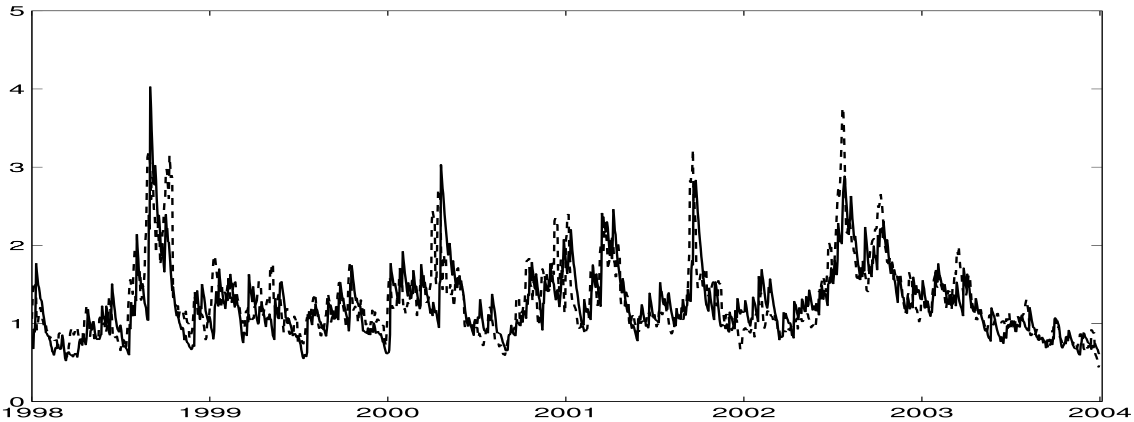
Among the ARFI models, the best forecast performance is attained by the SPARFI-DAXRL model, with a regression R^2 of 55.5% with the ARFI-DAXRL being a very close second with a R^2 of 55.4%, suggesting that incorporating structural changes adds little. The R^2-s clearly show an increasing pattern when adding non-linearities to the basic ARFI model. In general, allowing for the (pre-)announcement effects contributes most of all nonlinearities to the improvement in forecast performance over the linear ARFI model. Similarly important is the leverage effect. Including the leverage effect exogenously slightly improves over including the leverage effect in μ_t, the latter resulting in a regression R^2 of 54.6%. The MSPE, MAE and HMPSE show patterns which are less clear.

Despite the good performance of the (SP)ARFI-DAXRL model, it is not the best performing model overall. The HAR-DAXRL and AR10-DAXRL model both show slightly better results with the AR10-DAXRL being the best model with a R^2 of 56.2%. Figure 3.6 shows the one-day-ahead forecasts from the AR10-DAXRL model. The MCS p-values reported in the table allow for a comparison between the forecasts of the AR10-DAXRL model with those of all other models (on a one-to-one basis). The forecast encompassing p-values reported below the $\tilde{R}^2$ statistics in Table 3.6 show that the null that the forecasts from the AR10-DAXRL model encompass the forecasts from the Riskmetrics, GJR-GARCH as well as the ARFI models that do not allow for a leverage effect is rejected at conventional significance levels. Therefore, although judging from R^2 and the error metrics it could be suggested that explicitly modelling the long-memory component does

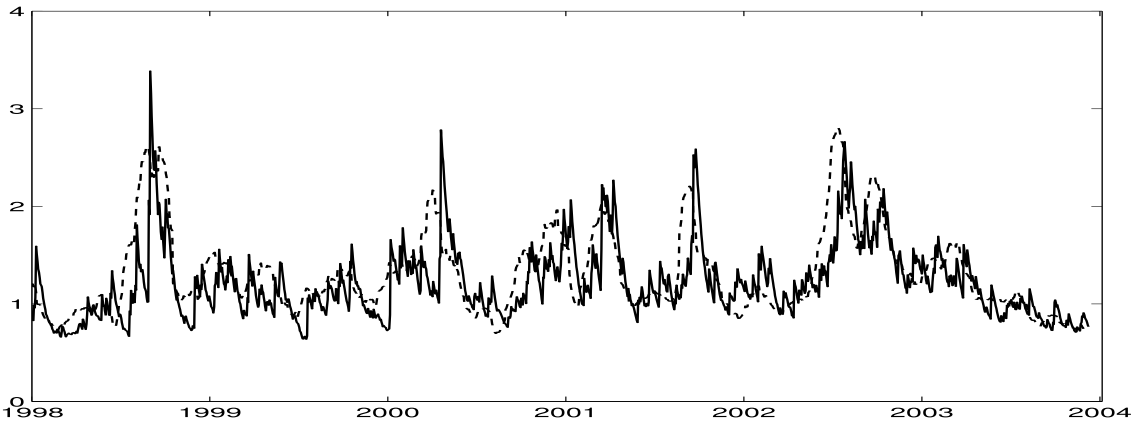
Figure 3.4: GJR-GARCH-DA(1,1) volatility forecasts



(a) 1-day



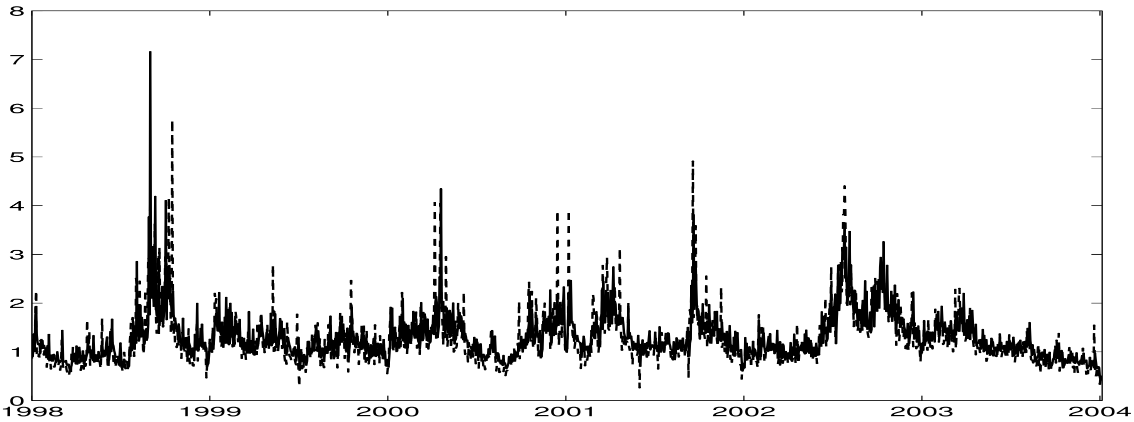
(b) 5-days



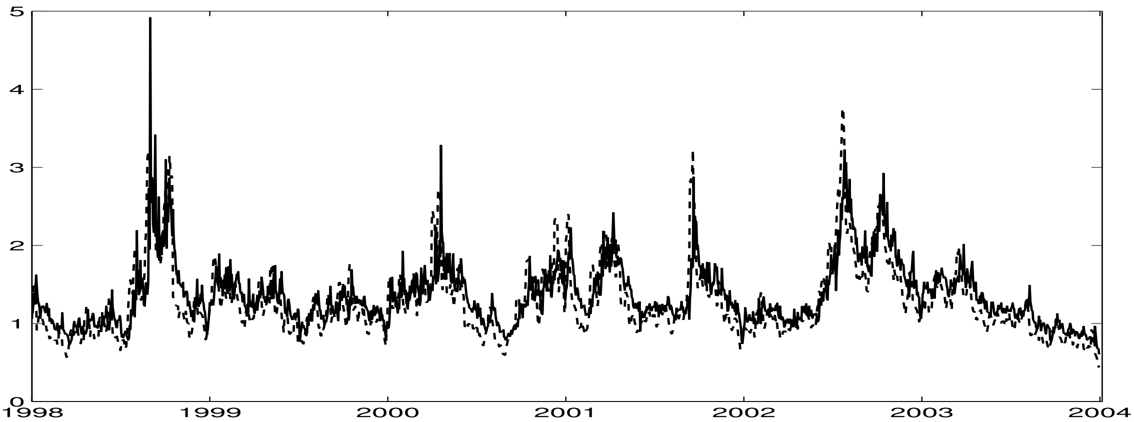
(c) 20-days

Notes: Out-of-sample forecasts $\widehat{s}_{t+h|t}$ of 1-day, 5-day and 20-day standard deviation obtained from a GJR-GARCH-DA(1,1) model (solid lines) and realizations $s_{t+h|t+1}$ (dashed lines) for the period from January 2, 1998, until December 29, 2003 (1510 observations for one-day ahead forecasts).

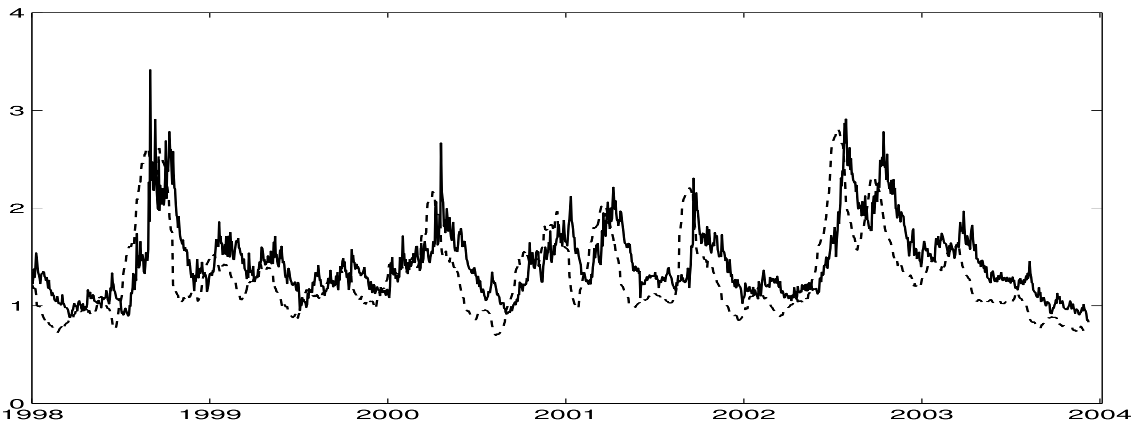
Figure 3.5: ARFI-DAXRL realized volatility forecasts



(a) 1-day



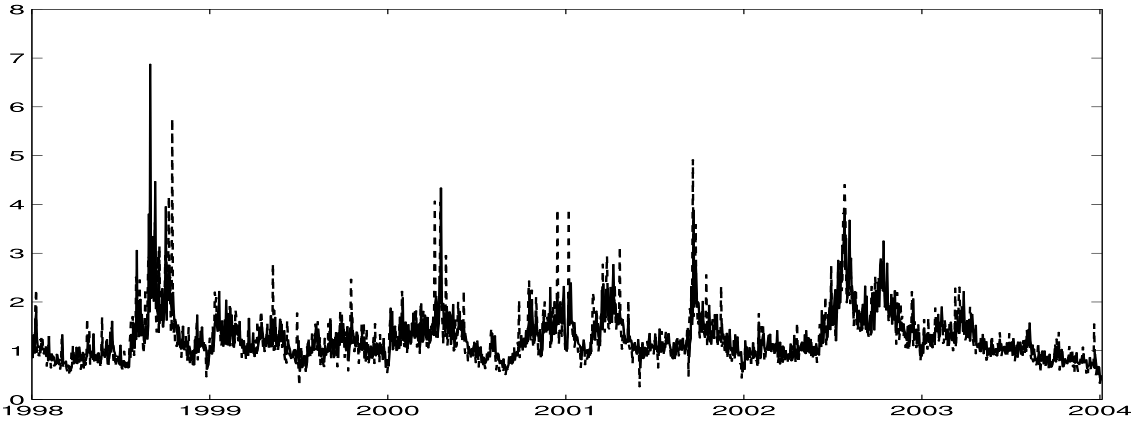
(b) 5-days



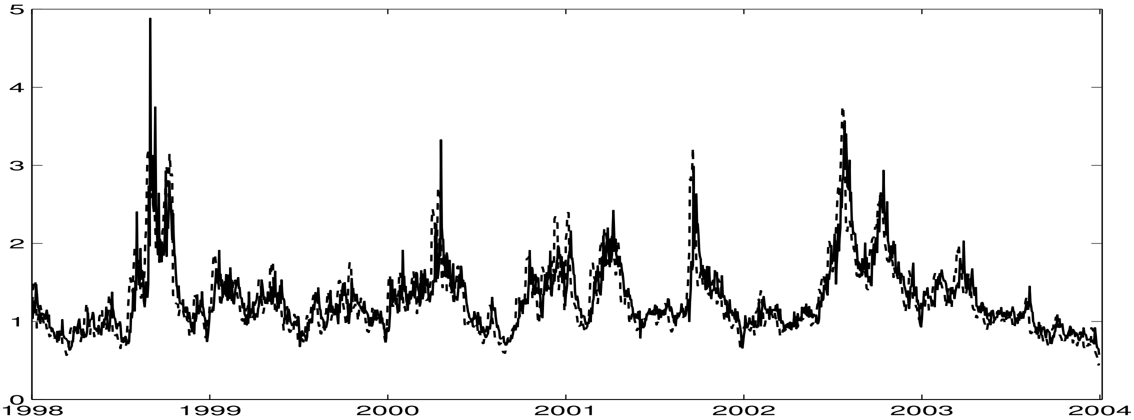
(c) 20-days

Notes: Out-of-sample forecasts $\widehat{s}_{t+h|t}$ of 1-day, 5-day and 20-day standard deviation obtained from the ARFI-DAXRL model (solid lines) and realizations $s_{t+h|t+1}$ (dashed lines) for the period from January 2, 1998, until December 29, 2003 (1510 observations for one-day ahead forecasts).

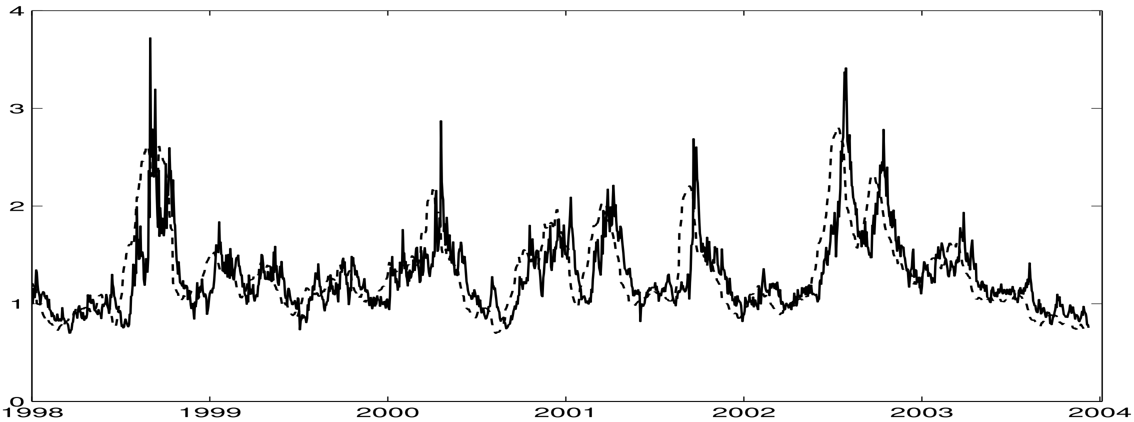
Figure 3.6: AR10-DAXRL realized volatility forecasts



(a) 1-day



(b) 5-days



(c) 20-days

Notes: Out-of-sample forecasts $\widehat{s}_{t+h|t}$ of 1-day, 5-day and 20-day standard deviation obtained from the AR10-DAXRL model (solid lines) and realizations $s_{t+h|t+1}$ (dashed lines) for the period from January 2, 1998, until December 29, 2003 (1510 observations for one-day ahead forecasts).

Table 3.7: Out-of-sample forecast evaluation, January 1998-December 2003 - standard deviation, five, ten and twenty-days ahead

	5-days ahead					10-days ahead					20-days ahead				
	MSPE	b_0	b_1	R^2	$\tilde{R}^2$	MSPE	b_0	b_1	R^2	$\tilde{R}^2$	MSPE	b_0	b_1	R^2	$\tilde{R}^2$
Riskmetrics	0.134	0.274	0.771	0.457	0.636	0.129	0.351	0.719	0.438	0.593	0.138	0.477	0.630	0.373	0.494
	[0.000]	(0.057)	(0.049)		[0.089]	[0.001]	(0.071)	(0.060)		[0.062]	[0.001]	(0.086)	(0.069)		[0.159]
GJR-G-DA(1,1)	0.116	0.299	0.788	0.535	0.640	0.113	0.335	0.777	0.508	0.598	0.118	0.378	0.763	0.443	0.503
	[0.006]	(0.052)	(0.045)		[0.003]	[0.021]	(0.069)	(0.061)		[0.030]	[0.057]	(0.087)	(0.075)		[0.029]
ARFI	0.096	−0.109	1.089	0.583	0.637	0.097	−0.086	1.074	0.531	0.592	0.109	0.032	0.985	0.405	0.493
	[0.037]	(0.066)	(0.057)		[0.056]	[0.080]	(0.089)	(0.076)		[0.042]	[0.228]	(0.128)	(0.105)		[0.050]
ARFI-D	0.094	−0.102	1.082	0.592	0.638	0.095	−0.079	1.068	0.538	0.592	0.108	0.040	0.978	0.410	0.494
	[0.081]	(0.064)	(0.055)		[0.058]	[0.150]	(0.088)	(0.075)		[0.036]	[0.255]	(0.126)	(0.104)		[0.066]
ARFI-DA	0.092	−0.099	1.078	0.600	0.638	0.094	−0.084	1.070	0.546	0.593	0.107	0.033	0.982	0.415	0.493
	[0.140]	(0.062)	(0.053)		[0.055]	[0.216]	(0.085)	(0.073)		[0.039]	[0.301]	(0.124)	(0.102)		[0.048]
ARFI-DAR	0.092	−0.095	1.076	0.599	0.637	0.094	−0.084	1.071	0.544	0.592	0.108	0.033	0.983	0.410	0.494
	[0.131]	(0.061)	(0.053)		[0.078]	[0.196]	(0.085)	(0.072)		[0.040]	[0.264]	(0.124)	(0.103)		[0.051]
ARFI-DARL	0.092	−0.048	1.031	0.597	0.635	0.095	−0.027	1.018	0.537	0.592	0.109	0.093	0.929	0.404	0.496
	[0.174]	(0.059)	(0.050)		[0.145]	[0.141]	(0.083)	(0.071)		[0.091]	[0.141]	(0.122)	(0.101)		[0.117]
ARFI-DAXRL	0.093	0.021	0.929	0.620	0.635	0.101	0.047	0.898	0.561	0.592	0.128	0.191	0.777	0.423	0.496
	[0.022]	(0.054)	(0.043)		[0.303]	[0.040]	(0.069)	(0.054)		[0.291]	[0.009]	(0.101)	(0.075)		[0.597]
ARFI-DXARL	0.094	0.011	0.933	0.618	0.635	0.103	0.034	0.902	0.558	0.592	0.131	0.174	0.783	0.421	0.496
	[0.026]	(0.055)	(0.043)		[0.317]	[0.023]	(0.071)	(0.055)		[0.293]	[0.007]	(0.104)	(0.076)		[0.518]
SPARFI-DAXRL	0.098	0.039	0.911	0.607	0.635	0.111	0.113	0.845	0.529	0.592	0.151	0.331	0.674	0.367	0.497
	[0.015]	(0.051)	(0.041)		[0.313]	[0.010]	(0.071)	(0.056)		[0.456]	[0.008]	(0.110)	(0.080)		[0.983]
HAR-DAXRL	0.088	0.131	0.906	0.620	0.635	0.091	0.192	0.864	0.570	0.592	0.107	0.329	0.765	0.460	0.494
	[0.356]	(0.054)	(0.046)		[0.190]	[0.289]	(0.066)	(0.056)		[0.304]	[0.077]	(0.087)	(0.073)		[0.833]
AR10-DAXRL	0.086	0.152	0.877	0.635		0.088	0.205	0.842	0.592		0.100	0.296	0.781	0.492	
		(0.050)	(0.042)				(0.062)	(0.052)				(0.080)	(0.064)		

Notes: The table presents estimates of regressions of realized standard deviation for the S&P 500 on a constant and either five, ten or twenty-days ahead forecasts from different models. The regression is $s_{t+h|t+1} = b_0 + b_1 \widehat{s}_{t+h|t,(1)}$ $[+ b_2 \widehat{s}_{t+h|t,(2)}] + u_t$, where $\widehat{s}_{t+h|t,(i)}$ is the forecast of h-days realized standard deviation with $h \in \{5, 10, 20\}$ from model i. The forecast evaluation period covers January 2, 1998-December 31, 2003 ($R = 1510$). See Table 3.6 for further details.

not seem necessary to capture the forecasting performance of realized volatility-based models, the ARFI models do have some incremental information not already contained in the nonlinear AR(10) model. For example, the resulting regression R^2 of optimally (with hindsight) combining the ARFI model with the AR10-DAXRL model is 58.7%.

The results for 5-day, 10-day and 20-day ahead forecasts of realized standard deviation are summarized in Table 3.7. The benefits of modelling the nonlinearities are also explicit for longer horizons. Comparing the ARFI and ARFI-DAXRL models, the regression R^2 increases from 58.3% to 62.0% for the 5-day horizon, from 53.1% to 56.1% for the 10-day horizon, and from 40.5% to 42.3% for the 20-day horizon. For the 20-day horizon the GARCH model with non-linearities now has a higher R^2 than the ARFI models. The AR(10) model that includes all non-linearities remains the best forecasting model at all horizons.

To compare forecast accuracy among multiple models simultaneously we apply the MCS procedure to the models listed in Table 3.6 with Table 3.8 showing the results. For all horizons, the model confidence sets consists of several models and, except for the 20-day horizon, never include the models based on daily returns or the linear ARFI model. The confidence sets always include models that incorporate non-linearities. Several of these, in particular the AR10-DAXRL model, belong to the model confidence for every horizon. Table 3.9 compares different models, all including the full array of nonlinearities, as well as the two models that are based on daily returns. The model confidence sets contain only a few models and always contain the AR(10) model.

3.5.5 Value-at-Risk forecasts

As an alternative method for evaluating the forecasts for the different models from a more economic point of view, we consider Value-at-Risk (VaR) estimates which are constructed using the volatility forecasts from the different models, see also Giot and Laurent (2004). Under the 1998 Market Risk Amendment (MRA) to the Basle Capital Accord, commercial banks are required to reserve capital to cover their market risk exposure. The required market risk capital for time $t+1$ (MRC_{t+1}) is determined by a bank's 99% VaR estimate calculated for a 10-day holding period (VaR_t^{10}) and is defined as the higher of the previous day's VaR estimate or the average of the estimates over the last sixty business days times a multiplication term

$$MRC_{t+1} = \max(VaR_t^{10}, S_t \times \frac{1}{60}\sum_{i=0}^{59} VaR_{t-i}^{10}), \tag{3.23}$$

Table 3.8: Model Confidence Sets, January 1998-December 2003 - standard deviation, one, five, ten and twenty-days ahead

	1-day ahead			5-days ahead			10-days ahead			20-days ahead		
	MSPE	T_R	T_{SQ}	MSPE	T_R	T_{SQ}	MSPE	T_R	T_{SQ}	MSPE	T_R	T_{SQ}
Riskmetrics	0.190	0.000	0.010	0.134	0.002	0.003	0.129	0.000	0.005	0.138	0.014	0.029
GJR-G-DA	0.160	0.090	0.118*	0.116	0.025	0.028	0.113	0.030	0.064	0.118	0.207*	0.257**
ARFI	0.142	0.090	0.331**	0.096	0.095	0.097	0.097	0.092	0.120*	0.109	0.207*	0.259**
ARFI-D	0.142	0.272**	0.563**	0.094	0.277**	0.230**	0.095	0.469**	0.287**	0.108	0.265**	0.290**
ARFI-DA	0.137	0.997**	0.994**	0.092	0.442**	0.321**	0.094	0.469**	0.368**	0.107	0.319**	0.300**
ARFI-DAR	0.137	0.997**	0.994**	0.092	0.277**	0.321**	0.094	0.469**	0.330**	0.108	0.265**	0.290**
ARFI-DAL	0.136	1.000**	1.000**	0.092	0.442**	0.321**	0.095	0.469**	0.329**	0.109	0.207*	0.259**
ARFI-DAXL	0.139	0.272**	0.639**	0.093	0.277**	0.290**	0.101	0.092	0.120*	0.128	0.067	0.126*
ARFI-DXAL	0.141	0.090	0.333**	0.094	0.095	0.159*	0.103	0.030	0.087	0.131	0.053	0.071
SPARFI-DAXL	0.139	0.272**	0.598**	0.098	0.095	0.075	0.111	0.030	0.035	0.151	0.014	0.014
HAR-DAXL	0.136	0.997**	0.994**	0.088	0.442**	0.356**	0.091	0.469**	0.368**	0.107	0.265**	0.290**
AR10-DAXL	0.136	0.997**	0.994**	0.086	1.000**	1.000**	0.088	1.000**	1.000**	0.100	1.000**	1.000**

Notes: The table presents model confidence sets for the selection of models reported in Table 3.6 for one, five, ten or twenty-days ahead forecasts. For each forecast horizon the MSPE is reported as well MCS p-values according to the test statistics T_R and T_{SQ}. Two stars indicate that a model belongs to the model set $\widehat{\mathcal{M}}^*_{0.25}$ whereas models with one stars belong to $\widehat{\mathcal{M}}^*_{0.10}$. The forecast evaluation period covers January 2, 1998-December 31, 2003 ($R = 1510$).

Table 3.9: Model Confidence Sets, January 1998-December 2003 - standard deviation, one, five, ten and twenty-days ahead

	1-day ahead			5-days ahead			10-days ahead			20-days ahead		
	MSPE	T_R	T_{SQ}	MSPE	T_R	T_{SQ}	MSPE	T_R	T_{SQ}	MSPE	T_R	T_{SQ}
Riskmetrics	0.190	0.001	0.002	0.134	0.002	0.000	0.129	0.001	0.004	0.138	0.010	0.002
GJR-G-DA	0.160	0.058	0.037	0.116	0.020	0.006	0.113	0.036	0.026	0.118	0.169*	0.091
ARFI-DAXRL	0.139	0.221*	0.377**	0.093	0.102*	0.075	0.101	0.073	0.056	0.128	0.020	0.027
SPARFI-DAXRL	0.139	0.221*	0.377**	0.098	0.088	0.026	0.111	0.036	0.018	0.151	0.010	0.002
AR1-DAXRL	0.155	0.058	0.074	0.142	0.020	0.002	0.169	0.001	0.004	0.195	0.010	0.000
SPAR1-DAXRL	0.157	0.001	0.007	0.157	0.002	0.000	0.180	0.000	0.000	0.197	0.001	0.000
AR10-DAXRL	0.136	0.786**	0.800**	0.086	1.000**	1.000**	0.088	1.000**	1.000**	0.100	1.000**	1.000**
SPAR10-DAXRL	0.138	0.694**	0.633**	0.094	0.102*	0.058	0.104	0.036	0.026	0.136	0.010	0.007
HAR-DAXRL	0.136	1.000**	1.000**	0.088	0.353**	0.364**	0.091	0.304**	0.297**	0.107	0.169*	0.091
SPHAR1-DAXRL	0.139	0.694**	0.633**	0.099	0.049	0.012	0.112	0.009	0.008	0.142	0.010	0.002

Notes: The table presents model confidence sets for different models, each with a '-DAXRL' specification and the Riskmetrics and GJR-GARCH(1,1)-DA model for one, five, ten or twenty-days ahead forecasts. For each forecast horizon the MSPE is reported as well MCS p-values according to the test statistics T_R and T_{SQ}. Two stars indicate that a model belongs to the model set $\widehat{\mathcal{M}}^*_{0.25}$ whereas models with one stars belong to $\widehat{\mathcal{M}}^*_{0.10}$. The forecast evaluation period covers January 2, 1998-December 31, 2003 ($R = 1510$).

using the one-step ahead volatility forecasts from the different models.[12]

Under the MRA, the accuracy of a bank's VaR estimates is judged by the accuracy of the 1-day ahead 99% VaR estimates[13]. We evaluate the accuracy of these estimates using the interval forecast evaluation techniques proposed by Christoffersen (1998) and the method set forth by Lopez (1999) which uses regulatory loss functions. The Christoffersen method is a test for conditional coverage, jointly testing the hypothesis that the percentage of times that the actual loss on a portfolio exceeds the VaR estimate (denoted by a VaR 'exception') equals one minus the significance level used in the VaR calculation (unconditional coverage) together with the hypothesis of serial independence for these VaR exceptions (independence).

Defining the indicator variable I_{t+1} as

$$I_{t+1} = \begin{cases} 1 & \text{if } r_{t+1} < VaR_t^1 \\ 0 & \text{if } r_{t+1} \geq VaR_t^1, \end{cases} \tag{3.24}$$

where r_{t+1} is the return over day $t+1$ and VaR_t^1 is the 100(1-α)% VaR estimate for day $t+1$ made on day t, a VaR exception corresponds with $I_{t+1} = 1$. The likelihood of observing x exceptions in a series of length R under the null hypothesis of accurate unconditional coverage is given by $L_0 = \alpha^x(1-\alpha)^{R-x}$. The corresponding likelihood under the alternative is $L_1 = \hat{\alpha}^x(1-\hat{\alpha})^{R-x}$, where $\hat{\alpha}$ is the "observed" probability of an exception, $\hat{\alpha} = x/R$. The null hypothesis of correct unconditional coverage can now be tested by means of the standard likelihood ratio statistic

$$LR_{uc} = 2[\log(L_1) - \log(L_0)], \tag{3.25}$$

which has an asymptotic $\chi^2(1)$ distribution.

The test statistic for the test of independence is also $\chi^2(1)$ distributed and is given by

$$LR_{ind} = 2[\log(L_2) - \log(L_1)], \tag{3.26}$$

where L_1 is the likelihood function under independence as given above, and L_2 is the likelihood function under the alternative of first-order Markov dependence. The latter is

[12] The value of the multiplication factor S_t is determined by the number of times that the actual losses on a portfolio exceed the 1-day 99% VaR estimates (so called VaR-exceptions). Three zones for an increasing number of exceptions are distinguished and the value of S_t increases from a minimum value of 3 to a possible maximum value of 4 across the different zones, see the Basle Committee on Banking Supervision (1996) for more details. In the evaluation of the VaR estimates we, however, fix S_t to the value 3. Note further that under the conditions of the MRA the VaR estimates are evaluated in dollar terms whereas we will consider the VaR in percentage terms.

[13] It is suggested by the Basle committee to use an evaluation period of at least 250 business days. Here we use the entire forecast evaluation sample.

given by $L_2 = (1-\pi_{01})^{R_{00}}\pi_{01}^{R_{01}}(1-\pi_{11})^{R_{10}}\pi_{11}^{R_{11}}$ where R_{ij} is the number of observations with value i followed by value j.

Testing correct conditional coverage boils down to testing correct unconditional coverage and independence simultaneously. The likelihood functions under the null and alternative are given by L_0 and L_2, respectively and, hence, the likelihood ratio statistic for correct conditional coverage is simply the sum of the two previous statistics,

$$LR_{cc} = LR_{uc} + LR_{ind}, \tag{3.27}$$

which is asymptotically $\chi^2(2)$ distributed.

The alternative method proposed by Lopez (1999) for evaluating VaR forecasts is based on the use of loss functions that are more closely related to the regulatory VaR requirements. By choosing a specific loss function, one can assign a numerical score to each individual VaR estimate that reflects specific concerns that one may have. For example, it seems natural to not only consider the number of VaR exceptions but also the magnitude of exceptions since the latter can be quite substantial. Therefore, we consider the loss function[14]

$$C_{t+1} = \begin{cases} 1 + (r_{t+1} - VaR_t^1)^2 & \text{if } r_{t+1} < VaR_t^1 \\ 0 & \text{if } r_{t+1} \geq VaR_t^1, \end{cases} \tag{3.28}$$

which includes the squared magnitude of the VaR exception $(r_{t+1} - VaR_t^1)^2$. Given a sample of R VaR estimates the total loss for the sample is given by $C = \sum_{i=1}^{R} C_{t+1}$.

To assess the relative performance of each volatility forecasting model, we compute for each model the average and standard deviation of the capital requirement MRC_{t+1} over the forecast evaluation period. Furthermore, we determine the unconditional coverage, $\hat{\alpha}$, together with the interval evaluation test statistics. Finally, we compute the total loss, C, for the sample of R one-day VaR estimates as well as the average score (defined as the total score divided by the number of exceptions) and the maximum daily score. The results are presented in Table 3.10.

We first of all observe that the average capital requirement is comparable across the different models. However, the long memory models typically have considerably less fluctuation in the required level of capital. This is confirmed graphically by Figure 3.7, which shows how the capital requirement evolves over time for the Riskmetrics, GJR-GARCH(1,1)-DA and ARFI-DAXRL models.

[14]Lopez (1999) discusses several possible loss functions one of which is the binomial loss function given in (3.24). The loss function given in the main text is in line with the guidelines by the Basle Committee which states that both the number as well the magnitude of exceptions are a matter of concern to regulators.

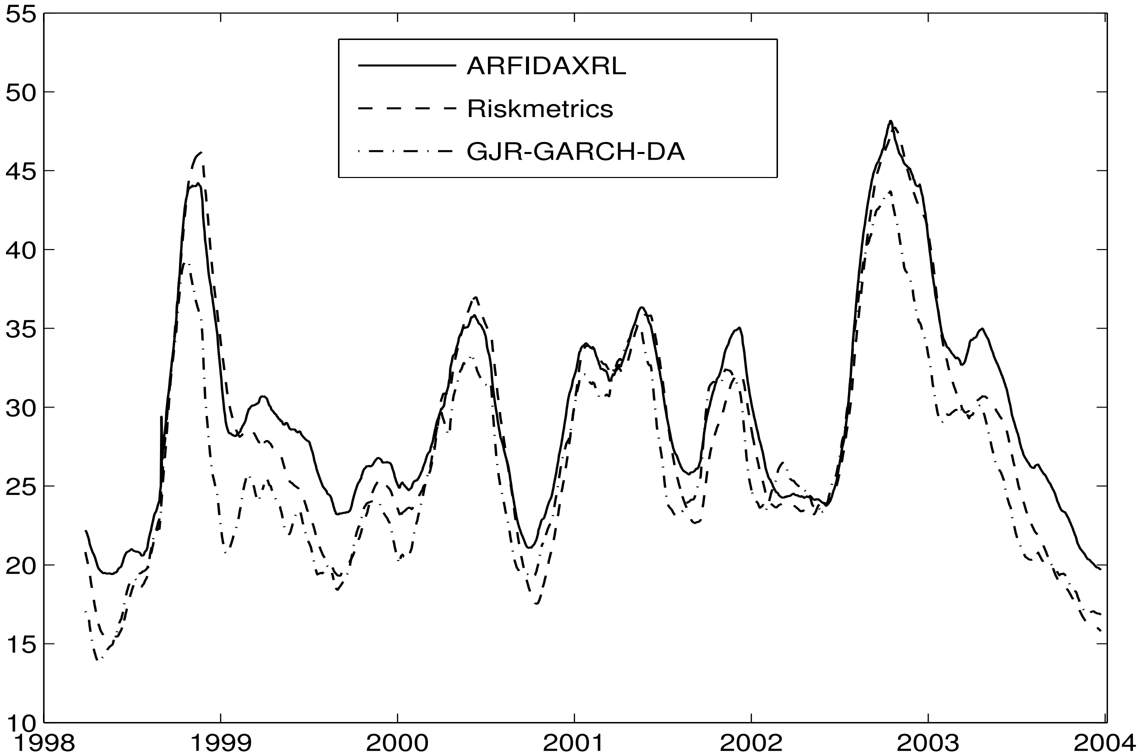
Table 3.10: Value-at-Risk evaluation for the out-of-sample period January 1998-December 2003

	MRC-μ	MRC-σ	$\hat{\alpha}$	LR_{uc}	LR_{ind}	LR_{cc}	C	$\bar{C}$	$\max(C_t)$
Riskmetrics	28.007	7.635	0.019	8.892	0.366	9.258	73.428	2.622	16.842
				(0.003)	(0.545)	(0.010)			
GJR-G-DA(1,1)	26.288	6.412	0.021	12.966	1.301	14.266	64.595	2.084	8.680
				(0.000)	(0.254)	(0.001)			
ARFI	27.360	5.569	0.016	4.494	0.723	5.217	53.464	2.228	11.584
				(0.034)	(0.395)	(0.074)			
ARFI-D	27.389	5.617	0.015	3.599	0.712	4.311	54.869	2.386	14.121
				(0.058)	(0.399)	(0.116)			
ARFI-DA	27.430	5.619	0.016	4.494	0.776	5.270	54.353	2.265	12.763
				(0.034)	(0.378)	(0.072)			
ARFI-DAR	27.411	5.590	0.017	6.537	0.912	7.448	61.123	2.351	17.095
				(0.011)	(0.340)	(0.024)			
ARFI-DARL	27.657	5.796	0.018	7.677	0.984	8.661	63.079	2.336	19.891
				(0.006)	(0.321)	(0.013)			
ARFI-DAXRL	29.697	6.585	0.015	3.599	0.712	4.311	50.652	2.202	15.244
				(0.058)	(0.399)	(0.116)			
ARFI-DXARL	29.870	6.538	0.016	4.494	0.776	5.270	52.049	2.169	15.902
				(0.034)	(0.378)	(0.072)			
HAR-DAXRL	29.619	5.881	0.017	6.537	0.912	7.448	55.398	2.131	16.339
				(0.011)	(0.340)	(0.024)			
ARX10-DAXRL	27.591	6.701	0.019	8.892	1.059	9.951	57.782	2.064	14.629
				(0.003)	(0.303)	(0.007)			

Notes: The table presents results for VaR estimates generated under the conditions of the Basle Committee MRA for the forecast evaluation period January 2, 1998-December 31, 2003 ($R = 1510$). The first two columns show the average and standard deviation of the required capital to cover market risk exposure (in percentage terms). Column 3 shows the average percentage number of exceptions defined as $\hat{\alpha} = x/R$ where x is the number of exceptions. Columns 4-6 show the interval forecast evaluation test statistics of correct unconditional coverage (uc), independence (ind) and correct conditional coverage (cc) (p-values are between brackets). Columns 7-9 give the total score C based on (3.28), the average score, $\bar{C} = C/x$ and the maximum individual score (all in percentage terms).

All models have a higher unconditional coverage than expected, leading to strong rejections of the null of correct unconditional coverage in all cases. By contrast, the null of independence is not rejected for any of the models under consideration. Based on the quadratic magnitude loss function, the nonlinear realized volatility models again perform

Figure 3.7: Market Risk Capital



Notes: The graph shows the required capital (in percentage terms) to cover market risk exposure which is calculated as $MRC_{t+1} = \max(VaR_t^{10}, S_t \times \frac{1}{60}\sum_{i=0}^{59} VaR_{t-i}^{10})$ based on volatility forecasts from the Riskmetrics, GJR-GARCH(1,1)-DA and ARFI-DAXRL models from January 2, 1998, until December 29, 2003 (1510 observations). VaR_t^{10} is the 99% VaR estimate for a 10-day holding period. The first sixty 1-day VaR estimates were used to construct the initial history needed to calculate MRC_{t+1}.

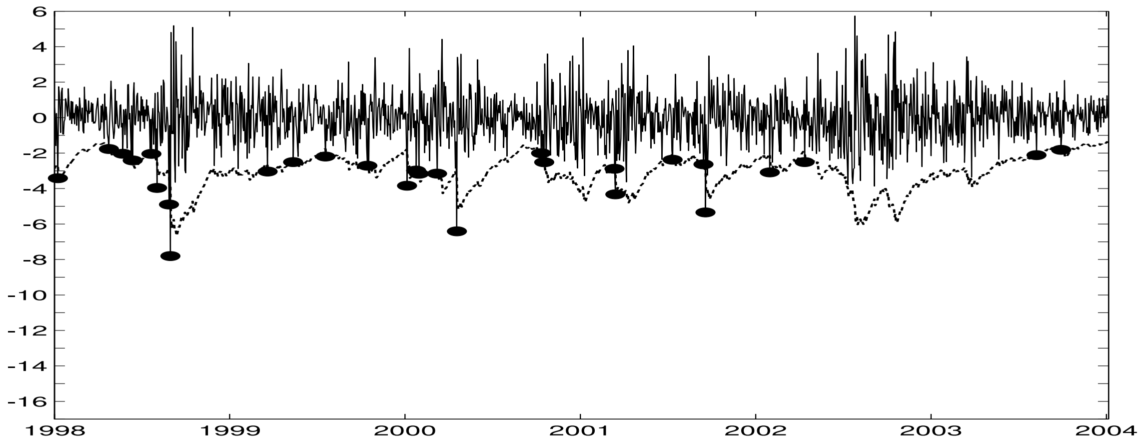
well when compared to the GARCH type models. Figure 3.8 shows the 1-day VaR estimates from the latter two models together with the AR10-DAXRL model.

3.6 Concluding remarks

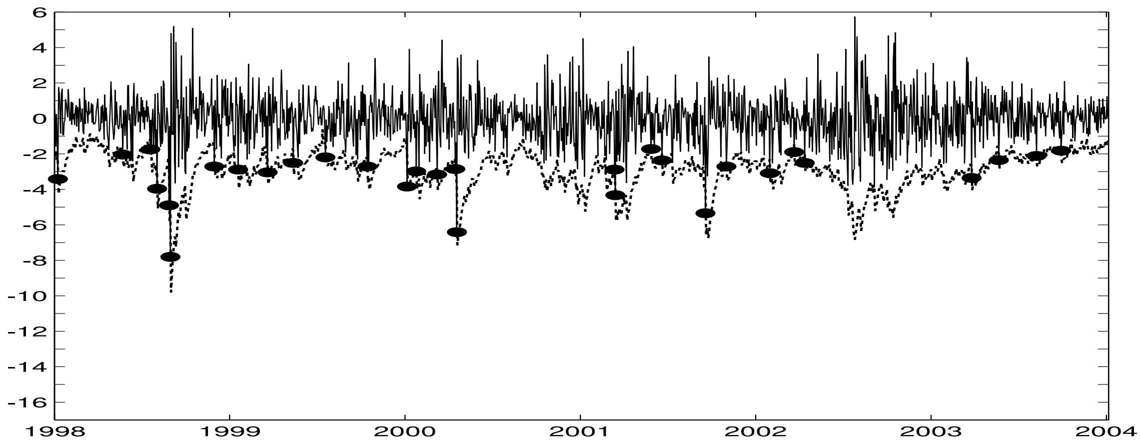
In this chapter we propose a nonlinear long-memory time series model for realized volatility that incorporates all well-known stylized facts from the (GARCH) volatility literature, in particular level shifts, day-of-the-week and holiday effects, announcement effects and leverage effects. The model, as well as several restricted versions, are estimated for the S&P 500 index.

The in-sample results show that all nonlinearities are highly significant and improve

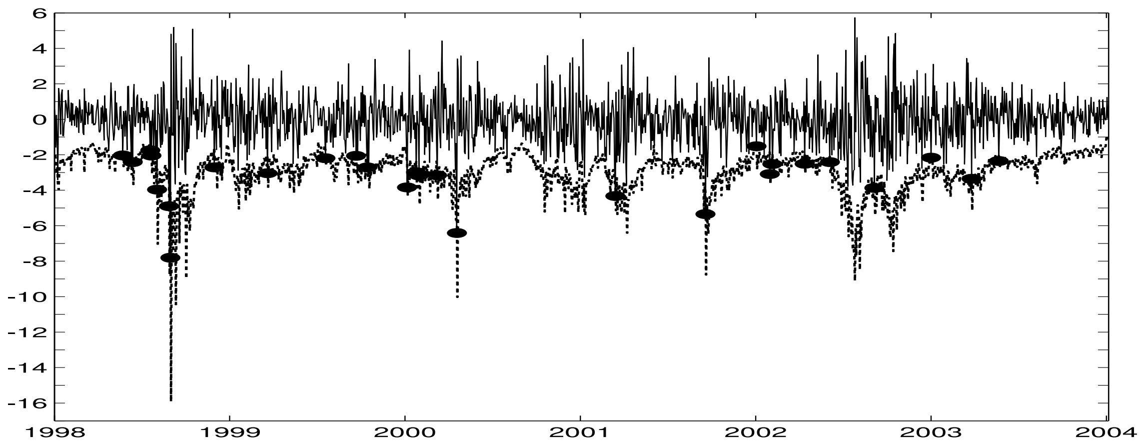
Figure 3.8: Value-at-Risk estimates



(a) Riskmetrics



(b) GJR-GARCH(1,1)-DA



(c) AR10-DAXRL

Notes: Realized returns (solid line) and 1-day 99% Value-at-Risk estimates based on volatility forecasts for the Riskmetrics, GJR-GARCH(1,1)-DA and AR10-DAXRL models (dotted lines) for the period from January 2, 1998, until December 29, 2003 (1510 observations). Black dots indicate VAR exceptions.

the description of the data. The out-of-sample results show that for forecasting horizons of up to 20 days, accounting for these nonlinearities improves the forecast performance compared to a linear ARFI model. Adding the same non-linearities to simpler autoregressive models is shown to lead to similar improvements for short horizons and more substantial improvements for longer horizons, indicating that incorporating nonlinearities has considerable benefits.

The most important nonlinearities for the S&P 500 index are (pre-)announcement effects and the leverage effect. The best way to incorporate the effects of lagged daily returns is to include them as exogenous regressors, i.e. outside the long memory filter in the case of the ARFI model. Not important for the forecast performance is to model the level shifts for the S&P 500 index.

Chapter 4

Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data

But which frequency to use?

4.1 Introduction

The work of Andersen and Bollerslev (1998a) has triggered a vast amount of research on the use of high-frequency data to measure, model and forecast volatility of financial asset returns. Most empirical studies on this topic of 'realized volatility' focus exclusively on the variance of individual asset returns, see Andersen *et al.* (2001a), Andersen *et al.* (2001b), Areal and Taylor (2002), Thomakos and Wang (2003), Martens *et al.* (2004), Pong *et al.* (2004), and Koopman *et al.* (2005), among others. Many financial applications such as risk management and portfolio construction, however, require estimates or forecasts of the entire covariance matrix, making covariances or correlations between returns on different assets at least as important. Yet only limited (empirical) research has addressed the merits of high-frequency data for potential economic or forecasting gains in a multivariate context. Andersen *et al.* (2003a) use a vector autoregressive (VAR) framework for the daily realized variances and covariance of two exchange rates (DEM/USD and YEN/USD) based on 30-minute returns, but they only consider the statistical accuracy of (co-)variance forecasts. Fleming *et al.* (2003) use five-minute returns on three actively traded futures contracts (S&P 500 index, Treasury bonds, and gold) to show that a mean-variance efficient investor would be willing to pay 50 to 200 basis points per annum for being able to use daily covariance matrix forecasts based on high-frequency intraday returns instead of daily returns. Similarly, Liu (2004) constructs and assesses the performance of the minimum variance portfolio and the minimum tracking error portfolio (tracking the S&P 500 index) using five-minute returns for the 30 Dow Jones index constituents.

These three studies have in common that they motivate the selected intraday sampling frequency as a trade-off between accuracy and potential biases due to market microstructure effects. The sensitivity of the results to the choice of sampling frequency used in constructing realized covariances is not investigated though. Martens (2004) demonstrates that non-trading, non-synchronous trading, and bid-ask bounce are indeed crucial determinants of the optimal sampling frequency that minimizes the Mean Squared Error (MSE) for measuring, and hence forecasting, the covariance matrix. The MSE is the sum of the squared bias and the variance of the realized (co-)variance. High sampling frequencies lead to a potentially large upward bias in realized variances due to bid-ask bounce and to a substantial downward bias in realized covariances due to non-synchronous trading. On the other hand, the variance of both realized variances and realized covariances usually decreases with higher sampling frequencies. As the degree of non-trading, non-synchronous trading, and bid-ask bounce varies widely across assets, the appropriate sampling frequency in a particular application needs to be investigated carefully.

In this chapter we examine the economic significance of determining the optimal sampling frequency, in the context of constructing mean-variance efficient portfolios from the individual constituents of the S&P 100 index. Our analysis builds on the framework developed in Fleming *et al.* (2001, 2003). In particular, we consider a risk-averse investor who constructs minimum variance portfolios and minimum tracking error portfolios with daily rebalancing, where portfolio risk is minimized either globally or subject to a fixed target return. We focus on pure volatility-timing strategies, in the sense that the portfolio weights are determined exclusively by forecasts of the daily conditional covariance matrix, which in turn is constructed using the realized covariance matrix with the sampling frequency of intraday returns ranging from one minute to 130 minutes. The economic value of using the optimal sampling frequency is assessed by comparing portfolio performance across this range of sampling frequencies. In particular, we consider the fee that the investor would be willing to pay to switch from one frequency to another.

We also examine how different bias- and variance-reduction techniques affect the choice of sampling frequency. First, we explore the usefulness of the two time-scales estimator proposed by Zhang *et al.* (2005),[1] which combines the realized covariance matrix constructed

[1] Zhang *et al.* (2005) focus solely on estimating the variance but here we apply their approach to covariances as well, as suggested by Zhang (2006).

using subsampling at a certain frequency and the realized covariance matrix constructed using the highest possible sampling frequency. Subsampling makes use of the fact that, for example, five-minute returns for a trading session starting at 9:30 could not only be measured using the intervals 9:30-9:35, 9:35-9:40, ..., but also using 9:31-9:36, 9:36-9:41, ..., etc. As explained in more detail below, subsampling can be used to reduce the variance of the realized covariance estimator. However, subsampling still renders a biased estimate of the true integrated covariance matrix with the bias being a function of the covariance matrix of the microstructure noise component in the intraday returns. The two time-scales estimator attempts to correct for this bias. Second, following the idea of Scholes and Williams (1977) for estimating (illiquid) stock betas, we investigate the merits of using leads and lags in measuring the realized covariances. For all these methods we also consider the effects of transaction costs and the holding period or portfolio rebalancing frequency.

Our main findings are as follows. For both minimum variance and minimum tracking error portfolios, using daily conditional covariance matrix forecasts based on high-frequency intraday returns instead of daily returns considerably improves portfolio performance. For the global minimum risk portfolios, the optimal sampling frequency for the S&P 100 constituents ranges between 30 and 65 minutes, considerably lower than the popular five-minute frequency. The same result occurs for minimum variance portfolios subject to a target return. Here, the Sharpe ratio increases from 0.6 to 0.8 going from daily to intraday returns, and a risk-averse investor would be willing to pay between 150 and 400 basis points per year to capture this gain in portfolio performance. In contrast, for the minimum tracking error portfolio subject to a target return the optimal sampling frequency appears to be much higher at one to two minutes. The performance gains compared to the use of daily returns are substantial, with the information ratio increasing from 0.1 to 0.4. The fee a risk-averse investor might pay for this enhanced performance ranges between 100 and 180 basis points per year.

The above findings are robust to the use of the two time-scales estimator and the lead-lag bias correction procedure. Both of these techniques marginally improve the performance for the minimum variance portfolios and the minimum tracking error portfolios. However, selecting the appropriate sampling frequency appears to be much more important than choosing between different bias- and variance-reduction techniques for the realized covariance matrices.

For the target return portfolios we find that turnover is lower when using intraday data, hence in the presence of transaction costs an investor is willing to pay even more for covariance matrix forecasts based on high-frequency data. The opposite is true for the target excess return portfolios. Lowering the rebalancing frequency from daily to weekly or monthly obviously reduces transaction costs, while at the same time having a similar or even better performance. Reducing the rebalancing frequency in the presence of transaction costs is especially beneficial for the minimum tracking error portfolios based on high-frequency data.

The issue of sampling frequency in the presence of market microstructure noise has been investigated quite heavily, but in the context of univariate realized volatility measurement, see Aït-Sahalia *et al.* (2005), Bandi and Russell (2005, 2006), Zhang *et al.* (2005), and Hansen and Lunde (2006), among others. In concurrent and independent work Bandi *et al.* (2008) derive the optimal sampling frequency to compute realized covariances that minimize the MSE. Applying this expression to our data we find an average optimal sampling frequency of 1.4 minutes, quite different from the 65-minute frequency that we find optimal with our approach. There are several differences between the approach of Bandi *et al.* (2008) and ours that might explain the large differences in results, and that make a direct comparison difficult. The most important difference is the fact that the optimal sampling frequencies optimize rather different criteria. Our optimal sampling frequencies aim to optimize economic criteria, including minimizing the risk or maximizing the return of stock portfolios. By contrast, the optimal sampling frequencies of Bandi *et al.* (2008) are designed to minimize the (daily) MSE of the realized (co-)variance estimator. Bandi *et al.* (2008) also allow their optimal sampling frequency to vary across (co-)variances and over time. This is convenient to handle the situation where the (relative) importance of microstructure noise displays cross-sectional and temporal variation. Within our approach, we aim for a single optimal sampling frequency that is the same for all (co-)variances and constant over time. The latter, however, can be relaxed by shortening the sampling period that is used to determine the optimal sampling frequency although selecting the appropriate period will necessarily be a subjective choice. A final explanation for the differences in optimal sampling frequencies that we find compared to Bandi *et al.* (2008) is discussed in Sheppard (2006) who shows that realized covariances can be substantially downward biased if the order of observation of prices is only weakly related to the order of price generation. Sheppard (2006) shows that due to this 'scrambling' effect the optimal sampling frequencies for realized covariances will be lower than the popular five-minute frequency.

Bandi *et al.* (2008) derive their optimal sampling frequency under the assumption that observed returns are equal to the true returns plus noise, with the noise being independent of the true return process. The same assumption is also used by Zhang *et al.* (2005) and Bandi and Russell (2005) to analyze the realized variance. This assumption may be considered somewhat restrictive in the context of measuring realized covariances using calendar time sampling, given that different stocks trade at different times. In that case, the noise term may contain previous true returns thereby violating the assumption that the noise is independent of the return process, see Lo and MacKinlay (1990) for extensive discussion. It would be interesting to examine whether an MSE-optimal sampling frequency can still be derived in a more general case where the noise is allowed to be correlated with the latent price process.[2]

The remainder of this chapter is organized as follows. Section 4.2 describes the data and the construction of the realized covariances. The mean-variance methodology is presented in Section 4.3. Results are discussed in Section 4.4. Section 4.5 concludes.

4.2 Data

The data set was obtained from Price-Data.com[3] and consists of open, high, low, and close transaction prices at the one-minute sampling frequency for the June 2004 S&P 100 index constituents, covering the period from April 16, 1997 until June 18, 2004 (1804 trading days). We disregard stocks for which the price series start at a later date, leaving 78 stocks for the analysis. The appendix provides a list of ticker symbols and company names. The data also comprise all (tick-by-tick) transaction prices of the S&P 500 index futures from April 16, 1997, through May 27, 2004. We follow the conventional practice of using the futures contract with the largest trading volume. This typically is the contract nearest to maturity, until a week before maturity when the next nearest contract takes over. Since the stock files miss April 9, 2003, and the futures files miss March 30, 2003 and May 3, 2004, this leaves 1788 common trading days from April 16, 1997, through May 27, 2004.

For each day t, we divide the trading session on the NYSE, which runs from 9:30 EST until 16:00 EST (390 minutes), into I intervals of equal length $h \equiv 1/I$, normalizing the daily interval to unity for ease of notation. For example, $I = 78$ for the five-minute

[2]We should note that the two time-scales estimator of Zhang *et al.* (2005) is derived under the same model assumptions (see equation (4) in Zhang *et al.* (2005). By using it in the same format for covariances we acknowledge that for example the weights for the covariance matrices at two frequencies may be suboptimal and that the estimator may be biased. Of course in our empirical work we can see whether despite these reservations the idea itself is useful for predicting covariances.

[3]`http://www.price-data.com/`

sampling frequency. Let p_{t-1+ih} denote the $(N \times 1)$ vector of log close transaction prices and let $r_{t-1+ih,h} \equiv p_{t-1+ih} - p_{t-1+(i-1)h}$ denote the $(N \times 1)$ vector of returns for the i^{th} intraday period on day t, for $i = 2, \ldots, I$, where $N = 78$ is the number of stocks. The return for the first intraday period, $r_{t-1+h,h}$, is defined as the difference between the log close and open transaction prices during that interval. The realized covariance matrix $V_{t,h}$ is defined as

$$V_{t,h} = r_{t,c-o} r'_{t,c-o} + \sum_{i=1}^{I} r_{t-1+ih,h} r'_{t-1+ih,h} \tag{4.1}$$

where $r_{t,c-o}$ is the $(N \times 1)$ vector of close-to-open (overnight) returns from day $t-1$ (close) to day t (open).[4] Martens (2002) documents that the overnight volatility represents an important fraction of total daily volatility, hence incorporating the cross-product of overnight returns as in (4.1) is important for accurately measuring (co-)variances, see also Fleming *et al.* (2003) and Hansen and Lunde (2005) for discussion. For the daily frequency the realized (co-)variance matrix V_t is defined as the outer product of the daily (close-to-close) returns, denoted by r_t, that is $V_t = r_t r'_t$.

Table 4.1, Panel A, illustrates some characteristic features of the daily realized variances and covariances by showing the mean (across stocks and across trading days) and variance for sampling frequencies of $390h = 1$, 2, 3, 5, 10, 15, 30, 65 and 130 minutes, such that in all cases the corresponding I intra-day intervals completely cover the 390-minute trading day. Several familiar patterns arise. First, the average realized variance increases with the sampling frequency (except for frequencies below 30 minutes). Bid-ask bounce induces negative autocorrelations in returns when prices are sampled more frequently leading to an upward bias in the realized variance. For example, the average variance using daily returns is 7.386 (corresponding to an annualized standard deviation of about 43%), whereas it is 9.494 for one-minute returns. Second, the average realized covariance decreases monotonically with the sampling frequency, where this downward bias can be attributed to non-synchronous trading, i.e. not every stock trades in each (intraday) interval or exactly at the end of each interval. The average covariance using one-minute returns is 0.826, whereas for daily data it is almost double at 1.568. Third, the variance of the realized (co-)variances becomes smaller for higher frequencies, simply because more data points are used. Hence in general for realized (co-)variances the bias increases and the variance decreases for higher sampling

[4]For obvious reasons the overnight return from 10 to 17 September, 2001 (the first trading day after 9/11) has been dropped.

Table 4.1: Mean and variance of the realized (co-)variance

Frequency	Realized Variance		Realized Covariance	
	Mean	Variance	Mean	Variance
Daily	7.386	1763	1.568	93.58
Panel A: Standard				
130 minutes	7.369	689.7	1.394	31.94
65 minutes	7.324	624.3	1.357	22.46
30 minutes	7.311	563.5	1.316	16.48
15 minutes	7.463	545.8	1.307	14.46
10 minutes	7.614	547.2	1.305	13.25
5 minutes	7.912	531.7	1.239	11.49
3 minutes	8.193	527.6	1.136	10.34
2 minutes	8.525	537.1	1.025	9.60
1 minute	9.494	597.0	0.826	8.73
Panel B: Two time-scales				
130 minutes	7.836	1090.6	1.462	37.05
65 minutes	7.295	760.9	1.418	22.94
30 minutes	7.150	606.4	1.414	16.82
15 minutes	7.233	570.1	1.420	14.97
10 minutes	7.315	562.8	1.407	13.90
5 minutes	7.446	554.9	1.361	12.24
3 minutes	7.500	534.5	1.297	11.32
2 minutes	7.524	521.4	1.231	10.80
Panel C: 1 lead and 1 lag				
130 minutes	7.422	758.6	1.418	36.73
65 minutes	7.368	667.0	1.389	26.27
30 minutes	7.329	595.6	1.351	18.97
15 minutes	7.342	552.7	1.332	15.51
10 minutes	7.420	545.6	1.334	14.23
5 minutes	7.611	532.3	1.300	12.67
3 minutes	7.812	538.3	1.257	11.56
2 minutes	8.020	536.3	1.193	10.70
1 minute	8.525	533.4	1.025	9.47

Notes: The table shows mean and variance of the realized (co-)variances at various sampling frequencies for 78 constituents of the S&P100 index from April 16, 1997, through May 27, 2004 (1788 trading days). For the realized variance the mean reflects the average taken over all 78 stocks and over all 1788 trading days. The variance is the average taken over the 78 sample variances of the realized variances. For the realized covariance the mean reflects the average taken over all 3003 pairs of stocks and over all 1788 trading days. The variance is the average taken over the 3003 sample variances of the realized covariances. In Panel A the "standard" realized covariance matrix $V_{t-1,h}$ given in (4.1) is used. Panel B is based on the two time-scales estimator $V_{t-1,h}^{TTS}$ given in (4.2), while Panel C shows results for the lead-lag corrected estimator $V_{t-1,h}^{LL}$ given in (4.3), with Bartlett-kernel weights $d_l = 1 - l/(q+1)$ and $q = 1$.

frequencies.[5]

One way to reduce the variance of realized covariances, given a particular sampling frequency, is to employ subsampling as first suggested in Zhang *et al.* (2005) in this context. In particular, the grid of x-minute intervals can be laid over the trading day in x different ways. For example, for the three-minute frequency rather than starting with the interval 9:30-9:33 one could also start with 9:31-9:34 or 9:32-9:35. In this way three 'subsamples' are created and each of these can be used to compute the realized covariance matrix. The final realized covariance matrix is then taken to be the average across subsamples. A practical problem with this procedure is how to treat the loose ends at the start and the end of the trading session. Here the start of the day is added to the overnight return, while the end of the day is omitted. The covariances measured during the trading session are proportionally inflated for the missing part of the trading session. Unreported summary statistics for the realized (co-)variances that are obtained with this procedure show that, in general, the effects of subsampling are ambiguous. There is a minor reduction in the variance of the realized covariances for the two- to 30-minute frequencies, but an increase in the variance of the realized variances, which becomes quite substantial for the lower sampling frequencies.

Zhang *et al.* (2005) suggest a bias-correction procedure for the subsampling estimator as described above using the realized covariance matrix obtained with the highest available sampling frequency. The essential argument is that subsampling still renders a biased estimate of the true integrated volatility with the bias being a function of the (co-)variance of the noise in the return processes. In fact, the realized (co-)variance estimator using the highest possible frequency consistently estimates this noise (co-)variance and can therefore be used to reduce and potentially even eliminate the bias of the subsampling estimator. Based on this idea, the two time-scales estimator $V_{t,h}^{TTS}$ is obtained as

$$V_{t,h}^{TTS} = \frac{I_{Max}}{I_{Max} - I} \left(V_{t,h}^{SubS} - \frac{I}{I_{Max}} V_{t,h}^{Max} \right), \tag{4.2}$$

where $V_{t,h}^{SubS}$ is the subsampling estimator using I returns over intervals of $390h$ minutes as described above, and $V_{t,h}^{Max}$ is the realized covariance matrix based on the highest possible sampling frequency with I_{Max} intraday return observations. In our case this is the one-minute frequency such that $I_{Max} = 390$.

[5] An exception is the realized variance at the one- and two-minute frequencies, where also the variance increases due to the increased importance of bid-ask bounce.

Summary statistics for the realized (co-)variances that are obtained with the two time-scales estimator are presented in Panel B of Table 4.1. The bias-correction procedure appears to work quite well, especially for the higher sampling frequencies, in the sense that the mean realized (co)variances get much closer to the mean values obtained with daily return observations. Note that this comes at the cost of increased variance of the realized (co)variances though, except for the realized variances at the one- and two-minute frequencies.

Finally, we examine whether the downward bias in the realized covariances can be reduced by adding lead and lagged covariances to the contemporaneous cross-product of returns in the spirit of Scholes and Williams (1977) and Cohen *et al.* (1983). Similarly, this might reduce the upward bias in the realized variance due to the negative autocorrelations in high-frequency returns, see Hansen and Lunde (2005, 2006). In particular, let $\Gamma_{t,h,l}$ denote the l-th order cross-covariance matrix of intraday h-period returns, that is

$$\Gamma_{t,h,l} = \sum_{i=1}^{I-l} r_{t-1+ih,h} r'_{t-1+(i-l)h,h}.$$

The realized covariance matrix with lead and lags is then obtained as

$$V_{t,h}^{LL} = V_{t,h} + \sum_{l=1}^{q} d_l \left(\Gamma_{t,h,l} + \Gamma'_{t,h,l}\right), \tag{4.3}$$

where $V_{t,h}$ is given by (4.1) and the weights d_l for the leads and lags are taken to be $d_l = 1 - l/(q+1)$. The use of these Bartlett-kernel weights guarantees that the realized covariance matrix $V_{t,h}^{LL}$ is positive definite, see Newey and West (1987) and Andrews (1991) for discussion of alternative weighting schemes that also achieve this objective. Precisely for this reason we do not consider the equal-weighting scheme ($d_l = 1$ for all $l = 1, \ldots, q$), as commonly used for estimating market betas of illiquid stocks and suggested by Zhou (1996) and Hansen *et al.* (2005b) in the context of realized variance and covariances, respectively. Barndorff-Nielsen *et al.* (2004) demonstrate that for realized variances the Bartlett-kernel estimator (4.3) and the subsampling estimator of Zhang *et al.* (2005) are almost identical.

Panel C of Table 4.1 present characteristics of $V_{t,h}^{LL}$ with $q = 1$.[6] As expected, the bias in both realized variances and realized covariances is reduced for all frequencies, although to a lesser extent compared to the two time-scales estimator. For example, the average

[6]We experimented with alternative values for q, which led to qualitatively similar findings. Detailed results are available upon request. The issue of determining the optimal value of q is beyond the scope of this chapter and is left for future research.

realized variance based on one-minute returns is reduced to 8.525, still considerably higher than the average daily squared return of 7.386. Similarly, the average realized covariance at the one-minute frequency is increased to 1.025, which comes closer to the average cross-product of daily returns (1.568) than the standard case. Note, however, that again the reduction in bias generally comes at the cost of increased variance. An exception is the one-minute frequency where not only the average realized variance is reduced and closer to the average daily squared return, but at the same time the variance is reduced from 597 to 533.

4.3 Methodology

4.3.1 Volatility-timing strategies

The benefits of high-frequency intraday data and the optimal way to employ these will be gauged by their economic value in the context of portfolio construction. In particular, we consider volatility timing strategies within the framework of conditional mean-variance analysis. We construct the minimum variance portfolio as well as the portfolio that minimizes variance given a set target return, which is denoted μ_P, allowing for daily rebalancing. To be precise, we solve the following two optimization problems for each day t:

$$\begin{aligned} &\min_{w_t} w_t'\Sigma_t w_t \\ &\text{s.t. } w_t'\iota = 1 \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} &\min_{w_t} w_t'\Sigma_t w_t \\ &\text{s.t. } w_t'\mu_t = \mu_P \quad \text{and} \quad w_t'\iota = 1 \end{aligned} \tag{4.5}$$

where w_t is the $(N \times 1)$ vector of portfolio weights, and ι denotes an $(N \times 1)$ vector of ones. In addition, μ_t is the $(N \times 1)$ vector with conditional expected returns for the individual stocks, that is $\mu_t \equiv \mathsf{E}[r_t|\mathcal{I}_{t-1}]$, where $\mathcal{I}_{t-1}$ denotes the information set available at the end of day $t-1$. Similarly, Σ_t is the $(N \times N)$ conditional covariance matrix, that is $\Sigma_t \equiv \mathsf{E}[(r_t - \mu_t)(r_t - \mu_t)'|\mathcal{I}_{t-1}]$. In order to concentrate on the use of high-frequency data for estimating and forecasting (co-)variances, we assume that μ_t is constant and, moreover, set it equal to the average returns in the complete out-of-sample period. Hence, we consider pure volatility-timing strategies, in the sense that the portfolio weights are determined exclusively by forecasts of the daily conditional covariance matrix Σ_t[7]. We return to these in Section 4.3.4 below[8].

The solution to the problem in (4.4), the weights for the fully invested minimum variance portfolio, is given by

$$w_{t,\mathrm{MVP}} = \frac{\Sigma_t^{-1}\iota}{\iota'\Sigma_t^{-1}\iota}. \tag{4.6}$$

For the solution of the problem in (4.5) first weights for the maximum Sharpe ratio portfolio are computed as

$$w_{t,\mathrm{MSR}} = \frac{\Sigma_t^{-1}\mu_t}{\iota'\Sigma_t^{-1}\mu_t} \tag{4.7}$$

and the weights for the target return portfolio are then provided by

$$w_{t,P} = \frac{\mu_{t,\mathrm{MSR}} - \mu_P}{\mu_{t,\mathrm{MSR}} - \mu_{t,\mathrm{MVP}}} w_{t,\mathrm{MVP}} + \frac{\mu_P - \mu_{t,\mathrm{MVP}}}{\mu_{t,\mathrm{MSR}} - \mu_{t,\mathrm{MVP}}} w_{t,\mathrm{MSR}} \tag{4.8}$$

where $\mu_{t,\mathrm{MVP}} = w'_{t,\mathrm{MVP}}\mu_t$ and $\mu_{t,\mathrm{MSR}} = w'_{t,\mathrm{MSR}}\mu_t$ are the expected returns on the minimum variance portfolio and the maximum Sharpe ratio portfolio, respectively.

In addition the above analysis is repeated using the conditional mean and covariance matrix for stock returns in excess of the S&P 500 futures returns. The solution to the problem in (4.4) then determines the minimum tracking error portfolio, i.e. the portfolio of the 78 S&P 100 stocks that tracks the S&P 500 index most closely. Similarly the solution to the problem in equation (4.5) then minimizes the tracking error given a certain target level of active return (i.e. portfolio return in excess of the S&P 500 return). The use of minimum tracking error portfolios is motivated by the analysis in Chan *et al.* (1999) who demonstrate that based on minimum variance portfolios it is difficult to distinguish between different covariance matrix estimates in the presence of a dominant (market) factor. Eliminating the dominant factor, in this case by switching to tracking error portfolios, largely solves this problem.

[7] In practice, an investor faces uncertainty about the vector of expected returns when constructing optimal portfolios. Fleming *et al.* (2001, 2003) examine the risk involved when estimating expected returns. See also Colacito and Engle (2006) for results on the ranking of different covariance matrix estimates, conditional on any vector of expected returns.

[8] As explained below, we require part of the sample period to initialize the conditional covariance matrix estimates, which in our case equals 122 trading days. This implies that the effective sample period available for portfolio construction and evaluation runs from October 8, 1997 until May 27, 2004 (1666 trading days).

4.3.2 The economic value of volatility timing

The performance of the portfolios in the different volatility timing strategies is evaluated using the ex-post daily stock returns r_t. For the minimum variance portfolio we consider the standard deviation, and for the target return portfolios we monitor the mean return, standard deviation, and Sharpe ratio, all based on ex-post returns. Similarly, for the minimum tracking error portfolio we consider the tracking error, and for the target active return portfolios we monitor the mean excess return, tracking error, and information ratio (excess return divided by tracking error), based on the ex-post daily excess returns.

Following Fleming *et al.* (2001, 2003), for the target return portfolios we assess the economic value of the different covariance matrix estimators in volatility timing strategies by determining the maximum performance fee a risk-averse investor would be willing to pay to switch from using one covariance matrix estimator to another. In particular, we assume the investor has a quadratic utility function given by

$$U(r_{t,P}) = W_0\left(1 + r_{t,P} - \frac{\gamma}{2(1+\gamma)}(1 + r_{t,P})^2\right), \tag{4.9}$$

where $r_{t,P} = w'_{t,P} r_t$ is the ex-post portfolio return, γ is the investor's relative risk aversion and W_0 is initial wealth. In order to compare two volatility timing strategies based on different covariance matrix estimators with portfolio returns denoted as r_{t,P_1} and r_{t,P_2}, we determine the maximum amount the investor is willing to pay to switch from the first strategy to the second. That is, we determine the value of Δ_γ such that

$$\sum_{t=1}^{T} U(r_{t,P_1}) = \sum_{t=1}^{T} U(r_{t,P_2} - \Delta_\gamma). \tag{4.10}$$

We interpret Δ_γ as a performance fee and report estimates in terms of basis points on an annual basis for $\gamma = 1$ and 10.

4.3.3 Transaction costs and rebalancing frequency

With daily rebalancing, the turnover of the volatility timing strategies is considerable, as shown in detail below. Hence, transaction costs play a non-trivial role and should be considered in evaluating the (relative) performance of different strategies. We handle this issue as follows. After rebalancing on day $t-1$, the i-th stock has been given a weight $w_{i,t-1}$ in the portfolio, $i = 1, \ldots, N$. The return on the i-th stock on day t is denoted as $r_{i,t}$ such that the portfolio return is $r_{t,P} = \sum_{i=1}^{N} w_{i,t-1} r_{i,t}$. At the moment just before rebalancing, denoted as t^-, the actual weight of the i-th stock in the portfolio therefore has changed to $w_{i,t^-} = w_{i,t-1}\frac{1+r_{i,t}}{1+r_{t,P}}$. The new weight $w_{i,t}$ for stock i follows from solving the investor's optimization problem, using time t information. The change in weight, or the required rebalancing, at time t is thus equal to $w_{i,t} - w_{i,t^-}$. We assume that transaction costs amount to a fixed percentage c on each traded dollar for any stock. Setting the initial wealth W_0 equal to 1 for simplicity, total transaction costs at time t are equal to

$$c_t = c\sum_{i=1}^{N} |w_{i,t} - w_{i,t^-}|,$$

such that the net portfolio return is given by $r_{t,P} - c_t$.

We report results for transaction cost levels between 2% and 20%, expressed in annualized percentage points. Note that this would be the reduction in the annualized portfolio return if the entire portfolio would have to be traded every day during a whole year, that is $\sum_{i=1}^{N} |w_{i,t} - w_{i,t^-}| = 1$ for all days t in a given year.[9]

A closely related issue is that of the portfolio rebalancing frequency. If daily turnover is substantial, transaction costs may eat away a considerable part of the portfolio performance and it may be better to rebalance the portfolio less frequently. In fact, given a certain level of transaction costs c one may attempt to determine the optimal rebalancing frequency, where a trade-off has to be made between updating the portfolio weights using the most recent covariance matrix information and incurring higher transaction costs. We consider this challenging problem to be beyond the scope of this chapter, though. We do provide some insight into the effect of the rebalancing frequency, by considering the portfolio performance if the holding period is set equal to a week or a month (or five and 21 trading days, respectively), as follows. We construct a new portfolio every day, but this is held on to for the next five (21) days. Hence, at any point in time the strategies effectively hold five (21) minimum variance portfolios, for example, each formed one day apart. To handle the problems concerned with overlapping returns, we calculate the overall return on day t as the average of all the portfolios that are held at that time.[10]

[9] Fleming *et al.* (2003) use a similar approach when assessing the effect of transaction costs.

[10] Note that our approach here differs from Fleming *et al.* (2003). Our method of holding multiple portfolios simultaneously is commonly applied in the literature on stock selection, see Jegadeesh and Titman (1993) and Rouwenvorst (1998), among many others.

4.3.4 Conditional covariance matrix estimators

Implementation of the portfolio construction methods discussed above requires estimates or forecasts of the conditional covariance matrix Σ_t. We closely follow Fleming *et al.* (2001, 2003) by using rolling volatility estimators for Σ_t, building on the work by Foster and Nelson (1996) and Andreou and Ghysels (2002b). The general rolling conditional covariance matrix estimator based on daily data is of the form

$$\widehat{\Sigma}_t = \sum_{k=1}^{\infty} \Omega_{t-k} \odot r_{t-k} r'_{t-k} \tag{4.11}$$

where Ω_{t-k} is a symmetric $(N \times N)$ matrix of weights, and $\odot$ denotes element-by-element multiplication. The weighting scheme is taken to be $\Omega_{t-k} = \alpha \exp(-\alpha k)\iota\iota'$, such that (4.11) can be rewritten as

$$\widehat{\Sigma}_t = \exp(-\alpha)\widehat{\Sigma}_{t-1} + \alpha \exp(-\alpha) r_{t-1} r'_{t-1}. \tag{4.12}$$

This choice is consistent with Foster and Nelson (1996) in that exponentially weighted estimators generally produce the smallest asymptotic MSE. In addition, using a single parameter (α) to control the rate at which the weights decay with lag length guarantees that $\widehat{\Sigma}_t$ is positive definite. One way of interpreting this weighting scheme is as a restricted multivariate GARCH model.[11] The optimal in-sample decay rate can therefore be estimated using (quasi) maximum likelihood for the model

$$r_t = \widehat{\Sigma}_t^{1/2} z_t \tag{4.13}$$

where $z_t \sim NID(0, I)$ and $\widehat{\Sigma}_t$ is given by (4.12). We estimate α using observations for the sample period October 8, 1997 until May 27, 2004 (1666 trading days). The reason for not using the sample from the first available day, April 16, 1997, onwards is that the covariance matrix estimate $\widehat{\Sigma}_t$ needs to be initialized. We use the first 122 observations as 'burn-in' period.

Given that the portfolios that subsequently are constructed using the weights $w_{\mathrm{MVP},t}$ from (4.6) and $w_{\mathrm{P},t}$ from (4.8) are evaluated over the same period that is used for estimating α, this raises the issue of data snooping. However, as noted by Fleming *et al.* (2001), the

[11] Fleming *et al.* (2003) show that actually using the (unrestricted) multivariate GARCH model leads to a better fit of the data as expected, but the covariance matrix forecasts result in worse portfolios than those obtained from the rolling covariance estimator. They cite the smoothness of the rolling estimator as the main reason for this.

statistical loss function used to estimate the decay parameter is rather different from the methods used to evaluate the performance of the various portfolios. Hence, look-ahead bias probably is not too big a problem. We return to this issue in Section 4.4.4.

Andersen *et al.* (2003a) and Barndorff-Nielsen and Shephard (2004a) show that intraday returns can be used to construct (co-)variance estimates that are more efficient than those based on daily returns. Sticking to the concept of rolling estimators and facilitating a direct comparison between daily and intraday data, it is most natural to replace the daily update $r_{t-1}r'_{t-1}$ in (4.12) by the realized covariance matrix $V_{t-1,h}$, that is, the conditional covariance matrix is estimated using high-frequency data as

$$\widehat{\Sigma}_{t,h} = \exp(-\alpha_h)\widehat{\Sigma}_{t-1,h} + \alpha_h \exp(-\alpha_h)V_{t-1,h} \tag{4.14}$$

where α_h can again be estimated by means of maximum likelihood for the model (4.13), but now using $\widehat{\Sigma}_{t,h}$ instead of $\widehat{\Sigma}_t$. In addition to the realized covariance matrix $V_{t-1,h}$ obtained from the 'basic' form given in (4.1), we implement (4.14) using the two time-scales estimator $V^{TTS}_{t-1,h}$ given in (4.2) and the lead-lag corrected estimator $V^{LL}_{t-1,h}$ given in (4.3). As mentioned before, we examine different sampling frequencies to construct the realized covariance matrix $V_{t-1,h}$, dividing the 390-minute NYSE trading session in (nonoverlapping) intervals of 1, 2, 3, 5, 10, 15, 30, 65 or 130 minutes.

We close this section by noting that the conditional covariance matrix estimate $\widehat{\Sigma}_{t,h}$ obtained from (4.14) may suffer from the biases in the realized covariance matrix $V_{t-1,h}$ due to market microstructure effects. For that reason, Fleming *et al.* (2003) propose a bias-correction method based on scaling the elements of $\widehat{\Sigma}_{t,h}$ with factors determined from the contemporaneous estimates of the daily-returns-based rolling estimator $\widehat{\Sigma}_t$ obtained from (4.12), see also Hansen and Lunde (2005, 2006). Although we considered this approach in our analysis, we were unable to obtain satisfactory results. The key problem with this bias adjustment procedure is that it adjusts each individual element in the covariance matrix separately, with a possibly different correction factor. Hence, whereas the unadjusted covariance matrix $\widehat{\Sigma}_{t,h}$ obtained from (4.14) is guaranteed to be positive definite, this does not hold for the bias-adjusted matrix. In the empirical application considered in Fleming *et al.* (2003) concerning three highly-liquid future contracts, this issue turns out not to be relevant, but for our application to the 78 S&P 100 stocks we ran into this problem quite frequently due to the large number of stocks. We tried to address this in several different ways but to no avail.

4.4 Results

4.4.1 Optimal decay rates

Table 4.2 shows the optimal decay rates α and α_h that maximize the likelihood of the model in equations (4.13) with (4.12) for daily returns and with (4.14) for intraday returns at the different sampling frequencies considered. Starting with total returns (as opposed

Table 4.2: Optimal decay parameters

Frequency	Standard		Two time-scales		1 lead, 1 lag	
	α	LogL	α	LogL	α	LogL
Panel A: Total Returns						
Daily	0.0070	$-300,492$	0.0070	$-300,492$	0.0070	$-300,492$
130 minutes	0.0119	$-276,376$	0.0129	$-275,476$	0.0111	$-276,939$
65 minutes	0.0149	$-274,580$	0.0156	$-273,578$	0.0137	$-274,782$
30 minutes	0.0204	$-273,747$	0.0200	$-272,644$	0.0179	$-273,519$
15 minutes	0.0273	$-273,802$	0.0256	$-272,707$	0.0231	$-273,186$
10 minutes	0.0329	$-274,121$	0.0293	$-273,071$	0.0273	$-273,261$
5 minutes	0.0481	$-275,004$	0.0356	$-274,024$	0.0375	$-273,774$
3 minutes	0.0678	$-275,975$	0.0386	$-274,939$	0.0493	$-274,407$
2 minutes	0.1025	$-276,985$	0.0385	$-275,729$	0.0643	$-275,164$
1 minute	0.2106	$-278,971$			0.1255	$-276,723$
Panel B: Excess Returns						
Daily	0.0070	$-298,480$	0.0070	$-298,480$	0.0070	$-298,480$
130 minutes	0.0119	$-274,455$	0.0130	$-273,449$	0.0112	$-275,033$
65 minutes	0.0151	$-272,712$	0.0158	$-271,507$	0.0138	$-272,907$
30 minutes	0.0208	$-271,926$	0.0204	$-270,572$	0.0183	$-271,669$
15 minutes	0.0282	$-272,100$	0.0263	$-270,702$	0.0238	$-271,381$
10 minutes	0.0342	$-272,516$	0.0304	$-271,133$	0.0282	$-271,522$
5 minutes	0.0514	$-273,562$	0.0373	$-272,238$	0.0393	$-272,157$
3 minutes	0.0757	$-274,663$	0.0406	$-273,293$	0.0529	$-272,925$
2 minutes	0.1178	$-275,675$	0.0402	$-274,192$	0.0713	$-273,751$
1 minute	0.2468	$-277,537$			0.1440	$-275,393$

Notes: The table shows the decay rates (α) that maximize the likelihood of the model in (4.13) and (4.12) for daily data and (4.13) and (4.14) for intraday data. In Panel A the model is estimated for total returns, whereas in Panel B the model is estimated for excess returns (stock returns minus S&P500 returns). The second and third column show the optimal decay rates and accompanying log-likelihood values when the covariance updates are based on the standard realized (co-)variances, the fourth and fifth column when the updates are based on the two time-scales estimator, and the final two columns when 1 lead and 1 lag of the (co-)variances are added to the contemporaneous (realized) covariances.

to returns in excess of the S&P 500 return) and the standard case (no two time-scales, no lead-lag correction), the optimal decay parameter increases monotonically from 0.0070 for daily data to 0.2106 for the one-minute frequency. This pattern implies that the update $V_{t-1,h}$ in (4.14) is given more weight when it is measured, presumably more accurately, at higher sampling frequencies. Fleming *et al.* (2003) report decay parameters of 0.031 and 0.064 for daily returns and five-minute returns, respectively, for the three liquid futures contracts they consider. The lower decay parameters at these frequencies obtained here for the 78 S&P 100 stocks are likely to be caused by having relatively more noise in the intra-day returns data and a well-known phenomenon in multivariate GARCH models (for daily returns) that the larger the number of assets, the lower the decay parameter, see Engle and Sheppard (2001) and Hafner and Franses (2003) for discussion.

The two time-scales estimator and the lead-lag correction reduce the bias but increase the variance of the realized covariance matrix for a particular sampling frequency. It appears that for both methods the latter is more important here, given that the decay parameters are lower for the corrected covariance matrices compared to the standard case. Note that the log-likelihood is improved, however, except when using the lead-lag correction for the lowest sampling frequencies. The decay parameters in Panel B, considering excess returns, are in general slightly higher in all instances, but otherwise the findings correspond to those for the total returns.

4.4.2 Portfolio performance

Table 4.3 shows the performance of the overall minimum variance portfolio, with weights defined in (4.6), and the minimum variance portfolio given an annualized target return of 10%,[12] with weights given by (4.8). For the overall minimum variance portfolio the optimal sampling frequency turns out to be 65 minutes in the standard case. The annualized standard deviation of 12.16% compares favorably to the 14.00% for daily data. For the popular five-minute frequency the standard deviation is 12.68%, clearly above the minimum. Also for the target return portfolios the 65-minute frequency is optimal, resulting in a Sharpe ratio of 0.786 compared to 0.596 for daily returns and 0.626 for five-minute returns. In terms of the performance fees (Δ_γ), an investor with low relative risk aversion ($\gamma = 1$) would be willing to pay 155 basis points per year to switch from the covariance matrix

[12]We examined the sensitivity of our results to the target return level by varying μ_P between 2% and 18%. These alternative target return levels led to qualitatively similar conclusions as those reported below. Detailed results are therefore not shown here, but are available on request.

Table 4.3: Out-of-sample performance - total returns

	Target return portfolio						Min. variance portfolio	
Frequency	μ_P	σ_P	SR	Δ_1	Δ_{10}	TO	σ_{MVP}	TO
Daily	8.84	14.83	0.596			16.8	14.00	16.4
Panel A: Standard								
130 minutes	10.24	13.13	0.780	163.3	376.9	13.5	12.46	12.9
65 minutes	10.12	12.87	0.786	154.5	399.1	13.2	12.16	12.5
30 minutes	8.16	12.99	0.628	−42.8	187.4	13.4	12.17	12.7
15 minutes	8.82	13.12	0.673	22.0	237.3	13.8	12.21	13.0
10 minutes	8.29	13.30	0.623	−33.9	160.1	14.2	12.38	13.4
5 minutes	8.56	13.69	0.626	−11.5	135.4	15.6	12.68	14.7
3 minutes	8.48	13.92	0.610	−22.7	95.6	17.4	12.84	16.5
2 minutes	7.67	14.18	0.541	−107.6	−22.3	21.1	13.06	20.0
1 minute	7.89	14.44	0.546	−89.6	−38.3	28.6	13.33	26.9
Panel B: Two time-scales								
130 minutes	9.27	13.04	0.711	67.8	292.6	13.2	12.26	12.7
65 minutes	8.82	12.88	0.685	25.0	268.3	12.4	12.10	11.9
30 minutes	8.16	12.88	0.633	−41.4	201.8	12.1	12.13	11.6
15 minutes	8.42	13.04	0.646	−17.3	208.0	12.4	12.24	11.8
10 minutes	8.78	13.17	0.667	17.3	226.2	12.6	12.33	12.0
5 minutes	9.18	13.49	0.680	52.1	222.5	12.9	12.57	12.2
3 minutes	9.14	13.73	0.665	44.9	186.2	12.8	12.73	12.0
2 minutes	9.21	13.91	0.662	49.8	168.3	12.4	12.88	11.6
1 minute								
Panel C: 1 lead and 1 lag								
130 minutes	10.23	13.11	0.780	162.8	378.8	13.6	12.44	13.0
65 minutes	10.25	12.87	0.797	168.0	412.9	13.3	12.16	12.6
30 minutes	9.07	12.81	0.708	50.8	302.7	13.2	12.01	12.6
15 minutes	8.41	12.89	0.653	−16.3	226.3	13.4	12.05	12.7
10 minutes	7.81	13.01	0.600	−78.2	149.6	13.7	12.15	12.9
5 minutes	8.54	13.28	0.643	−8.9	187.0	14.4	12.35	13.6
3 minutes	8.50	13.55	0.627	−15.9	147.6	15.6	12.57	14.8
2 minutes	8.05	13.80	0.583	−64.6	68.6	17.1	12.75	16.2
1 minute	8.06	14.10	0.571	−68.2	27.0	23.2	12.99	22.0

Notes: The table shows the out-of-sample performance of the overall minimum variance portfolio, with weights given in (4.6), and the minimum variance portfolio given a target level of return of 10%, with weights given in (4.8), constructed using rolling covariance matrix forecasts based on various sampling frequencies and based on different ways of measuring the realized covariance matrix (standard, two time-scales, and 1 lead and 1 lag). For the target return portfolios, we report the mean return (μ_P) and standard deviation (σ_P) in annualized percentage points, the Sharpe ratio (SR), the annualized basis points fee (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the daily returns covariance matrix estimate to the intraday returns of the optimal portfolios, and average daily turnover (TO) in percentage points. For the minimum variance portfolios, we report the standard deviation (σ_{MVP}) in annualized percentage points average daily turnover (TO) in percentage points.

estimate based on daily returns to the realized covariance matrix obtained with 65-minute returns. An investor with high relative risk aversion ($\gamma = 10$) would be willing to pay as much as 399 basis points.

The results in Panel B of Table 4.3, using the two time-scales estimator, show only a marginal improvement for the overall minimum variance portfolio with a standard deviation of 12.10% compared to 12.16% before, both at the 65-minute sampling frequency. The same conclusion holds for all other frequencies except 15 minutes. For the target return portfolios, however, the results are ambiguous, in the sense that for sampling frequencies of 10 minutes and higher the two time-scales estimator leads to a higher Sharpe ratio but for lower sampling frequencies portfolio performance worsens. At the optimal frequency of 130 minutes the Sharpe ratio is lower at 0.711, compared to 0.786 for the 65-minute frequency in the standard case. The performance fees Δ_γ show the same pattern.

The lead-lag correction in (4.3) leads to a higher optimal sampling frequency of 30 minutes for the minimum variance portfolio. The 12.01% annualized standard deviation is slightly better than the 12.16% and 12.10% at the optimal 65-minute frequency in the standard and two time-scales cases, respectively. In fact, the lead-lag bias-correction leads to a reduction in volatility of the minimum variance portfolio at all frequencies, such that the 10-minute sampling frequency now leads to approximately the same level of volatility as the optimal 65-minute frequency in the standard case. Hence, using the lead-lag correction allows for a substantially higher sampling frequency before the increased noise level due to the use of leads and lags offsets this advantage. For the target return portfolios, the optimal sampling frequency remains at 65 minutes as in the standard case, although the corresponding Sharpe ratio is somewhat higher (0.797 compared to 0.786). The same applies to the performance fees Δ_γ, which increase to 168 and 412 basis points per year for low and high relative risk aversion, respectively (compared to 155 and 399).

The performance of the minimum tracking error portfolios is shown in Table 4.4. Using the standard realized covariance matrix, the tracking error is minimized at 4.43% using the 30-minute frequency compared to 4.75% for daily data and 4.92% at the popular five-minute frequency. The two time-scales estimator provides a further improvement with the minimum tracking error equal to 4.18% at the 65-minute frequency. Finally, using one lead and lag results in a higher optimal sampling frequency of 15 minutes as for the minimum variance portfolio, with a marginally lower tracking error at 4.35%. Hence here we do observe that bias-correction further improves the performance.

Table 4.4: Out-of-sample performance - excess returns

	Target active return portfolio						Min. TE portfolio	
Frequency	μ_P	TE_P	IR	Δ_1	Δ_{10}	TO	TE_{MTE}	TO
Daily	0.52	4.77	0.110			5.2	4.75	5.2
Panel A: Standard								
130 minutes	0.45	4.55	0.099	−6.4	3.0	6.2	4.53	6.2
65 minutes	0.28	4.51	0.063	−22.7	−11.9	6.7	4.48	6.7
30 minutes	0.09	4.48	0.019	−42.4	−30.1	7.6	4.43	7.5
15 minutes	0.57	4.52	0.127	6.2	16.9	8.8	4.46	8.8
10 minutes	0.75	4.67	0.161	23.2	27.8	9.8	4.59	9.8
5 minutes	0.89	4.99	0.179	35.7	26.3	12.6	4.92	12.6
3 minutes	1.85	5.23	0.353	129.9	109.3	16.6	5.18	16.6
2 minutes	2.36	5.42	0.436	180.5	150.8	22.4	5.40	22.5
1 minute	1.39	5.79	0.241	81.7	33.3	36.2	5.78	36.3
Panel B: Two time-scales								
130 minutes	0.54	4.26	0.126	3.7	24.4	4.4	4.24	4.4
65 minutes	0.44	4.21	0.105	−5.7	17.1	4.4	4.18	4.3
30 minutes	0.49	4.22	0.116	−0.9	21.5	4.7	4.19	4.6
15 minutes	0.20	4.32	0.046	−30.4	−11.8	5.5	4.28	5.4
10 minutes	0.56	4.43	0.126	5.0	19.1	6.2	4.38	6.1
5 minutes	1.13	4.78	0.236	60.5	60.1	7.9	4.72	7.9
3 minutes	1.81	5.16	0.351	126.9	109.6	9.5	5.11	9.5
2 minutes	2.28	5.51	0.414	172.1	138.0	11.1	5.45	11.1
1 minute								
Panel C: 1 lead and 1 lag								
130 minutes	0.18	4.58	0.038	−34.0	−25.9	6.2	4.56	6.1
65 minutes	0.32	4.50	0.071	−19.0	−7.8	6.3	4.48	6.3
30 minutes	0.41	4.44	0.092	−9.9	3.9	6.8	4.40	6.8
15 minutes	0.62	4.40	0.141	11.4	26.8	7.5	4.35	7.5
10 minutes	0.63	4.48	0.140	11.7	24.1	8.1	4.42	8.0
5 minutes	0.72	4.58	0.156	20.1	28.1	9.4	4.51	9.4
3 minutes	1.57	4.77	0.329	104.7	104.7	11.4	4.71	11.3
2 minutes	1.95	4.97	0.394	142.2	133.7	13.7	4.91	13.7
1 minute	2.16	5.33	0.406	161.3	135.9	22.2	5.31	22.3

Notes: The table shows the out-of-sample performance of the overall minimum tracking error portfolio, with weights given in (4.6), and the minimum tracking error portfolio given a target level of return of 1%, with weights given in (4.8), constructed using rolling covariance matrix forecasts based on various sampling frequencies and based on different ways of measuring the realized covariance matrix (standard, two time-scales, and 1 lead and 1 lag). For the target active return portfolios, we report the mean active return (μ_P) and tracking error (TE_P) in annualized percentage points, the information ratio (IR), the annualized basis points fee (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the daily returns covariance matrix estimate to the intraday returns of the optimal portfolios, and average daily turnover (TO) in percentage points. For the minimum tracking error portfolios, we report the tracking error (TE_{MTE}) in annualized percentage points average daily turnover (TO) in percentage points.

Table 4.4 also demonstrates that for the active portfolio manager with an annualized target excess return of five percent the optimal sampling frequency is much higher than for total returns. The ex-post information ratio (excess return divided by tracking error) is optimal for the two-minute frequency in the standard case at 0.436 compared to an information ratio of 0.110 at the daily frequency. Risk-averse investors would be willing to pay between 151 and 181 basis points per year to make use of the two-minute frequency realized covariance matrix. The optimal frequency using one lead and one lag is even the one-minute frequency, but it results in a slightly lower information ratio of 0.406 and slightly lower performance fees of 136 and 161 basis points. The two time-scales estimator also results in an optimal frequency of two minutes but with an information ratio of 0.414 and performance fees of 138 and 172 basis points, below the optimum in the standard case. Comparing the information ratios at other frequencies with the corresponding results in the standard case, again we find ambiguous results. The information ratio declines for sampling frequencies of 15 minutes and higher, but it increases for lower sampling frequencies, while the same is observed for the performance fees Δ_γ. Note that this pattern is the complete opposite of that found for the target return portfolios in Panel B of Table 4.3.

In sum, the general conclusion from Tables 4.3 and 4.4 when computing the minimum variance portfolio or minimum tracking error portfolio is that the two time-scales estimator and the lead-lag bias correction marginally improve the out-of-sample performance. We emphasize, however, that selecting the appropriate sampling frequency appears to be much more important than choosing between different bias- and variance-reduction techniques for the realized covariance matrices. For example, the reduction in volatility of the minimum variance portfolio when going from the popular five-minute frequency to the optimal 65-minute frequency in the standard case (from 12.68% to 12.16%) is more than three times as large as the additional reduction achieved by applying the lead-lag bias correction at the 30-minute frequency (which further reduces volatility to 12.01%).

In general we would like to express a warning note on the target return results in Tables 4.3 and 4.4. The actual return pattern at the various frequencies is anything but smooth and hence subject to a certain degree of ‘luck’. Obviously these results depend both on the quality of the expected (excess) returns and the covariance matrix forecasts, making a direct comparison of the quality of the covariance forecasts more difficult than is the case for the minimum variance and minimum tracking error portfolios.

4.4.3 Transaction costs and rebalancing frequency

Table 4.3 shows that daily turnover in the volatility timing strategies is considerable, ranging between 12 and 17 percent for most sampling frequencies. In case the 'standard' realized covariance matrix or the lead-lag correction is used at the one-minute sampling frequency, turnover increases even to around 25%. For the tracking error portfolios in Table 4.4, turnover is below 10% for sampling frequencies below five minutes, but rapidly increases when returns are sampled more frequently. Given that the optimal sampling frequency for the target excess return portfolios was found to be two minutes, transaction costs may be substantial and should be taken into account when assessing the portfolio performance.

Panel A of Table 4.5 shows the performance of the target return portfolios with daily rebalancing for transaction cost levels (c in the first column) between 2% and 20%, in annualized percentage points as explained before. Results are shown for portfolios based on covariance matrix estimates obtained with daily returns and with intraday returns at the optimal frequency, which turns out to be 65 minutes irrespective of the transaction costs level. No bias-corrections are applied to the realized covariance matrix in this case. As expected, transaction costs reduce the portfolio return while the portfolio variance is largely unaffected, leading to a monotonic decline of the Sharpe ratio as the level of transaction costs increases. Note that the reduction in returns and Sharpe ratio is larger for the portfolios based on covariance matrix estimates obtained with daily returns. Therefore the difference in Sharpe ratios with the portfolios based on high-frequency intraday returns actually becomes larger, such that the performance fees increase to 245 and 487 basis points for $\gamma = 1$ and 10 in case transaction costs amount to 20%. This is not surprising of course, given that daily turnover for these portfolios equals 16.8 and 13.2%, respectively (see Table 4.3).

Panel A of Table 4.6 reveals that transaction costs have more dramatic effects for the target excess return portfolios. Daily turnover for the portfolio based on covariance matrix estimates obtained with the optimal two-minute returns is more than four times as high as for the portfolio based on daily returns, at 22.4% compared to 5.2%. The reduction in the mean active return and the information ratio therefore is much more pronounced for the intraday returns based strategy, such that the performance fee actually becomes negative for transaction costs in excess of 10%. Hence, if transaction cost levels are considerable, it does not pay off to use high-frequency intraday returns to estimate the covariance matrix.

Table 4.5: Transaction costs and rebalancing frequency - total returns

	Daily returns			Intraday returns					
c	μ_P	σ_P	SR	μ_P	σ_P	SR	h	Δ_1	Δ_{10}
Panel A: Daily rebalancing									
0	8.84	14.83	0.596	10.12	12.87	0.786	65	154.5	399.1
2	8.49	14.83	0.572	9.84	12.87	0.765	65	162.7	407.0
4	8.12	14.83	0.548	9.56	12.87	0.743	65	171.0	415.1
6	7.75	14.82	0.523	9.27	12.87	0.720	65	179.6	423.5
8	7.37	14.82	0.497	8.98	12.87	0.698	65	188.3	431.9
10	6.98	14.82	0.471	8.68	12.87	0.674	65	197.2	440.6
12	6.58	14.82	0.444	8.37	12.87	0.651	65	206.3	449.5
14	6.18	14.82	0.417	8.06	12.87	0.626	65	215.6	458.6
16	5.76	14.82	0.389	7.74	12.87	0.601	65	225.1	467.9
18	5.33	14.82	0.360	7.41	12.87	0.576	65	234.9	477.4
20	4.90	14.81	0.330	7.08	12.87	0.550	65	244.9	487.2
Panel B: Weekly rebalancing									
0	9.62	14.88	0.646	10.36	13.17	0.787	130	98.4	314.8
2	9.43	14.88	0.634	10.22	13.17	0.776	130	103.1	319.4
4	9.25	14.88	0.621	10.08	13.17	0.766	130	107.9	324.2
6	9.05	14.88	0.608	9.94	13.17	0.755	130	112.8	329.0
8	8.86	14.88	0.595	9.80	13.17	0.744	130	117.8	333.9
10	8.66	14.88	0.582	9.65	13.17	0.732	130	122.9	339.0
12	8.45	14.88	0.568	9.49	13.17	0.721	130	128.1	344.1
14	8.24	14.88	0.554	9.34	13.17	0.709	130	133.5	349.4
16	8.03	14.88	0.540	9.18	13.17	0.697	130	138.9	354.8
18	7.81	14.88	0.525	9.01	13.17	0.684	130	144.5	360.3
20	7.58	14.88	0.510	8.85	13.17	0.672	130	150.3	366.0
Panel C: Monthly rebalancing									
0	10.53	14.95	0.704	10.36	13.31	0.779	130	7.0	217.1
2	10.43	14.95	0.698	10.29	13.31	0.774	130	9.3	219.4
4	10.34	14.95	0.691	10.22	13.31	0.768	130	11.6	221.7
6	10.24	14.95	0.685	10.15	13.31	0.763	130	14.0	224.1
8	10.14	14.95	0.678	10.07	13.31	0.757	130	16.4	226.5
10	10.04	14.95	0.671	9.99	13.31	0.751	130	18.9	229.0
12	9.93	14.95	0.664	9.91	13.31	0.745	130	21.5	231.5
14	9.83	14.95	0.657	9.83	13.30	0.739	130	24.1	234.1
16	9.72	14.95	0.650	9.75	13.30	0.733	130	26.7	236.7
18	9.60	14.95	0.642	9.67	13.30	0.727	130	29.5	239.4
20	9.49	14.95	0.635	9.58	13.30	0.720	130	32.3	242.2

Notes: The table shows the out-of-sample performance of the minimum variance portfolio given a target level of return of 10%, with weights given in (4.8), constructed using rolling covariance matrix forecasts based on daily returns and on intraday returns at the sampling frequency that maximized the information ratio, based on the 'standard' way of measuring the realized covariance matrix. We report the mean return (μ_P) and standard deviation (σ_P) in annualized percentage points, the Sharpe ratio (SR), and the annualized basis points fee (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the daily returns covariance matrix estimate to the intraday returns of the optimal portfolios. The column headed c indicates the level of transaction costs, expressed in annualized percentage points, which correspond with the reduction in the annualized portfolio return if the entire portfolio would have to be traded every day during the whole year. The column headed h indicates the optimal sampling frequency, expressed as the length of the corresponding return interval in minutes.

Table 4.6: Transaction costs and rebalancing frequency - excess returns

	Daily returns			Intraday returns					
c	μ_P	TE_P	IR	μ_P	TE_P	IR	h	Δ_1	Δ_{10}
Panel A: Daily rebalancing									
0	0.52	4.77	0.110	2.36	5.42	0.436	2	180.5	150.8
2	0.41	4.77	0.085	1.91	5.42	0.352	2	146.6	116.7
4	0.29	4.77	0.061	1.44	5.42	0.266	2	111.9	81.9
6	0.17	4.77	0.035	0.97	5.42	0.178	2	76.5	46.4
8	0.05	4.77	0.010	0.48	5.43	0.089	2	40.3	10.1
10	−0.08	4.76	−0.017	0.08	5.24	0.015	3	13.5	−7.7
12	−0.21	4.76	−0.044	−0.30	5.24	−0.057	3	−11.3	−32.7
14	−0.34	4.76	−0.072	−0.68	5.24	−0.131	3	−36.7	−58.2
16	−0.48	4.76	−0.100	−1.08	5.24	−0.206	3	−62.7	−84.3
18	−0.61	4.76	−0.129	−1.48	5.24	−0.283	3	−89.3	−111.1
20	−0.76	4.76	−0.159	−1.90	5.24	−0.362	3	−116.6	−138.5
Panel B: Weekly rebalancing									
0	0.23	4.81	0.048	2.36	5.29	0.447	2	211.0	189.3
2	0.17	4.81	0.035	2.16	5.29	0.408	2	196.6	174.9
4	0.11	4.81	0.023	1.96	5.29	0.369	2	182.0	160.2
6	0.05	4.81	0.010	1.74	5.29	0.330	2	167.0	145.2
8	−0.01	4.81	−0.002	1.53	5.29	0.289	2	151.7	129.9
10	−0.08	4.81	−0.016	1.30	5.18	0.251	3	135.8	119.5
12	−0.14	4.81	−0.029	1.13	5.18	0.217	3	124.7	108.3
14	−0.21	4.81	−0.043	0.95	5.18	0.183	3	113.3	96.9
16	−0.27	4.81	−0.057	0.76	5.18	0.147	3	101.7	85.3
18	−0.34	4.81	−0.071	0.57	5.18	0.111	3	89.8	73.3
20	−0.42	4.81	−0.086	0.38	5.18	0.073	3	77.5	61.1
Panel C: Monthly rebalancing									
0	0.22	4.88	0.044	3.45	5.28	0.654	1	321.5	303.0
2	0.19	4.88	0.038	3.36	5.28	0.636	1	315.2	296.7
4	0.16	4.88	0.032	3.27	5.28	0.619	1	308.8	290.3
6	0.13	4.88	0.026	3.17	5.28	0.601	1	302.3	283.8
8	0.10	4.87	0.020	3.07	5.28	0.582	1	295.6	277.1
10	0.07	4.87	0.013	2.97	5.28	0.563	1	288.8	270.3
12	0.03	4.87	0.007	2.87	5.28	0.544	1	281.8	263.3
14	0.00	4.87	0.000	2.77	5.28	0.524	1	274.7	256.2
16	−0.03	4.87	−0.007	2.66	5.28	0.504	1	267.4	248.9
18	−0.07	4.87	−0.014	2.55	5.28	0.483	1	259.9	241.4
20	−0.10	4.87	−0.021	2.44	5.28	0.462	1	252.3	233.7

Notes: The table shows the out-of-sample performance of the minimum tracking error portfolio given a target level of return of 1%, with weights given in (4.8), constructed using rolling covariance matrix forecasts based on daily returns and on intraday returns at the sampling frequency that maximized the information ratio, based on the 'standard' way of measuring the realized covariance matrix. We report the mean active return (μ_P) and tracking error (TE_P) in annualized percentage points, the information ratio (IR), the annualized basis points fee (Δ_γ) an investor with quadratic utility and constant relative risk aversion of γ would pay to switch from the daily returns covariance matrix estimate to the intraday returns of the optimal portfolios. The column headed c indicates the level of transaction costs, expressed in annualized percentage points, which correspond with the reduction in the annualized portfolio return if the entire portfolio would have to be traded every day during the whole year. The column headed h indicates the optimal sampling frequency, expressed as the length of the corresponding return interval in minutes.

Also note that the optimal sampling frequency becomes lower at three minutes for high levels of transaction costs.

Transaction costs may be reduced by rebalancing the portfolio less frequently. The effects on portfolio performance are shown in Panels B and C of Table 4.5 for the target return portfolios. First note that, as expected, the volatility of the portfolio increases when the portfolio holding period increases, but only slightly. Somewhat surprisingly, the portfolio return increases considerably and comes much closer to the target return of 10% when the rebalancing frequency decreases. This corresponds with the findings of Fleming *et al.* (2003). Turning to the effects of transaction costs, we find that the reduction in returns and Sharpe ratio is indeed much less pronounced when rebalancing the portfolio weekly or monthly rather than daily. Again, turnover is higher for the portfolios based on covariance matrix estimates obtained with daily returns, such that the maximum fee investors are willing to pay to switch to covariance matrix estimates obtained with intraday returns increases with the level of transaction costs. Also note that the magnitude of the performance fee Δ_γ declines when the rebalancing frequency becomes lower. This is due to the fact that the improvement in performance when going from daily to weekly or monthly rebalancing is relatively larger for the portfolio based on daily returns.

In order to assess the economic value of rebalancing less frequently more directly, we compute the performance fee Δ_γ that an investor is willing to pay to switch from daily rebalancing to weekly (or monthly) rebalancing for a given level of transaction costs. For the daily and weekly rebalanced portfolios based on intraday returns at the optimal sampling frequencies and annualized transaction costs equal to 10%, we find that Δ_γ is equal to 94 and 47 basis points for $\gamma = 1$ and 10, respectively. These performance fees even increase to 128 and 64 points when comparing the daily and monthly rebalancing frequencies.

Finally, the benefits of rebalancing less frequently become very clear from Panels B and C of Table 4.6 for the target excess return portfolio. Although the active return is still reduced due to transaction costs in case of weekly or monthly rebalancing, it remains higher for the portfolio based on intraday returns than for the daily returns portfolio even in case of transaction costs up to 20%. Given that the levels of ex-post tracking error do not differ very much, the IR remains higher as well, and investors are willing to pay considerable fees to make use of the high-frequency returns portfolio.

Again we compute the performance fee Δ_γ using portfolios with daily and weekly (or monthly) holding periods for a given level of transaction costs to evaluate the economic gains from rebalancing less frequently directly. With annualized transaction costs equal to 10%, we find that an investor is willing to pay 200 basis points to switch from daily to weekly rebalancing for both low and high relative risk aversion. Comparing the daily and monthly rebalancing frequencies, Δ_γ is equal to 280 and 276 basis points for $\gamma = 1$ and 10, respectively. It would be interesting to include the rebalancing frequency in the portfolio optimization problem. Obviously this is difficult to achieve and beyond the scope of this chapter.

4.4.4 Genuine out-of-sample forecasting

Fleming *et al.* (2001, 2003) suggest that determining the decay parameters α and α_h in (4.12) and (4.14), respectively, using maximum likelihood on the full sample does not lead to serious data snooping problems because the final evaluation criterion (maximizing return or minimizing risk) differs from the likelihood objective function. To test the validity of this argument, and to test a true out-of-sample strategy, we proceed as follows. First we find the decay parameters that maximize the performance of the various portfolios over the first 250 days following the initial burn-in period, i.e. the values of α and α_h that minimize the (relative) variance or maximizes the Sharpe (or information) ratio. These decay parameters are then used to estimate the conditional covariance matrices $\widehat{\Sigma}_t$ and $\widehat{\Sigma}_{t,h}$ for the first day following the in-sample period, for which optimal portfolio weights are then constructed using (4.6) and (4.8). This procedure is repeated using an expanding in-sample estimation window where each time a new observation is added. This not only implies that the decay parameter varies over time, but also that the portfolio performance thus obtained is truly out-of-sample. Since we lose an additional 250 days at the start of the sample, for comparison we re-estimated the decay parameter using maximum likelihood for the shorter sample of 1416 trading days and constructed the corresponding portfolio weights and performance.

The results are presented in Table 4.7. For both the minimum variance and minimum tracking error portfolios the results are re-assuring, in the sense that the optimal sampling frequency is still 65 and 30 minutes, respectively. Also the performance itself is similar to that of the standard case. By contrast, for the target return portfolios the results do change considerably. In the total return case the optimal sampling frequency is now 10 minutes instead of 65, and the Sharpe ratio has deteriorated from 0.640 to 0.554. In the excess return case the optimal sampling frequency is now one minute instead of two, but with a better information ratio at 0.457 versus 0.373. Perhaps most revealing, the optimal decay parameters are much lower when determined using in-sample portfolio performance

Table 4.7: Out-of-sample α's

	Target return portfolio				Minimum risk portfolio			
	α_h		SR/IR		α_h		σ_P/TE_P	
Frequency	Mean	St.Dev	Perf.	LogL	Mean	St.Dev	Perf.	LogL
Panel A: Total Returns								
Daily	0.001	0.003	0.507	0.482	0.004	0.002	13.770	13.669
130 minutes	0.069	0.037	0.534	0.626	0.014	0.002	12.534	12.229
65 minutes	0.087	0.107	0.262	0.640	0.018	0.003	11.937	11.945
30 minutes	0.035	0.020	0.436	0.519	0.024	0.003	11.986	11.972
15 minutes	0.097	0.130	0.261	0.603	0.034	0.005	12.061	12.034
10 minutes	0.352	0.034	0.554	0.561	0.042	0.006	12.220	12.207
5 minutes	0.312	0.086	0.434	0.597	0.047	0.005	12.471	12.466
3 minutes	0.223	0.189	0.471	0.578	0.080	0.030	12.592	12.556
2 minutes	0.215	0.194	0.452	0.526	0.107	0.076	12.797	12.755
1 minute	0.361	0.116	0.478	0.516	0.124	0.076	13.002	12.971
Panel B: Excess returns								
Daily	0.041	0.004	0.357	−0.031	0.006	0.001	4.858	4.871
130 minutes	0.004	0.006	0.167	0.122	0.008	0.003	4.593	4.581
65 minutes	0.002	0.002	0.298	0.161	0.008	0.004	4.595	4.554
30 minutes	0.002	0.009	0.181	0.096	0.011	0.005	4.512	4.477
15 minutes	0.001	0.001	0.376	0.198	0.012	0.006	4.516	4.483
10 minutes	0.004	0.012	0.203	0.312	0.012	0.006	4.630	4.629
5 minutes	0.001	0.002	0.395	0.234	0.010	0.005	4.814	4.949
3 minutes	0.009	0.010	0.239	0.348	0.010	0.005	4.970	5.220
2 minutes	0.024	0.041	0.269	0.373	0.011	0.005	5.095	5.514
1 minute	0.031	0.028	0.453	0.184	0.015	0.006	5.249	5.864

Notes: The table shows the out-of-sample performance of the overall minimum volatility (tracking error) portfolio, with weights given in (4.6), and the minimum variance portfolio given an annualized target level of (active) return of 10% (1%), with weights given in (4.8), constructed using rolling covariance matrix forecasts based on various sampling frequencies and based on the 'standard' realized covariance matrix. Panel A shows results for total returns and Panel B for excess returns (stock returns minus S&P 500 returns). The optimal decay parameters are determined by optimizing portfolio performance using an expanding window period (starting with 250 days). Columns 2 and 3, and 6 and 7, report the mean and standard deviation of the resulting estimates of α_h. Columns 4 and 8, headed 'Perf.', show the Sharpe ratio and volatility (panel A) or the information ratio and tracking error (panel (B) for the resulting portfolios. Columns 5 and 9, headed 'LogL', show the SR/IR and σ_P/TE_P for portfolios constructed with decay parameters for the conditional covariance matrix that are estimated by maximizing the log-likelihood over the complete out-of-sample period.

than when estimated with maximum likelihood (except for the target return portfolios, when performance is measured by the Sharpe ratio). This holds especially for the higher sampling frequencies. To verify that this is not an artefact of using different decay rates over time, we also did a datasnooping exercise with a constant decay parameter equal to the value that maximizes performance (rather than the log-likelihood) over the entire out-of-sample period. These results (not reported here) confirm that performance-based decay rates are much lower than the ones based on the log-likelihood. In addition, this enhances the performance at those frequencies. Hence the log-likelihood procedure tends to give too much weight to the updates. A logical explanation for this is that the noise pattern of the updates suits the log-likelihood when standardizing equally noise daily returns, but more smoothing is needed (lower decay parameters) for forecasting the covariance matrices.

4.5 Concluding remarks

Existing studies that use high-frequency intra-day data to measure and forecast the daily covariance matrix make ad-hoc choices with regard to the sampling frequency. The presence of bid-ask bounce and non-synchronous trading creates a trade-off between higher sampling frequencies leading to lower variances of the (co-)variance measures due to having more data, and lower sampling frequencies reducing the impact of these market microstructure effects. Popular ad-hoc choices to strike a balance between the resulting bias and variance of the realized covariance estimates are the five- and 30-minute sampling frequencies.

In this chapter we show that choosing the optimal sampling frequency is crucial for the out-of-sample performance of portfolios constructed using realized covariances. Even for the relatively liquid stocks that comprise the S&P 100 index the optimum is more likely to be in the neighbourhood of an hour rather than five or 30 minutes.

We also investigated the use of bias- and variance-reduction methods for computing the realized covariances. Both the two time-scales estimator and the lead-lag bias-correction procedure result in a marginal improvement over the standard realized covariance matrix estimator at the same frequency. Transaction costs were shown to affect portfolio performance considerably, and in particular they imply that rebalancing the portfolio less frequently may be beneficial.

Several interesting topics for further research come to mind. First, it would be interesting to explore other ways to correct for biases in realized covariances due to non-synchronous trading. Second, it may be worthwhile to allow the sampling frequency to vary over time, to take into account changes in trading intensity. The S&P stocks considered here, for example, were traded much more frequently at the end of the sample period than in the beginning, see also Bandi *et al.* (2008). Third, Andersen *et al.* (2003a) suggest that with more and more assets eventually a factor model will be needed, see Andersen *et al.* (2001a) and Hafner *et al.* (2006) for additional motivation and discussion and Bollerslev and Zhang (2003) for an application using the Fama-French three-factor model. Fourth, it would be interesting to examine the effects of restrictions on the portfolio weights, which we did not consider here. As shown by Jagannathan and Ma (2003), imposing short-selling constraints and a maximum weight constraint, for example, may enhance portfolio performance, even if the restrictions are wrong. Finally, in this chapter we considered the popular approach that makes use of artificially constructed equidistant prices in calendar time, in part because the empirical data set was constrained to the close of each minute rather than all transaction prices. It would be interesting to see empirical work on the scale of this chapter with many stocks that considers transaction time sampling rather than calendar time sampling, for example using the covariance estimators of Harris *et al.* (1995), De Jong and Nijman (1997) and Hayashi and Yoshida (2005). Martens (2004) provides an overview and comparison in a simulation setting, while Hansen *et al.* (2005b) discuss theoretical issues related to such estimators including bias correction procedures.

Appendix

4A S&P 100 constituents on June 18, 2004

The 100 constituents of the S&P 100 index on June 18, 2004. The 78 stocks marked with a * are included in the analysis. For these stocks there is a complete set of one-minute open-high-low-close prices from April 16, 1997, through May 27, 2004 (1788 trading days).

Table 4.8: S&P 100 constituents

Symbol	Issue name	Symbol	Issue name
AA*	ALCOA INC	IBM*	INTL BUS MACHINE
AEP*	AMER ELEC PWR	INTC*	INTEL CORP
AES*	THE AES CORP	IP	INTL PAPER CO
AIG*	AMER INTL GROUP	JNJ*	JOHNSON&JOHNSON
ALL*	ALLSTATE CP	JPM*	JP MORGAN CHASE
AMGN*	AMGEN	KO*	COCA COLA CO
AOL	AOL TIME WARNER	LEH*	LEHMAN BROS
ATI	ALLEGHENY TECH	LTD*	LIMITED BRANDS
AVP	AVON PRODS INC	LU*	LUCENT TECH
AXP*	AMER EXPRESS CO	MAY*	MAY DEPT STORES
BA*	BOEING CO	MCD*	MCDONALDS CORP
BAC*	BANK OF AMERICA	MDT*	MEDTRONIC INC
BAX*	BAXTER INTL INC	MEDI	MEDIMMUNE INC
BCC*	BOISE CASCADE	MER*	MERRILL LYNCH
BDK*	BLACK & DECKER	MMM*	3M COMPANY
BHI*	BAKER HUGHES INC	MO*	ALTRIA GROUP
BMY*	BRISTOL MYERS SQ	MRK*	MERCK & CO
BNI*	BURL NTHN SANTA	MSFT*	MICROSOFT CP
BUD*	ANHEUSER BUSCH	MWD	MORGAN STANLEY
C*	CITIGROUP	NSC*	NORFOLK SOUTHERN
CCU*	CLEAR CHANNEL	NSM*	NATL SEMICONDUCT
CI*	CIGNA CORP	NXTL*	NEXTEL COMMS
CL*	COLGATE PALMOLIV	ONE*	BANK ONE CORP
CPB*	CAMPBELL SOUP CO	ORCL*	ORACLE CORP
CSC	COMPUTER SCIENCE	PEP*	PEPSICO INC
CSCO*	CISCO SYSTEMS	PFE*	PFIZER INC
DAL*	DELTA AIR LINES	PG	PROCTER & GAMBLE
DD*	DU PONT CO	ROK*	ROCKWELL AUTOMAT
DIS*	WALT DISNEY CO	RSH	RADIOSHACK
DOW	DOW CHEMICAL CO	RTN	RAYTHEON CO
EK*	EASTMAN KODAK	S*	SEARS ROEBUCK
EMC*	EMC CORP	SBC*	SBC COMMS
EP	EL PASO CORP	SLB*	SCHLUMBERGER LTD
ETR*	ENTERGY CP	SLE*	SARA LEE CORP
EXC	EXELON CORP	SO*	SOUTHERN CO
F	FORD MOTOR CO	T*	AT&T CORP
FDX	FEDEX CORP	TOY*	TOYS R US CORP
G*	GILLETTE CO	TXN*	TEXAS INSTRUMENT
GD*	GENERAL DYNAMICS	TYC*	TYCO INTL
GE*	GENERAL ELEC CO	UIS*	UNISYS CORP
GM*	GENERAL MOTORS	USB	US BANCORP
GS	GOLDM SACHS GRP	UTX*	UNITED TECH CP
HAL*	HALLIBURTON CO	VIAb	VIACOM CL B
HCA	HCA INC	VZ	VERIZON COMMS
HD*	HOME DEPOT INC	WFC*	WELLS FARGO & CO
HET*	HARRAHS ENTER	WMB*	WILLIAMS COS INC
HIG*	HARTFORD FINL	WMT*	WAL-MART STORES
HNZ*	H J HEINZ CO	WY	WEYERHAEUSER CO
HON*	HONEYWELL INTL	XOM	EXXON MOBIL
HPQ*	HEWLETT-PACKARD	XRX*	XEROX CORP

Part B

Modeling and Forecasting Interest Rates

Chapter 5

Examining the Nelson-Siegel Class of Term Structure Models

In-sample fit versus out-of-sample forecasting performance

5.1 Introduction

Accurate estimates of the current term structure of interest are of crucial importance in many areas of finance. Equally important is the ability to forecast the future term structure. It is not surprising therefore that substantial research effort has been devoted to the questions of how to optimally estimate, model and forecast the term structure of interest rates. One class of models that has the potential of providing satisfactory answers to these questions is that of the Nelson-Siegel models.

Nelson and Siegel (1987) proposed to fit the term structure using a flexible, smooth parametric function. They demonstrated that their proposed model is capable of capturing many of the typically observed shapes that the yield curve assumes over time. Since then various extensions have been proposed that incorporate additional flexibility with a popular extension being the Svensson (1994) model. Despite the drawback that they lack theoretical underpinnings, the Bank of International Settlements (BIS, 2005) reports that currently nine out of thirteen central banks which report their curve estimation methods to the BIS use either the Nelson-Siegel or the Svensson model to construct zero-coupon yield curves. As the Nelson-Siegel model is also widely used among practitioners, this ranks it among the most popular term structure estimation methods.

Recently, Diebold and Li (2006) have shown that the three-factor Nelson-Siegel model can also be used to construct accurate term structure forecasts. By using a straightforward two-step estimation procedure they demonstrate that the model performs well, relative to competing models, especially for longer forecast horizons. Mönch (2006a) partially confirms these results and Fabozzi, Martellini, and Priaulet (2005) show that the Nelson-Siegel model produces forecasts that are not only statistically accurate but also economically meaningful as these can be used to generate substantial investment returns.

Due to these successes it is not surprising that the Nelson-Siegel model is increasingly being used in other applications as well. For example, Diebold, Rudebusch, and Aruoba (2006b) use the model to study the interactions between the macro economy and the yield curve (see also Diebold, Piazzesi, and Rudebusch, 2005) whereas Diebold, Ji, and Li (2006a) apply it to identify systematic risk sources and to construct a generalized duration measure.

Much of this research focuses, however, solely on the original three-factor Nelson-Siegel model. Extensions such as the Svensson model have not yet been investigated for their out-of-sample performance whereas extensions like those of Björk and Christensen (1999) have been left nearly unexamined altogether. This chapter tries to fill this gap. In particular, I examine several models within the Nelson-Siegel class for their in-sample fitting and out-of-sample forecasting performance. Diebold and Li (2006) show that the dynamic three-factor Nelson-Siegel model had the potential of performing well in both areas. This motivates a closer examination of the various extended Nelson-Siegel models. It is unclear, however, if, or to what extent, models that are capable of parsimoniously fitting the term structure in-sample should also necessarily render accurate out-of-sample forecasts. No-arbitrage models for example typically fit the term structure quite accurately but forecast poorly (Duffee, 2002). An important task is therefore to try and evaluate the trade-off between in-sample fit and out-of-sample forecasting performance. More flexible models will most likely improve the in-sample fit but the question thus is to what extent these can also produce better out-of-sample results. In order to address this question for the class of Nelson-Siegel models I use a sample of U.S. Treasury zero-coupon bond yields consisting of twenty years of monthly data. I determine which features of the extended models help to improve the term structure fit. To gauge the out-of-sample performance I construct yield forecasts for short and long-term horizons and compare these with forecasts from several competitor models. In addition to looking at different Nelson-Siegel specifications I also examine in detail the benefits of alternative model estimation techniques, discuss several potential estimation and identification issues and propose solutions on how to tackle these.

The results can be summarized as follows. First of all I show that the more flexible models fit the term structure more accurately than the three-factor Nelson and Siegel

(1987) model. This is not a surprising result in itself. What is interesting though is that a similar fit can be obtained as that of the popular Svensson (1994) model by extending the three-factor model with a second slope factor as in Björk and Christensen (1999). The advantage of the four-factor model is that it is easier to estimate than the Svensson model as it is less hampered by potential non-identification issues when estimating the factors.

In addition to an improved in-sample fit, I also demonstrate that the four-factor model produces accurate out-of-sample forecasts. In fact, the four-factor model outperforms the random walk benchmark and AR and VAR competitor models as well as all other Nelson-Siegel specifications, including the three-factor model. The best results are obtained by simultaneously taking into account cross-sectional and time-series information about yields when estimating the model and using an AR specification for the factor dynamics. The four-factor model forecasts increasingly well for all maturities when the forecast horizon lengthens. The outperformance relative to the random walk is substantial as it reduces the RMSPE by often as much as 10% or more. Subsample analysis shows that, unlike the performance of for example the three-factor model, the four-factor model is consistently producing highly accurate forecasts.

The remainder of this chapter is structured as follows. In Section 5.2 I give a short review of term structure estimation methods. Section 5.3 discusses the various Nelson-Siegel models in detail and Section 5.4 is devoted to the estimation of these models. Section 5.5 describes the data. The in-sample results are presented in Section 5.6 whereas Section 5.7 shows the out-of-sample forecast results. Section 5.8 concludes and offers some directions for further research.

5.2 Term structure estimation methods

The term structure of interest rates describes the relationship between interest rates and time to maturity. The standard way of measuring the term structure of interest rates is by means of the spot rate curve, or yield curve[1], on zero-coupon bonds. The reason behind this is that yields-to-maturity on coupon-bearing bonds suffer from the 'coupon-effect' (see Caks, 1977) which implies that two bonds which are identical in every respect except for bearing different coupon-rates can have a different yield-to-maturity. The problem with zero-coupon yields on the other hand, is that these can only be directly observed from

[1]The yield-to-maturity and the spot rate on a zero-coupon bond are the same. Because in this chapter I focus solely on zero-coupon bond interest rates I use both terms interchangeably.

Treasury Bills which have maturities of twelve months or less. Longer maturity zero-coupon yields need to be derived from coupon-bearing Treasury Notes and Bonds. In practice, we can therefore not observe the entire term structure of interest rates directly. We need to estimate it using approximation methods[2] Term structure estimation methods are designed for the purpose of approximating one of three equivalent representations of the term structure: the spot rate curve, discount curve and forward rate curve. Once we have a representation for one of these we can automatically derive the other representations. In the remainder of this section I briefly discuss the three curves and fix notation[3]. For convenience, I assume throughout that all rates are continuously compounded.

The forward rate curve characterizes forward rates as a function of maturity. A forward rate $f_t(\tau, \tau^*)$ is the interest rate of a forward contract on an investment which is initiated τ periods in the future and which matures τ^* periods beyond the start date of the contract. We obtain the *instantaneous* forward rate $f_t(\tau)$ by letting the maturity of such a forward contract go to zero:

$$\lim_{\tau^* \downarrow 0} f_t(\tau, \tau^*) = f_t(\tau) \tag{5.1}$$

The instantaneous-maturity forward rate curve represent forward rates on infinitesimal-maturity forward contracts which are initiated τ periods in the future for $\tau \in [0, \infty)$.

Given the forward curve, we can determine the spot rate (or yield) on a zero-coupon bond with τ periods to maturity, denoted by $y_t(\tau)$, by taking the equally weighted average over the forward rates:

$$y_t(\tau) = \frac{1}{\tau} \int_0^\tau f_t(m) \mathrm{d}m \tag{5.2}$$

The discount curve, $P_t(\tau)$, which denotes the present value of a zero-coupon bond that pays out a nominal amount of \$1 after τ periods, can in turn be obtained from the spot rate curve by

$$P_t(\tau) = \exp\left[-\tau y_t(\tau)\right] \tag{5.3}$$

[2]. In the U.S. zero-coupon rates are to a certain extent more directly available through the use of STRIPS. The Treasury STRIPS program, which started in 1985, allows an investor to split coupon-bearing Treasury Notes and Bonds into a basket of zero-coupon securities, see Sack (2000) for a discussion. Due to a limited investor demand for short- and medium-term zero-coupon securities there are some concerns, however, about the liquidity of shorter-term STRIPS and it is therefore still more common practice to estimate the yield curve using coupon-bearing bonds.

[3]For a more elaborate discussion see, e.g., Svensson (1994).

The final relationship we have links forward rates directly to the discount curve and is given by

$$f_t(\tau) = -\frac{1}{P_t(\tau)}\frac{\mathrm{d}P_t(\tau)}{\mathrm{d}\tau} = y_t(\tau) + \tau\frac{\mathrm{d}y_t(\tau)}{\mathrm{d}\tau} \tag{5.4}$$

We can move from one curve to the other by using the relationships specified in (5.2)-(5.4).

Various methods have been proposed to estimate the term structure from (quoted) bond prices. A popular approach is the bootstrapping procedure by Fama and Bliss (1987) which consists of sequentially extracting forward rates from bond prices with successively longer maturities. The Fama and Bliss (1987) approach exactly prices all bonds included in the procedure and assumes that the forward rate between observed maturities is constant. The dataset I analyze in this chapter consists of Fama-Bliss interest rates. Other term structure estimation methods use for example cubic splines (McCulloch, 1975), exponential splines (Vasicek and Fong, 1982), polynomials functions (Chambers *et al.*, 1984), parametric methods (Nelson-Siegel, see e.g. Bliss, 1997) or non-parametric methods (Linton *et al.*, 2001). Studies such as Bliss (1997), Ferguson and Raymar (1998) and Jeffrey *et al.* (2006) compare several different estimation methods and demonstrate the pros and cons of the various methods.

Once the decision has been made as to which method to use to construct an estimate of the term structure, the next step is to build a model to describe the evolution of the term structure over time. Popular models are no-arbitrage affine models, e.g. the one-factor models by Vasicek (1977) and Cox *et al.* (1985) or multi-factor models as specified and analyzed in Duffie and Kan (1996), Dai and Singleton (2000) and De Jong (2000). In this study I focus solely on the class of Nelson-Siegel models. Diebold and Li (2006) show that the Nelson-Siegel model not only provides a good in-sample fit of the term structure but also produces accurate out-of-sample interest rate forecasts for a 6 and 12-month forecast horizon. Diebold and Li (2006) only consider the original three-factor Nelson and Siegel (1987) model, however. The purpose of this chapter is to examine a broader class of Nelson-Siegel models. This includes for example the four-factor specifications proposed by Svensson (1994) and Björk and Christensen (1999).

5.3 Nelson-Siegel class of models

5.3.1 Three-factor base model

Nelson and Siegel (1987) suggest to fit the forward rate curve at a given date with a mathematical class of approximating functions. The functional form they advocate uses Laguerre functions which consist of the product between a polynomial and an exponential decay term. The resulting Nelson-Siegel approximating forward curve can be assumed to be the solution to a second order differential equation with equal roots for spot rates[4]

$$f_t(\tau) = \beta_{1,t} + \beta_{2,t} \exp\left(-\frac{\tau}{\lambda_t}\right) + \beta_{3,t} \left(\frac{\tau}{\lambda_t}\right) \exp\left(-\frac{\tau}{\lambda_t}\right) \tag{5.5}$$

The parameters $\beta_{t,1}$, $\beta_{t,2}$ and $\beta_{t,3}$ are determined by initial conditions and λ_t is a constant associated with the equation. By averaging over forward rates, as in (5.2), we obtain the spot rate curve

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left[\frac{1-\exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)}\right] + \beta_{3,t} \left[\frac{1-\exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)} - \exp\left(-\frac{\tau}{\lambda_t}\right)\right] \tag{5.6}$$

There are several reasons why the Nelson-Siegel model is such a popular term structure estimation method. First of all, it provides a parsimonious approximation of the yield curve using only a small number of parameters (contrary to for example spline methods). Together, the three components $[1,\ \frac{1-\exp(-\tau/\lambda_t)}{(\tau/\lambda_t)},\ \frac{1-\exp(-\tau/\lambda_t)}{(\tau/\lambda_t)} - \exp\left(-\tau/\lambda_t\right)]$, give the model enough flexibility to capture a range of monotonic, humped and S-type shapes typically observed in yield data. Second of all, the model produces forward and yield curves which have the desirable property of starting off from an easily computed instantaneous short rate value of $\beta_{1,t} + \beta_{2,t}$ and levelling off at a finite infinite-maturity value of $\beta_{1,t}$, that is constant[5]:

$$\lim_{\tau \downarrow 0} y_t(\tau) = \beta_{1,t} + \beta_{2,t}; \qquad \lim_{\tau \to \infty} y_t(\tau) = \beta_{1,t} \tag{5.7}$$

Finally, the three Nelson-Siegel components have a clear interpretation as short, medium and long-term components. These labels are the result of each element's contribution to the yield curve. Figure 5.1[a] depicts the value of each component as a function of maturity. The *long-term* component is the component on $\beta_{1,t}$ because it is constant at 1 and therefore

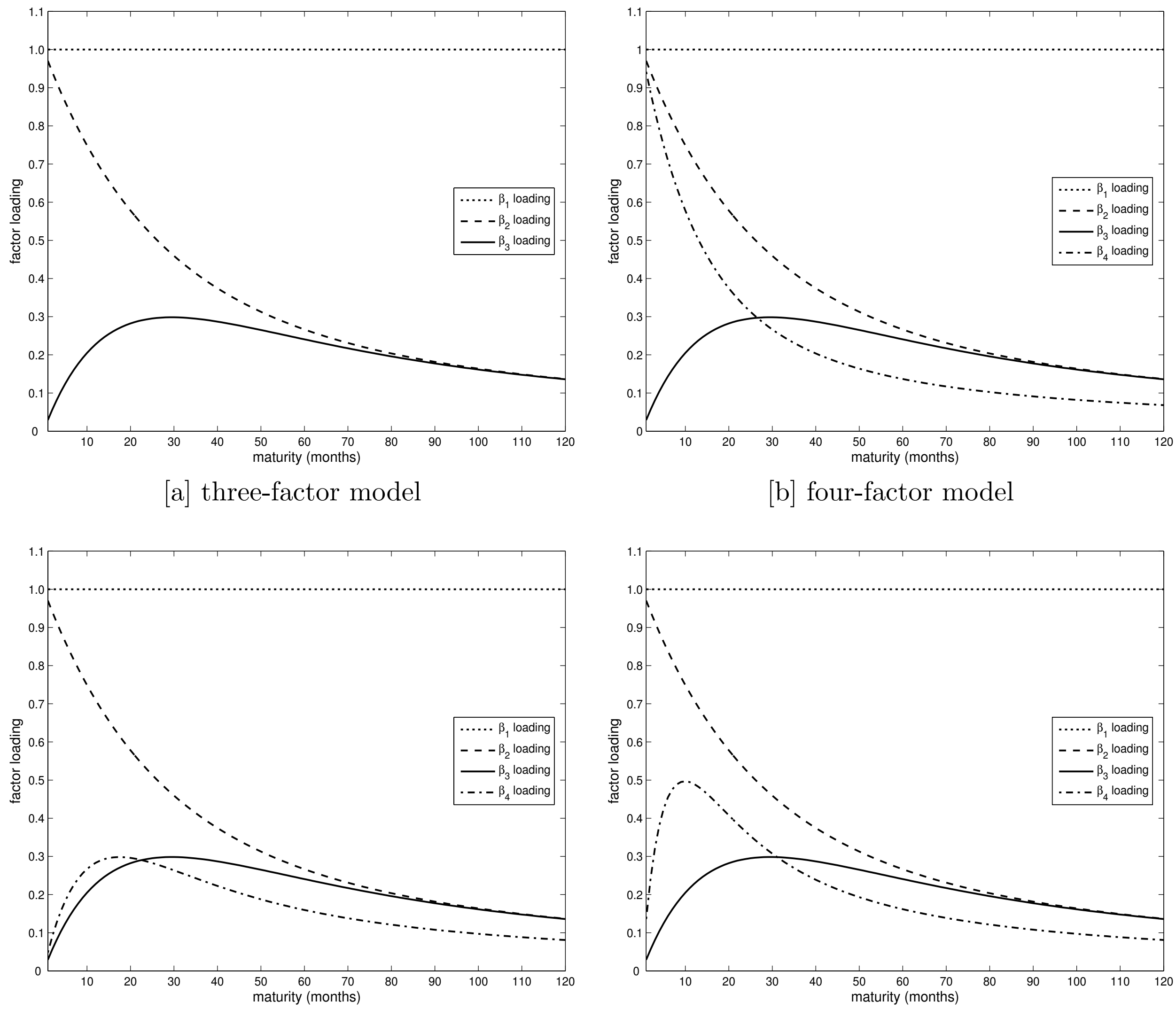
[4]For the specification of the Nelson-Siegel model I follow Fabozzi *et al.* (2005), Diebold and Li (2006) and Diebold, Rudebusch, and Aruoba (2006b) although I specify the decay parameter(s) the same way as in Nelson and Siegel (1987).

[5]The limiting behavior of the spot curve and the forward curve are the same. The Nelson-Siegel discount curve, which follows from combining (5.6) with (5.3) is given by

$$P_t(\tau) = \exp\left\{(-\beta_{1,t}\tau - \beta_{2,t}[1-\exp\left(-\lambda_t\tau\right)] - \beta_{3,t}[1-\exp\left(-\lambda_t\tau\right) - \tau\exp\left(-\lambda_t\tau\right)]\right\},$$

The discount curve starts at 1 and converges to zero for infinite maturities as required.

Figure 5.1: Nelson-Siegel factor loadings



[a] three-factor model

[b] four-factor model

[c] Svensson model

[d] Adjusted Svensson model

Notes: The graph depicts the factor loadings for β_1 (dotted line), β_2 (dashed line), β_3 (solid line) and β_4 (dash-dotted line) for the [a] three-factor, [b] four-factor, [c] Svensson and [d] Adjusted Svensson Nelson-Siegel model. The factor loadings are plotted using $\lambda = 16.42$ for the 3-factor and 4-factor models. For the (Adjusted) Svensson model it holds that $\lambda_1 = 16.42$ and $\lambda_2 = 9.73$ which ensures that the maturities at which the two curvature factors reach their maximum is at least twelve months apart.

the same for every maturity. The component on $\beta_{2,t}$ is $\left[\frac{1-\exp(-\tau/\lambda_t)}{(\tau/\lambda_t)}\right]$ and is designated as the *short-term* component. It starts at 1 but then decays to zero at an exponential rate. The rate of decay is determined by the parameter λ_t. Smaller values for λ_t induce a faster decay to zero. The *medium-term* component is $\left[\frac{1-\exp(-\tau/\lambda_t)}{(\tau/\lambda_t)} - \exp\left(-\tau/\lambda_t\right)\right]$ which starts at 0, increases for medium maturities and then decays to zero again thereby creating a hump-shape. The decay parameter λ_t determines at which maturity this component reaches its maximum.

Although the Nelson-Siegel model was in essence designed to be a static model which does not account for the intertemporal evolution of the term structure, Diebold and Li (2006) show that the coefficients $\beta_{1,t}, \beta_{2,t}$ and $\beta_{3,t}$ can be interpreted as three latent dynamic factors[6]. Moreover, the authors show that the labels level, slope and curvature are appropriate for these factors. The long-term factor $\beta_{1,t}$ governs the level of the yield curve whereas $\beta_{2,t}$ and $\beta_{3,t}$ govern its slope and curvature respectively.

By casting the Nelson-Siegel model into a dynamic framework, Diebold and Li (2006) further show that the model is capable of replicating the main empirical facts of the term structure of interest rates over time: the average curve is upward sloping and concave, yield dynamics are highly persistent with long maturity rates being more persistent than short-maturity rates, and interest rate volatility is decreasing for longer maturities. Due to its attractive properties and its widespread use by central banks and practitioners I regard the three-factor model in (5.6) as the Nelson-Siegel base model. Note that Diebold and Li (2006) as well as for example Dolan (1999), Fabozzi *et al.* (2005) and Mönch (2006a) first fix λ_t to a pre-specified value and then proceed with analyzing the three-factor model. Here, I estimate λ_t as well as fixing it.

Although the base model can already capture a wide range of shapes, it cannot handle all the shapes that the term structure assumes over time. As an attempt to remedy this problem, several more flexible Nelson-Siegel specifications have been proposed in the literature to better fit more complicated shapes, mainly shapes with multiple minima and/or maxima. These extended Nelson-Siegel models achieve the increase in flexibility by introducing either additional factors, further decay parameters, or by a combination of both. In the remainder of this section I discuss which of these specifications I will examine for their in-sample fit and out-of-sample predictive accuracy.

[6] The short, medium and long-term components can therefore also be interpreted as factor loadings.

5.3.2 Alternative Nelson-Siegel specifications

5.3.3 Two-factor model

The first model I consider is a restriction rather than an extension of the three-factor model. Litterman and Scheinkman (1991), among many other studies, show that the variation in interest rates can be explained by only a small number of underlying common factors. Typically, the first three principal component factors are already sufficient since these explain the bulk of interest rate variance but also because they have the intuitive interpretation as level, slope and curvature factors from the manner in which these factors affect the yield curve. The third factor has usually very little to add, however, (typically only a few percentage points) to the amount of interest rate variance that is already captured by the first two factors[7]. For this reason, authors such as Bomfim (2003a) and Rudebusch and Wu (2003) consider two-factor affine models to explain interest rate dynamics whereas Diebold, Piazzesi, and Rudebusch (2005) examine a two-factor Nelson-Siegel model. Compared to the three-factor Nelson-Siegel model, the two-factor model only contains the level and slope factor:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left[\frac{1 - \exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)} \right] \tag{5.8}$$

Diebold, Piazzesi, and Rudebusch (2005) argue that since the first two principal components explain nearly all variation in interest rates, a two-factor model may suffice to forecast the term structure. They also argue, however, that two factors will most likely not be enough to accurately fit the entire yield curve[8].

5.3.4 Björk and Christensen (1999) four-factor model

The three-factor Nelson-Siegel model can be extended in various ways to increase its flexibility. From an estimation point of view, the easiest approach is to introduce additional factors. Björk and Christensen (1999) propose to add a fourth factor to the approximating

[7]For the dataset I use here, the first and second factor explain 95.6% and 4% each of the variance in yield levels. The third factor explains only an additional 0.23%.

[8]Diebold, Piazzesi, and Rudebusch (2005) show how to impose no-arbitrage restrictions on this two-factor Nelson-Siegel model. As argued by Diebold *et al.* (2006b) no-arbitrage restrictions are likely to be, at least approximately, satisfied in the U.S. Treasury data analyzed here. Therefore, imposing no-arbitrage restrictions are unlikely to improve the performance of the two-factor model reported in Section 5.6 and 5.7. Its poor performance seems primarily due to the limited number of only two factors.

forward curve in (5.5):

$$f_t(\tau) = \beta_{1,t} + \beta_{2,t} \exp\left(-\frac{\tau}{\lambda_t}\right) + \beta_{3,t} \left(\frac{\tau}{\lambda_t}\right) \exp\left(-\frac{\tau}{\lambda_t}\right) + \beta_{4,t} \exp\left(-\frac{2\tau}{\lambda_t}\right) \tag{5.9}$$

The four-factor Nelson-Siegel yield curve is then given by

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left[\frac{1 - \exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)}\right] + \beta_{3,t} \left[\frac{1 - \exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)} - \exp\left(-\frac{\tau}{\lambda_t}\right)\right] + \beta_{4,t} \left[\frac{1 - \exp\left(-\frac{2\tau}{\lambda_t}\right)}{\left(\frac{2\tau}{\lambda_t}\right)}\right] \tag{5.10}$$

The fourth component, $\left[\frac{1-\exp(-2\tau/\lambda_t)}{(2\tau/\lambda_t)}\right]$, resembles the second component as it also mainly affects short-term maturities. The difference is that it decays to zero at a faster rate which can be seen from Figure 5.1[b]. The factor $\beta_{4,t}$ can therefore be interpreted as a second slope factor. As a result, the four-factor Nelson-Siegel model captures the slope of the term structure by the (weighted) sum of $\beta_{2,t}$ and $\beta_{4,t}$. The instantaneous short rate in (5.7) is for the four-factor model therefore equal to $y_t(0) = \beta_{1,t} + \beta_{2,t} + \beta_{4,t}$. Diebold, Rudebusch, and Aruoba (2006b) report that the four-factor model marginally improves the in-sample fit of the term structure but they do not consider out-of-sample forecasting.

Björk and Christensen (1999) also consider a five-factor model:

$$f_t(\tau) = \beta_{1,t} + \beta_{2,t} \left(\frac{\tau}{\lambda_t}\right) + \beta_{3,t} \exp\left(-\frac{\tau}{\lambda_t}\right) + \beta_{4,t} \left(\frac{\tau}{\lambda_t}\right) \exp\left(-\frac{\tau}{\lambda_t}\right) + \beta_{5,t} \exp\left(-\frac{2\tau}{\lambda_t}\right)$$

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left(\frac{\tau}{2\lambda_t}\right) + \beta_{3,t} \left[\frac{1-\exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)}\right] + \beta_{4,t} \left[\frac{1-\exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)} - \exp\left(-\frac{\tau}{\lambda_t}\right)\right] + \beta_{5,t} \left[\frac{1-\exp\left(-\frac{2\tau}{\lambda_t}\right)}{\left(\frac{2\tau}{\lambda_t}\right)}\right]$$

and Diebold, Rudebusch, and Aruoba (2006b) report that adding two additional factors again only leads to a negligible improvement in in-sample fit. The problem with the five-factor model, however, is that it contains a component which is linear in τ. Consequently, the model implies linearly increasing long-maturity spot and forward rates. This is problematic and I therefore do not consider the five-factor model here.

5.3.5 Bliss (1997) three-factor model

A second option to make the Nelson-Siegel more flexible is through relaxing the restriction that the slope and curvature component should be governed by the same decay parameter

λ_t. Bliss (1997) estimates the term structure of interest rates with the three-factor Nelson-Siegel model but allows for two different decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$[9]. The forward curve and spot rate curves are then given by

$$f_t(\tau) = \beta_{1,t} + \beta_{2,t} \exp\left(-\frac{\tau}{\lambda_{1,t}}\right) + \beta_{3,t}\left(\frac{\tau}{\lambda_{2,t}}\right)\exp\left(-\frac{\tau}{\lambda_{2,t}}\right) \tag{5.11}$$

and

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t}\left[\frac{1-\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)}{\left(\frac{\tau}{\lambda_{1,t}}\right)}\right] + \beta_{3,t}\left[\frac{1-\exp\left(-\frac{\tau}{\lambda_{2,t}}\right)}{\left(\frac{\tau}{\lambda_{2,t}}\right)} - \exp\left(-\frac{\tau}{\lambda_{2,t}}\right)\right] \tag{5.12}$$

Obviously, the Bliss Nelson-Siegel model will only be different from the base model if $\lambda_{1,t} \neq \lambda_{2,t}$.

Nelson and Siegel (1987) also consider an approximating forward curve with different decay parameters[10]. The forward curve is again derived as the solution to a second-order differential equation but now with real and unequal roots. Their forward rate curve is given by

$$f_t(\tau) = \beta_{1,t} + \beta_{2,t} \exp\left(-\frac{\tau}{\lambda_{1,t}}\right) + \beta_{3,t}\exp\left(-\frac{\tau}{\lambda_{2,t}}\right)$$

We need the additional factor $\left(\frac{\tau}{\lambda_{2,t}}\right)$ in (5.11) to obtain the curvature factor. Otherwise, the model contains two slope factors which are different only if the decay parameters are different. The model would then closely resemble the two-factor model.

5.3.6 Svensson (1994) four-factor model

A popular term-structure estimation method among central banks (see BIS, 2005) is the four-factor Svensson (1994) model. Svensson (1994) proposes to increase the flexibility and fit of the Nelson-Siegel model by adding a second hump-shape factor with its own separate decay parameter. The resulting four-factor forward curve is given by:

$$f_t(\tau) = \beta_{1,t}+\beta_{2,t}\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)+\beta_{3,t}\left(\frac{\tau}{\lambda_{1,t}}\right)\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)+\beta_{4,t}\left(\frac{\tau}{\lambda_{2,t}}\right)\exp\left(-\frac{\tau}{\lambda_{2,t}}\right) \tag{5.13}$$

[9] Bliss (1997) calls this the 'Extended Nelson-Siegel' but as several more extensions are considered here as well I simply refer to it as the 'Bliss' model.

[10] Nelson and Siegel (1987) try to fit this model to their sample of yields which only consists of maturities up until one year. They report that the model is over-parameterized and therefore use the forward curve in (5.5). Bliss (1997) remarks that over-parametrization should not pose any problem when also longer-maturity yields are fitted, which is also the case here.

The resulting equation for the zero-coupon yield curve is then

$$y_t(\tau)=\beta_{1,t}+\beta_{2,t}\left[\frac{1\text{–}\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)}{\left(\frac{\tau}{\lambda_{1,t}}\right)}\right]+\beta_{3,t}\left[\frac{1\text{–}\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)}{\left(\frac{\tau}{\lambda_{1,t}}\right)}-\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)\right]+\beta_{4,t}\left[\frac{1\text{–}\exp\left(-\frac{\tau}{\lambda_{2,t}}\right)}{\left(\frac{\tau}{\lambda_{2,t}}\right)}-\exp\left(-\frac{\tau}{\lambda_{2,t}}\right)\right] \tag{5.14}$$

The fourth component, $\left[\frac{1\text{–}\exp(-\tau/\lambda_{2,t})}{(\tau/\lambda_{2,t})}-\exp(-\tau/\lambda_{2,t})\right]$, introduces a second medium-term component to the model which is depicted by the dash-dotted line in Figure 5.1[c]. The Svensson Nelson-Siegel model can more easily fit term structure shapes with more that one local maximum or minimum along the maturity spectrum. As the fourth component mainly affects medium-term maturities, the limiting results in (5.7) also hold for the Svensson model.

5.3.7 Adjusted Svensson (1994) four-factor model

A potential problem with the Svensson model is that it is highly non-linear which can make the estimation of the model difficult, see Bolder and Stréliski (1999) for a discussion. A multicolinearity problem arises when the decay parameters $\lambda_{1,t}$ and $\lambda_{2,t}$ assume similar values. When this happens, the Svensson model reduces to the three-factor base model but with a curvature factor equal to the sum of $\beta_{3,t}$ and $\beta_{4,t}$. Only the sum of these parameters can then still be estimated efficiently, not the individual parameters[11].

One way to try and cure this multicolinearity problem is to make sure that the two medium-term components are different when $\lambda_{1,t} \simeq \lambda_{2,t}$. I therefore propose an 'Adjusted' Svensson model which is given by the following forward and zero curves:

$$f_t(\tau)=\beta_{1,t}+\beta_{2,t}\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)+\beta_{3,t}\left(\frac{\tau}{\lambda_t}\right)\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)+\beta_{4,t}\left[\exp\left(-\frac{\tau}{\lambda_{2,t}}\right)+\left(\frac{2\tau}{\lambda_{2,t}}-1\right)\exp\left(-\frac{2\tau}{\lambda_{2,t}}\right)\right] \tag{5.15}$$

and

$$y_t(\tau)=\beta_{1,t}+\beta_{2,t}\left[\frac{1\text{–}\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)}{\left(\frac{\tau}{\lambda_{1,t}}\right)}\right]+\beta_{3,t}\left[\frac{1\text{–}\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)}{\left(\frac{\tau}{\lambda_{1,t}}\right)}-\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)\right]+\beta_{4,t}\left[\frac{1\text{–}\exp\left(-\frac{\tau}{\lambda_{2,t}}\right)}{\left(\frac{\tau}{\lambda_{2,t}}\right)}-\exp\left(-\frac{2\tau}{\lambda_{2,t}}\right)\right]$$

[11] One example of this multicolinearity effect can be seen in Gimeno and Nave (2006). When applying the Svensson model to estimate the zero-yield curve from Spanish Treasury Bonds, Gimeno and Nave report that $\beta_{3,t}$ and $\beta_{4,t}$ display clear structural streaks and often take on large values but with opposite signs. The sum of the two parameters is stable across time, however (see Figure 3[a] in Gimeno and Nave, 2006). The reason for this becomes apparent from their Figure 2 in which it is shown that the extreme factor estimates correspond to samples for which the estimated values for $\lambda_{1,t}$ and $\lambda_{2,t}$ are very similar.

(5.16)

The adjustment to the second curvature component ensures that multicolinearity is no longer an issue. The adjusted component also starts at 0 but then increases for medium maturities at a faster rate than the first curvature component and returns to zero faster as well. The dash-dotted line in Figure 5.1[d] depicts the fourth component as a function of maturity. The difference between the two additional curvature components in the Svensson and Adjusted Svensson model can be seen by comparing Figure 5.1[c] with Figure 5.1[d].

5.3.8 General specification

The different Nelson-Siegel specifications that I examine are all nested and can therefore be captured in one general model set-up. In particular, consider the following state-space representation:

$$Y_t = X_t\beta_t + \varepsilon_t \tag{5.17}$$

$$\beta_t = \mu + \Phi\beta_{t-1} + \nu_t \tag{5.18}$$

The measurement equations in (5.17) specify the vector of yields, which contains N different maturities, $Y_t = [y_t(\tau_1)\ldots y_t(\tau_N)]'$, as the sum of a Nelson-Siegel spot rate curve, $X_t\beta_t$, plus a vector of yield errors which are assumed to be independent across maturities but with different variance terms, $\sigma^2(\tau_i)$. The Nelson-Siegel spot rate curves are those discussed in the previous sections with β_t being the $(K \times 1)$ vector of factors and X_t the $(N \times K)$ matrix of factor loadings which are potentially time-varying if the decay parameter(s) are estimated alongside the factors. Each of the Nelson-Siegel models in sections 5.1-5.7 is a special case of (5.17) with a different number of factors and/or a different specification for the factor loadings.

If we are only interested in fitting the term structure then the measurement equations are sufficient. However, in order to construct term structure forecasts we also need a model for the factor dynamics. I follow the dynamic frame-work of Diebold and Li (2006) and Diebold, Rudebusch, and Aruoba (2006b) by specifying first-order autoregressive processes for the factors as in the state equations (5.18). These can be either individual AR(1) processes or one multivariate VAR(1) process[12]. The vector μ and matrix Φ have dimensions

[12]Here I use a straightforward linear specification of the measurement and state equations. More complex specifications, such as a Markov Switching approach, are used in for example Bernadell *et al.* (2005).

$(K \times 1)$ and $(K \times K)$ respectively. The model is completed by assuming that the measurement equation and state equation error vectors are orthogonal and normally distributed:

$$\begin{bmatrix} \varepsilon_t \\ \nu_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{N\times 1} \\ 0_{K\times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) \tag{5.19}$$

where H is a $N \times N$ matrix which I assume to be diagonal throughout the analysis. For the state equation covariance matrix Q I make the assumption that it is either a diagonal $(K \times K)$ matrix or a full matrix, depending on the estimation procedure which I discuss next.

5.4 Estimation procedures

There are several approaches to estimating the latent factors and parameters in the Nelson-Siegel state-space representation. These approaches depend crucially on whether the measurement and state equations are estimated separately or simultaneously and on the assumptions regarding the decay parameters[13].

The most straightforward approach is used in for example Fabozzi *et al.* (2005) and Diebold and Li (2006) and consists of a two-step procedure. In the first step the measurement equations are treated as a cross-sectional model and Least Squares is used to estimate the parameters for every month separately. In the second step time series models are specified and fitted for the factors. A second, somewhat more demanding estimation approach is a one-step procedure in which all the parameters in the state-space system are estimated simultaneously. This approach uses the Kalman filter to estimate the factors and is proposed in Diebold, Rudebusch, and Aruoba (2006b). Here I use the one-step as well as the two-step estimation procedures and in this section I discuss both techniques in detail. Specific details regarding the estimation are given in Appendix A.

5.4.1 Two-step approach with a fixed decay parameter

Diebold and Li (2006) suggest to fix λ_t in the three-factor model to a pre-specified value which is the same for every t, instead of treating it as an unknown parameter. By doing so, the nonlinear measurement equations become linear in the state vector which can then be

[13] I only use frequentist maximum likelihood techniques to estimate parameters. Mönch (2006b) and De Pooter *et al.* (2007) consider Bayesian estimation of the three-factor model. Whereas a Bayesian approach would also account for parameter uncertainty, I do not pursue it here and leave this for further research.

estimated using straightforward cross-sectional OLS. The decay parameter λ_t determines the (medium-term) maturity at which the factor loading on the curvature factor $\beta_{3,t}$ is at its maximum. The value of 16.42 that Diebold and Li (2006) use for λ_t is such that this maximum is reached at a 30-month maturity. Larger values for λ_t produces slower decaying factor loadings with the curvature factor achieving it maximum at a longer maturity and vice versa. Although other authors have used different values as well, I follow Diebold and Li (2006) and set λ_t equal to 16.42.

The first step of the estimation produces time-series of estimated values for each of the K factors; $\{\beta_{i,t}\}_{t=1}^{T}$ for $i = 1, \dots, K$. The next step is to estimate the factor dynamics of the state equations. I estimate separate AR(1) models for each factor, thus assuming that Φ and Q are both diagonal, as well as a joint VAR(1) by assuming that Φ and Q are full matrices instead.

I apply the two-step estimation approach with a fixed decay parameter only to estimate the two, three and four-factor Nelson-Siegel specifications. The remaining models have two decay parameters and would therefore require finding two appropriate values to choose for $\lambda_{1,t}$ and $\lambda_{2,t}$, which is difficult. I use the notation 'NS2' to indicate the two-step estimation procedure. I denote the two, three and four-factor models by **NS2-2**, **NS2-3** and **NS2-4** respectively and add suffixes '-AR' and '-VAR' to indicate the time-series model specification for the state equations.

5.4.2 Two-step approach with estimated decay parameters

When the decay parameters are estimated alongside the factors, the estimation of the now nonlinear measurement equations in the first step becomes more challenging and requires nonlinear least squares. However, the increased flexibility of the model as a result of the additional parameter can nevertheless make this a worthwhile exercise to undertake. I therefore also estimate the two, three and four-factor models when treating λ_t as a parameter and I denote these by **NS2-2-$\boldsymbol{\lambda}$**, **NS2-3-$\boldsymbol{\lambda}$** and **NS2-4-$\boldsymbol{\lambda}$** with suffixes '-AR' and '-VAR'. The Bliss and (Adjusted) Svensson models all have two decay parameters, $\lambda_{1,t}$ and $\lambda_{2,t}$ which should even further improve the fit of the Nelson-Siegel model due to the increased flexibility of the factor loadings. The two-step estimation procedure can also be applied to these models and I use the notation **NS2-B**, **NS2-S** and **NS2-AS** for the Bliss, Svensson and Adjusted Svensson model respectively. Note that in the second step I do not model the dynamics of the decay parameters explicitely. Instead, in order to construct forecasts I use the median of their in-sample estimated values, see Section 5.7.1 for further details.

5.4.3 Restrictions on the decay parameters

The nonlinear estimation procedure can result in factor estimates which can sometimes be very extreme. An example is shown and discussed in Gimeno and Nave (2006) for the Svensson model. Gimeno and Nave report extreme (and often offsetting) values for factor estimates. Bolder and Stréliski (1999) also address numerical problems and estimation issues when estimating the Svensson model.

The nonlinear model structure seems to pose serious difficulties for optimization procedures to arrive at reasonable estimates. An additional reason, which to my knowledge seems to have been overlooked in the literature surprisingly, is the behavior of the factor loadings when the decay parameters take on extreme values. When this happens multicollinearity problems can occur and some of the factors are then no longer uniquely identified. To understand why this is the case we need to examine the factor loadings as functions of λ_t. We have the following straightforward limiting results

$$\lim_{\lambda_t \downarrow 0}\left[\frac{1-\exp(-\tau/\lambda_t)}{(\tau/\lambda_t)}\right]=0; \qquad \lim_{\lambda_t \downarrow 0}\left[\frac{1-\exp(-\tau/\lambda_t)}{(\tau/\lambda_t)}-\exp(-\tau/\lambda_t)\right]=0 \tag{5.20}$$

$$\lim_{\lambda_t \to \infty}\left[\frac{1-\exp(-\tau/\lambda_t)}{(\tau/\lambda_t)}\right]=1; \qquad \lim_{\lambda_t \to \infty}\left[\frac{1-\exp(-\tau/\lambda_t)}{(\tau/\lambda_t)}-\exp(-\tau/\lambda_t)\right]=0 \tag{5.21}$$

The results in (5.20) imply that for very small values of λ_t the slope and curvature factors will be near non-identification which can result in extreme estimates[14]. For large values of λ_t, as indicated by (5.21), curvature factors are nearly non-identified. Furthermore, the level and slope factors are jointly identified, but no longer identified separately and can therefore take on extreme, offsetting, values[15].

If we are only interested in fitting the term structure at a given point in time in step one, these non-identification issues do not necessarily cause problems. Although the factor estimates can be extreme, the models still accurately fit the term structure. The real problem occurs when estimating dynamics in step two as the time-series of the factors can potentially be plagued by outliers. In order to prevent extreme factor estimates, I impose restrictions on the decay parameters. By only allowing the curvature factor loading to

[14] Note that in the Bliss and (Adjusted) Svensson models non-identification issues arise when either $\lambda_{1,t}$ or $\lambda_{2,t}$ tends to zero and even more so when both parameters tend to zero.

[15] This explains the peaks with opposite signs in the level and slope estimates in Figure 1 of Gimeno and Nave (2006).

reach its maximum for maturities between one and five years, the decay parameters are restricted to lie in the interval [6.69, 33.46][16]. I impose one additional restriction on the Svensson and Adjusted Svensson models separately.

For the Svensson model I restrict the loading on the second curvature factor, $\beta_{4,t}$, to reach its maximum for a maturity which is at least twelve months shorter than the corresponding maturity for the first curvature loading. Specifically it comes down to the following minimum distance restriction: $\lambda_{1,t} \geq \lambda_{2,t} + 6.69$. This restriction prevents the case where $\beta_{3,t}$ and $\beta_{4,t}$ are only jointly identified but not individually. Note that the two curvature components in the Svensson model, and therefore $\lambda_{1,t}$ and $\lambda_{2,t}$, as well as their role in the restriction, are interchangeable. In the Adjusted Svensson model the curvature factor loadings are different so there is no need to impose any minimum distance between the two decay parameters. I do, however, again force the first curvature hump to be to the right of the second curvature hump by imposing the restriction $\lambda_{1,t} \geq \lambda_{2,t}$.

5.4.4 One-step state-space approach

The alternative to the two-step approach is to estimate all parameters simultaneously. By using the prediction-error decomposition of the likelihood we can estimate parameters by maximum likelihood and apply the Kalman filter to obtain optimal factor estimates. The likelihood for the state-space system in (5.17)-(5.18) is given by

$$\mathcal{L} = \sum_{t=1}^{T} \left[-\frac{1}{2} \ln (2\Pi) - \frac{1}{2} \ln (|f_{t|t-1}|) - \frac{1}{2} \eta'_{t|t-1} f_{t|t-1}^{-1} \eta_{t|t-1} \right] \tag{5.22}$$

which is a function of the parameter set $\Theta = (\lambda_1, \lambda_2, \beta_t, \mu, \Phi, H, Q)$. The likelihood is comprised of the $(N \times 1)$ yield prediction error vector; $\eta_{t|t-1} \equiv y_t - y_{t|t-1}$ where $y_{t|t-1}$ is the vector of in-sample yield forecasts given information up to time $t-1$, and of the $(N \times N)$ conditional covariance matrix of the prediction errors; $f_{t|t-1} \equiv \mathbb{E}[\eta_{t|t-1}\eta'_{t|t-1}]$, see Kim and Nelson (1999) for further details. Note that the decay parameters are assumed

[16] Recall that Diebold and Li (2006) fix the decay parameter such that the maximum is reached at a maturity of two and a half years. Note that Gürkanyak, Sack, and Wright (2006) do not impose restrictions on the Svensson model when estimating the U.S. Treasury yield curve. They find that the second hump is located at much longer maturities (beyond twenty years). However, Gürkanyak *et al.* (2006) estimate the term structure using bonds with maturities up to thirty years. I only use maturities up to ten years and the domain of the curvature humps of one to five years seems therefore reasonable and sufficiently wide in order not to be too restrictive. Some experimentation with using wider domains indeed resulted in factor estimates that were more ‘extreme’.

to be constant over time[17]. As these can now be estimated using information from both the cross-section as well as the time-series of yields it is much less likely that they will take on extreme values. Furthermore, because the dynamics of the factors are explicitly taken into account when optimizing the likelihood, it does not seem necessary anymore to impose the earlier restrictions on the decay parameters. I estimate all the models using this one-step procedure with the decay parameters being estimated alongside the factors and the remaining parameters and allowing for Q to be a full matrix. I denote the results for the different models using this approach by **NS1-2**, **NS1-3**, **NS1-4**, **NS1-B**, **NS1-S** and **NS1-AS**.

Diebold, Rudebusch, and Aruoba (2006b) favor the one-step over the two-step estimation approach because parameters are estimated simultaneously which ensures that the uncertainty of all parameters is taken into account at the same time. The drawback, however, is that the number of parameters to estimate is substantial in the state-space model. For example, for the four-factor Svensson model with a VAR(1) specification for the state equations, the total number of parameters for the dataset used here equals 49 (two decay parameters, four parameters in μ, 16 parameters in Φ, $N = 17$ parameters in H, four variance and six covariance terms in Q). In order to reduce the number of parameters I therefore also try two alternative specifications for the system of state equations in (5.18). Apart from specifying VAR(1) dynamics for the factors by assuming Φ to be a full matrix as in Diebold, Rudebusch, and Aruoba (2006b) I also consider AR(1) dynamics using a diagonal Φ. Additionally I specify random walk factor dynamics by setting μ equal to zero and Φ equal to the identity matrix[18]. I distinguish the different dynamics by using the suffixes '-VAR', '-AR' and '-RW' for respectively VAR(1), AR(1) and random walk dynamics. An overview of all model abbreviations used is given in Table 5.1.

5.5 Data

The dataset available here consists of end-of-month continuously compounded U.S. zero-coupon bond forward rates. I compute constant maturity spot rates by averaging these forwards rates as in (5.2). The forward rates are constructed from filtered average bid-ask

[17] Huse (2007), on the contrary, argues that for the three-factor model the decay parameter should also be treated as a dynamic factor and that it should be modelled accordingly. I do not consider this approach here.

[18] Diebold and Li (2006) find for the three-factor model that the null of a unit root in the factor dynamics cannot be rejected for $\beta_{1,t}$ and $\beta_{2,t}$. Fabozzi *et al.* (2005) find similar results and therefore model first differences of the level and slope factors.

Table 5.1: Model abbreviations

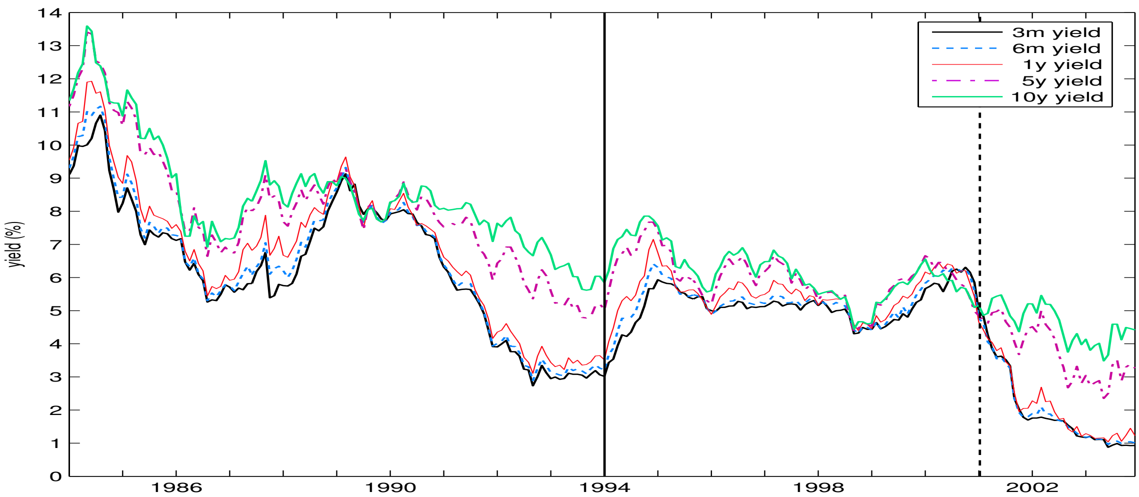
Two-step models	model description	factor dynamics
NS2-2-AR	2-factor NS model with λ fixed to 16.42	AR(1) per factor
NS2-2-VAR	2-factor NS model with λ fixed to 16.42	VAR(1) for factors
NS2-3-AR	3-factor NS model with λ fixed to 16.42	AR(1) per factor
NS2-3-VAR	3-factor NS model with λ fixed to 16.42	VAR(1) for factors
NS2-4-AR	4-factor NS model with λ fixed to 16.42	AR(1) per factor
NS2-4-VAR	4-factor NS model with λ fixed to 16.42	VAR(1) for factors
NS2-2-λ-AR	2-factor NS model treating λ as parameter	AR(1) per factor
NS2-2-λ-VAR	2-factor NS model treating λ as parameter	VAR(1) for factors
NS2-3-λ-AR	3-factor NS model treating λ as parameter	AR(1) per factor
NS2-3-λ-VAR	3-factor NS model treating λ as parameter	VAR(1) for factors
NS2-4-λ-AR	4-factor NS model treating λ as parameter	AR(1) per factor
NS2-4-λ-VAR	4-factor NS model treating λ as parameter	VAR(1) for factors
NS2-B-AR	3-factor Bliss NS model treating λ_1, λ_2 as parameters	AR(1) per factor
NS2-B-VAR	3-factor Bliss NS model treating λ_1, λ_2 as parameters	VAR(1) for factors
NS2-S-AR	4-factor Svensson NS model treating λ_1, λ_2 as parameters	AR(1) per factor
NS2-S-VAR	4-factor Svensson NS model treating λ_1, λ_2 as parameters	VAR(1) for factors
NS2-AS-AR	4-factor Adjusted Svensson model treating λ_1, λ_2 as parameters	AR(1) per factor
NS2-AS-VAR	4-factor Adjusted Svensson model treating λ_1, λ_2 as parameters	VAR(1) for factors
One-step models	**model description**	**factor dynamics**
NS1-2-RW	2-factor NS model	random walk per factor
NS1-2-AR	2-factor NS model	AR(1) per factor
NS1-2-VAR	2-factor NS model	VAR(1) for factors
NS1-3-RW	3-factor NS model	random walk per factor
NS1-3-AR	3-factor NS model	AR(1) per factor
NS1-3-VAR	3-factor NS model	VAR(1) for factors
NS1-4-RW	4-factor NS model	random walk per factor
NS1-4-AR	4-factor NS model	AR(1) per factor
NS1-4-VAR	4-factor NS model	VAR(1) for factors
NS1-B-RW	3-factor Bliss NS model	random walk per factor
NS1-B-AR	3-factor Bliss NS model	AR(1) per factor
NS1-B-VAR	3-factor Bliss NS model	VAR(1) for factors
NS1-S-RW	4-factor Svensson NS model	random walk per factor
NS1-S-AR	4-factor Svensson NS model	AR(1) per factor
NS1-S-VAR	4-factor Svensson NS model	VAR(1) for factors
NS1-AS-RW	4-factor Adjusted Svensson NS model	random walk per factor
NS1-AS-AR	4-factor Adjusted Svensson NS model	AR(1) per factor
NS1-AS-VAR	4-factor Adjusted Svensson NS model	VAR(1) for factors

Notes: The table gives model abbreviations used in the subsequent tables and graphs. 'NS' stands for Nelson-Siegel model. For the two-step models (top panel), the factors are estimated in a first step using least squared applied to the cross-section of yields in each month. In the second step, the dynamics of the estimated factors from the first step are estimated. For the one-step models (bottom panel), all parameters are estimated simultaneously as a state-space model using the Kalman filter. The table shows model abbreviations in the first column, model descriptions in the second column and the specification for the factor dynamics in the third column.

price quotes on U.S. Treasury securities using the Fama and Bliss (1987) bootstrap method as outlined in Bliss (1997)[19]. The price quotes are taken from the CRSP government bond files. CRSP filters the available quotes by taking out illiquid bonds and bonds with option features. Similar to Diebold and Li (2006), Diebold, Rudebusch, and Aruoba (2006b) and Mönch (2006a), I use unsmoothed Fama-Bliss yields[20].

I estimate the class of Nelson-Siegel models using data for the sample period 1984:1 - 2003:12 ($T = 240$ observations) and I use the following $N = 17$ maturities in the estimation: $\tau = 3, 6, 9, 12, 15, 18, 21, 24$ and 30 months as well as 3, 4,..., 10 years. I start my dataset after the Volcker period to allow for a fair comparison with the results in Diebold and Li (2006) and Mönch (2006a). Note that the forecasting results reported by De Pooter *et al.* (2007) for the three-factor Nelson-Siegel model with both the two-step and one-step estimation procedure are based on a much longer span of data (1970:1 - 2003:12).

Figure 5.2: U.S. zero-coupon yields

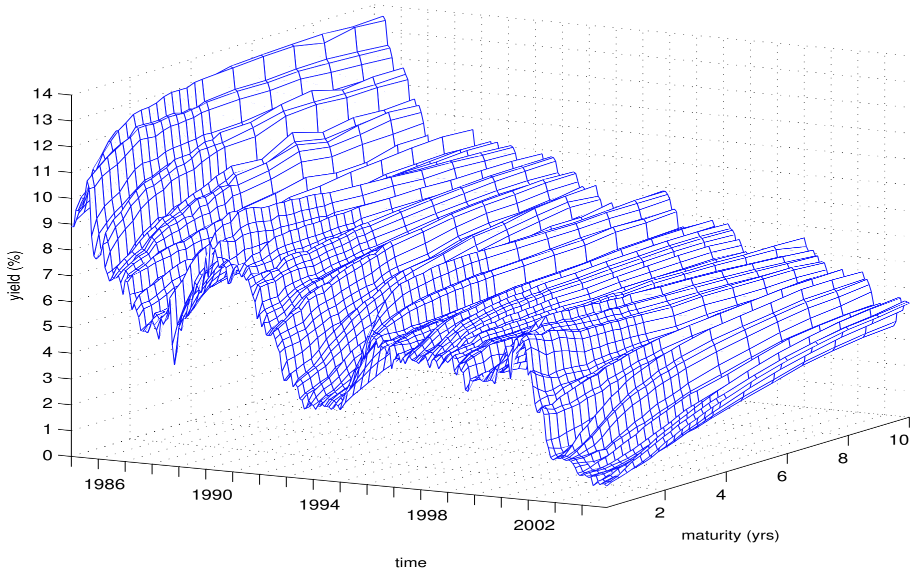


Notes: The figure shows time-series plots for a subset of maturities of end-of-month U.S. zero coupon yields constructed using the unsmoothed Fama and Bliss (1987) bootstrap method. Sample period is January 1984 - December 2003 (240 observations). The solid vertical line indicates the start of the forecasting sample (January 1994 - December 2003). The dotted line divides the forecast sample into two subsamples (January 1994 - December 2000 and January 2001 - December 2003).

[19]I kindly thank Robert Bliss for providing me with the unsmoothed Fama-Bliss forward rates and the programs to construct the spot rates.

[20]The reason for using unsmoothed Fama-Bliss yields is that Bliss (1997) finds, using parametric and nonparametric tests, that the Fama-Bliss method does best overall in terms of estimating the term structure in comparison with other popular estimation methods.

Figure 5.3: U.S. zero-coupon yields



Notes: The figure shows a 3-dimension plot of the panel of end-of-month U.S. zero coupon yields constructed using the unsmoothed Fama and Bliss (1987) bootstrap method. Sample period is January 1984 - December 2003 (240 observations).

Table 5.2: Summary statistics

maturity	mean	stdev	skew	kurt	min	max	JB-p	ρ_1	ρ_{12}	ρ_{24}
1-month	5.051	2.090	0.045	2.864	0.794	10.727	0.845	0.968	0.552	0.121
3-month	5.286	2.175	-0.005	2.843	0.876	10.905	0.849	0.977	0.563	0.148
6-month	5.434	2.226	0.024	2.951	0.958	11.169	0.962	0.977	0.559	0.162
1-year	5.707	2.290	0.051	3.056	1.040	11.928	0.946	0.976	0.562	0.190
2-year	6.083	2.281	0.234	3.351	1.299	12.777	0.201	0.974	0.555	0.229
3-year	6.365	2.211	0.357	3.465	1.618	13.115	0.031	0.973	0.561	0.268
4-year	6.589	2.168	0.481	3.524	1.999	13.268	0.003	0.972	0.572	0.297
5-year	6.711	2.125	0.599	3.613	2.351	13.410	0.000	0.972	0.574	0.319
6-year	6.878	2.108	0.666	3.574	2.663	13.493	0.000	0.973	0.589	0.335
7-year	6.967	2.061	0.761	3.727	3.003	13.554	0.000	0.972	0.577	0.333
8-year	7.058	2.019	0.751	3.650	3.221	13.596	0.000	0.972	0.590	0.359
9-year	7.106	2.001	0.753	3.584	3.389	13.529	0.000	0.973	0.599	0.371
10-year	7.102	1.982	0.783	3.598	3.483	13.595	0.000	0.973	0.600	0.373

Notes: The table shows summary statistics for end-of-month unsmoothed continuously compounded U.S. zero-coupon yields. The results shown are for annualized yields (expressed in precentages). The sample period is January 1984 - December 2003 (240 observations). Reported are the mean, standard deviation, skewness, kurtosis, minimum, maximum, the p-value of the Jarque-Bera test statistic for normality and the 1^{st}, 12^{th} and 24^{th} sample autocorrelation.

Figure 5.2 shows time-series plots for a subset of the maturities and illustrates how yield levels and spreads vary substantially throughout the sample. For example, for the period from 1994 onwards, which is the period I use to evaluate the models' forecasting performance, we can distinguish a stable period (mid 1990s till the end of 2000) but also a period where short term interest rates fell by roughly 4%, resulting in a sharp increase in the term spread (the last three years of the sample). It is clear from Figure 5.3 that not only the level of the term structure fluctuates over time but also its slope and curvature. The curve takes on various forms ranging from nearly flat to (inverted) S-type shapes. Table 5.2 reports summary statistics for yield levels for various maturities. The stylized facts common to yield curve data are clearly present: the sample average curve is upward sloping and concave, volatility is decreasing with maturity, autocorrelations are very high and increasing with maturity. The null of normality is rejected for medium and longer term maturities due to positive skewness and excess kurtosis but can be accepted for shorter maturities. Correlations between yields of different maturities are high (80% or above), especially for close-together maturities.

5.6 In-sample fit results

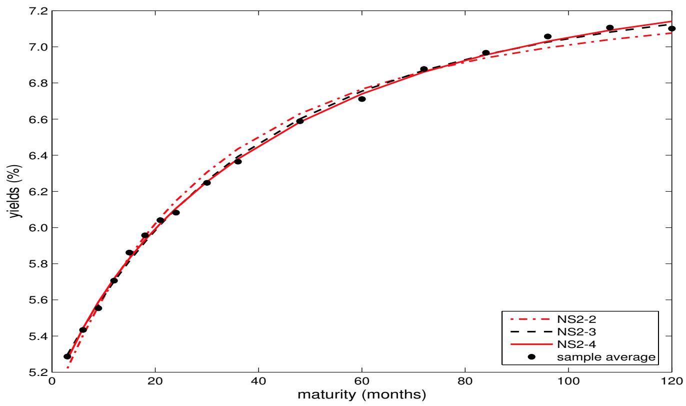
5.6.1 In-sample fit

In this section I discuss the results of fitting the term structure using the class of Nelson-Siegel models. I only focus on the fit from step one of the two-step estimation procedure due to the fact that the one-step procedure potentially also uses (future) time-series information which is unavailable if we want to fit the term structure at a given point in time. We can expect that more flexible models result in a better fit. However, as the increased flexibility can be obtained both by additional decay parameters as well as by additional factors, the question is which of the two modelling options improves the fit more.

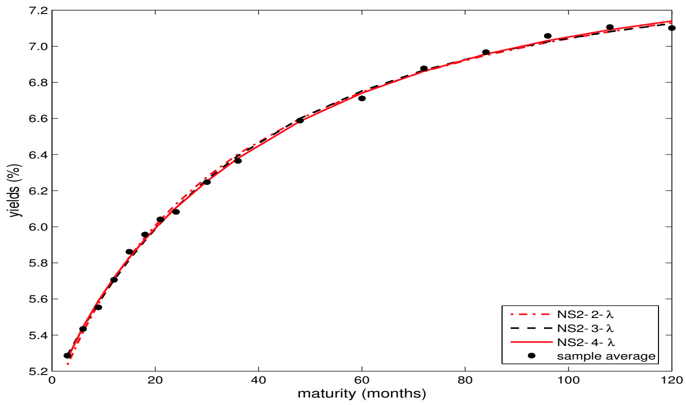
Figure 5.4 shows that all models accurately fit the average curve. The only exception is the two-factor model with fixed decay parameter, shown in Panel [a], most likely because it lacks a curvature component. Nevertheless, freeing up the decay parameter seems to provide sufficient additional flexibility as the two-factor average curve now becomes virtually indistinguishable from the three and four-factor models (Panel [b]).

Whereas the average fit may be nearly identical across the different models, Figure 5.5 on the other hand shows that the fit in individual months can be quite different. Shown in

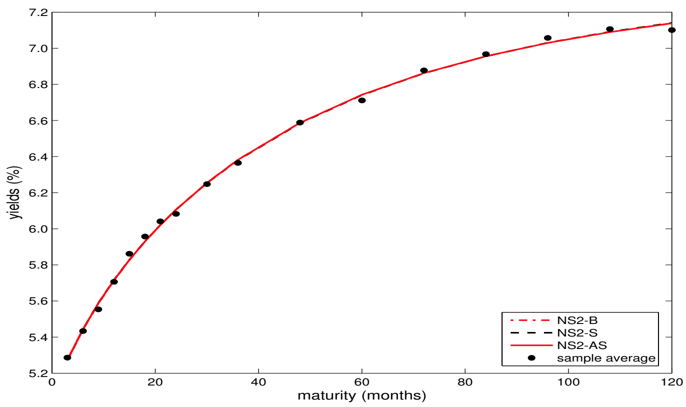
Figure 5.4: Fitted average yield curve



[a] NS2-2, NS2-3, NS2-4



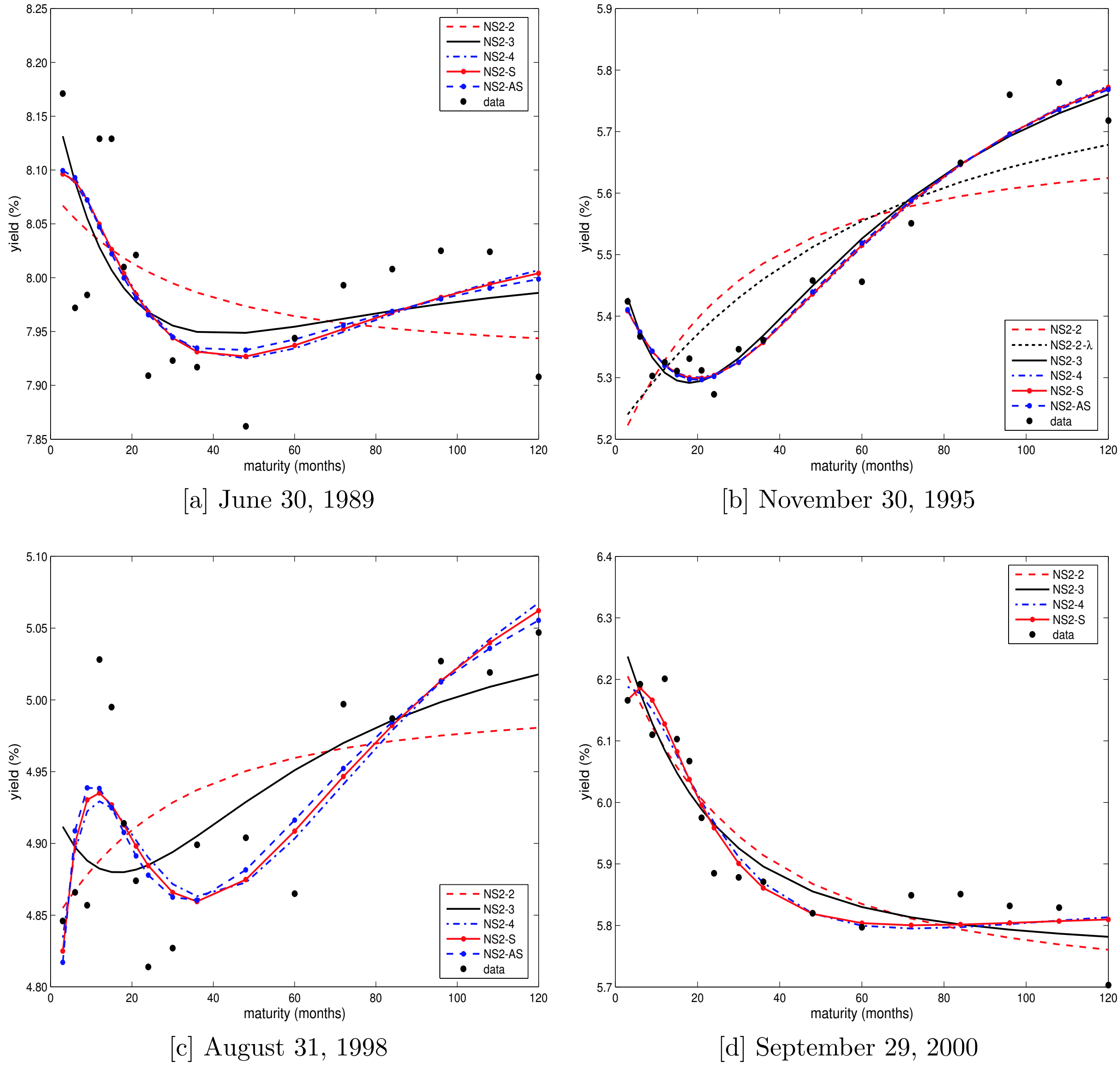
[b] NS2-2-λ, NS2-3-λ, NS2-4-λ



[c] NS2-B, NS2-S, NS2-AS

Notes: The graph shows the average fitted curve for different Nelson-Siegel models. Panel [a] shows the estimated average curve for the two-factor, three-factor and four-factor. Panel [b] shows the average curve for the same models but where now λ is estimated alongside (β_1 β_2 β_3 β_4). Finally, Panel [c] depicts the average curve for the three-factor Bliss model, the four-factor model Svensson extension (second column) and the Adjusted Svensson model. The dots in each graph are the actual sample averages. The solid and (dash-)dotted lines depict the fitted lines. The sample period is 1984:1 - 2003:12 (240 observations).

Figure 5.5: Fitted yield curve for specific months



[a] June 30, 1989

[b] November 30, 1995

[c] August 31, 1998

[d] September 29, 2000

Notes: The graph depicts the actual yield curve (black dots) and the fitted yield curve for a subset of models. Shown are four months from the full sample 1984:1 - 2003:12 (240 observations): [a] June 30, 1989, [b] November 30, 1995, [c] August 31, 1998 and [d] September 29, 2000. The fitted curve is shown for the two-factor, three-factor and four-factor model with fixed λ, the two-factor model where λ is estimated alongside β_1 and β_2, the Svensson model and the Adjusted Svensson model.

Figure 5.5 are the actual term structures in four specific months of the sample. These four months are an example of the various different term structure shapes that occur in the data. Whereas for November 1995 and September 2000 the shape is respectively S-shaped and downward sloping, for June 1989 and August 1998 the shapes are more difficult to describe. The lines in each panel show the fit of various models. The two-factor model in particular has difficulties fitting the more complex curves, but the three-factor model also does not seem flexible enough judging from, for example, Panel [c]. Graphically, the best fit is obtained with the four-factor model and the (Adjusted) Svensson models which give very similar fitted curves.

Table 5.3 reports detailed in-sample results for all models, which have been estimated with the restrictions on the decay parameters in place. The best fitting models, as judged by a number of standard criteria given in the table (standard deviation of yield errors, root mean squared fit error, mean absolute fit error, minimum and maximum fit error) are represented by the bold numbers. The results can be summarized by making the following observations.

The models that achieve the best fit overall are indeed the most flexible models, in particular the (Adjusted) Svensson model. For nearly every maturity shown in the table, the Svensson models are the most accurate on all criteria, including having the lowest persistence in yield errors. Except from the two-factor model, all models perform relatively similar, however, which agrees with the results in Dahlquist and Svensson (1996) and Diebold, Rudebusch, and Aruoba (2006b) who demonstrate that the three-factor model fits the term structure well compared to more elaborate models. It is nonetheless interesting to examine how the results of the remaining models compare to those of the Svensson models, but in particular how they compare amongst each other. For the two and three-factor models we can judge which extension yields the largest gain; estimating λ_t or adding a factor. From columns two to six in Table 5.3 it becomes clear that for the two-factor model adding a (curvature) factor improves the in-sample fit much more than by estimating λ_t alongside the level and slope factors. The results for the three-factor model lead to the same conclusion although the improvement when going from the three to the four-factor model is much less substantial than going from the two to the three-factor model. Estimating λ_t instead of using the fixed value of 16.42 improves the fit for each model although in absolute terms the benefits are minor (tens of basis points). Another comparison to make is that between the three-factor model with estimated λ_t and the Bliss model as the latter does not impose that the slope and curvature factor are determined by the same decay

Table 5.3: In-sample fit: restricted decay parameters

maturity	NS2-2	NS2-3	NS2-4	NS2-2-λ	NS2-3-λ	NS2-4-λ	NS2-B	NS2-S	NS2-AS
					Mean Error				
3-month	6.55	-0.97	1.47	5.06	-0.26	1.65	1.09	1.50	1.36
6-month	3.06	-0.93	-0.86	1.67	-0.75	-0.97	-0.86	-0.92	-0.93
1-year	-0.23	0.50	-0.98	-0.52	0.22	-1.07	-0.70	-0.98	-0.87
2-year	-6.52	-2.17	-2.34	-4.30	-2.42	-2.23	-2.34	-2.31	-2.28
5-year	-5.50	-4.36	-2.86	-3.55	-4.13	-2.95	-3.10	-2.92	-3.12
10-year	2.53	-2.40	-3.97	-2.82	-2.44	-3.87	-3.75	-3.94	-3.69
					Standard Deviation				
3-month	32.06	8.59	3.12	17.09	5.25	2.74	3.98	2.59	**2.23**
6-month	15.88	4.12	4.11	9.07	4.20	3.94	4.05	3.96	**3.91**
1-year	9.23	8.11	5.49	8.26	6.70	5.46	6.17	5.42	**5.33**
2-year	16.53	4.72	4.70	8.57	4.77	4.45	4.57	**4.40**	4.45
5-year	7.21	5.90	4.74	6.14	5.40	4.62	5.11	4.67	**4.60**
10-year	20.07	7.12	5.15	10.26	6.08	4.93	5.30	**4.73**	4.89
					Root Mean Squared Error				
3-month	32.72	8.64	3.45	17.82	5.26	3.20	4.12	3.00	**2.62**
6-month	16.17	4.23	4.19	9.22	4.27	4.06	4.14	4.07	**4.02**
1-year	9.24	8.12	5.58	8.28	6.70	5.56	6.21	5.51	**5.40**
2-year	17.77	5.20	5.25	9.59	5.35	4.98	5.14	**4.96**	5.00
5-year	9.07	7.34	5.54	7.09	6.80	**5.48**	5.98	5.50	5.55
10-year	20.23	7.52	6.50	10.64	6.55	6.27	6.50	6.16	**6.12**
					Mean Absolute Error				
3-month	25.56	6.39	2.76	12.64	3.71	2.55	3.06	2.34	**1.93**
6-month	12.19	3.06	3.03	6.19	3.10	2.95	3.00	**2.94**	2.97
1-year	7.45	6.37	4.51	6.58	5.24	4.42	4.84	4.35	**4.25**
2-year	13.71	3.70	3.75	6.81	3.75	3.58	3.63	**3.56**	3.64
5-year	7.45	6.13	4.33	5.88	5.53	4.21	4.75	**4.19**	4.29
10-year	15.71	5.91	5.25	8.31	5.21	5.06	5.14	4.98	**4.94**
					Minimum Error				
3-month	-86.81	-34.51	-11.65	-49.15	-22.38	-7.83	-16.67	-5.98	**-4.13**
6-month	-40.00	-14.11	-13.82	-20.69	-13.29	-12.66	-13.02	**-12.63**	-12.72
1-year	-20.74	-18.33	-17.91	-20.75	-16.61	-20.70	-20.31	-20.07	**-19.69**
2-year	-46.90	-19.05	-20.10	-27.83	-21.67	-18.32	-19.22	-17.15	**-16.75**
5-year	-27.39	-19.89	-20.23	-18.38	-17.20	-23.54	**-17.20**	-23.59	-23.46
10-year	-37.58	-25.57	-18.38	-41.31	-18.39	**-18.37**	-18.39	-19.09	-19.28
					Maximum Error				
3-month	75.52	21.75	10.60	54.67	12.43	9.03	12.44	8.97	**8.97**
6-month	44.64	21.81	22.10	34.28	22.16	22.09	22.20	21.22	**18.55**
1-year	28.83	26.69	13.48	22.14	21.99	12.03	17.86	11.96	**11.89**
2-year	36.97	**16.64**	16.97	22.28	18.98	16.98	18.67	18.91	18.33
5-year	20.39	18.62	12.37	19.48	13.39	**10.20**	13.39	10.87	11.53
10-year	53.93	16.41	7.96	24.75	17.02	7.97	9.25	8.05	**7.45**
					$\widehat{\rho}_1$				
3-month	0.907	0.754	0.483	0.817	0.689	0.483	0.473	0.417	0.435
6-month	0.875	0.276	0.270	0.688	0.248	0.278	0.271	0.244	0.322
1-year	0.659	0.582	0.386	0.615	0.510	0.390	0.417	0.378	0.369
2-year	0.913	0.649	0.628	0.759	0.613	0.615	0.625	0.597	0.622
5-year	0.805	0.740	0.644	0.746	0.696	0.642	0.606	0.602	0.609
10-year	0.889	0.627	0.488	0.706	0.550	0.442	0.408	0.405	0.438
					$\widehat{\rho}_{12}$				
3-month	0.347	0.087	0.102	0.281	0.023	0.192	0.052	0.174	0.119
6-month	0.430	0.203	0.188	0.304	0.159	0.239	0.215	0.220	0.240
1-year	0.347	0.296	0.370	0.311	0.338	0.356	0.330	0.353	0.356
2-year	0.295	0.129	0.132	0.091	0.129	0.100	0.104	0.102	0.108
5-year	0.099	0.046	-0.092	-0.060	-0.099	-0.112	-0.116	-0.122	-0.121
10-year	0.394	0.297	0.305	0.190	0.205	0.298	0.215	0.264	0.250

Notes: The table show in-sample fit error statistics for the full sample 1984:1-2003:12 (240 observations). The statistics are expressed in basis points. Results are shown for the models with λ_t fixed to 16.42 [NS2-2, NS2-3, NS-4], with λ estimated (but restricted) [NS-2-λ, NS-3-λ, NS-4-λ], the Bliss extension [NS2-B] and the adjusted Svensson model [NS2-(A)S]. The statistics $\widehat{\rho}_1$ and $\widehat{\rho}_{12}$ represent the 1st and 12th autocorrelation of the yield errors. For selected statistics, bold numbers indicate the best performing model.

Table 5.4: In-sample fit: unrestricted decay parameters

maturity	NS2-2-λ	NS2-3-λ	NS2-4-λ	NS2-B	NS2-S	NS2-AS
			Mean Error			
3-month	4.36	-0.41	1.67	1.52	1.33	0.85
6-month	0.34	-0.82	-0.98	-0.88	-0.37	0.10
1-year	-1.30	0.27	-1.10	-0.99	-1.15	-1.10
2-year	-3.45	-2.33	-2.23	-2.31	-2.31	-2.43
5-year	-1.95	-4.32	-2.91	-2.91	-2.94	-2.83
10-year	-5.42	-2.15	-3.93	-3.82	-3.49	-3.63
			Standard Deviation			
3-month	11.18	5.12	2.82	2.96	2.32	**2.29**
6-month	6.03	4.14	3.97	3.93	3.70	**3.45**
1-year	8.14	6.57	5.41	5.65	5.33	**5.15**
2-year	6.43	4.67	4.55	4.40	4.40	**4.37**
5-year	6.37	5.11	4.55	4.52	**4.37**	4.39
10-year	8.63	5.67	4.94	4.38	**4.20**	4.33
			Root Mean Squared Error			
3-month	12.00	5.13	3.28	3.33	2.67	**2.45**
6-month	6.04	4.22	4.09	4.03	3.72	**3.45**
1-year	8.25	6.58	5.52	5.73	5.45	**5.27**
2-year	7.30	5.22	5.06	**4.96**	**4.96**	5.00
5-year	6.66	6.69	5.41	5.38	5.26	**5.22**
10-year	10.19	6.07	6.31	5.81	**5.46**	5.65
			Mean Absolute Error			
3-month	8.72	3.58	2.62	2.58	1.91	**1.58**
6-month	4.54	3.03	2.97	2.90	2.75	**2.59**
1-year	6.73	5.16	4.46	4.57	4.24	**4.14**
2-year	5.42	3.67	3.65	**3.56**	**3.56**	3.67
5-year	5.40	5.39	4.21	4.16	**3.95**	3.96
10-year	7.89	4.88	5.12	4.79	**4.45**	4.63
			Minimum Error			
3-month	-41.06	-22.44	-8.14	-7.63	**-3.96**	-12.38
6-month	-19.68	-13.33	-12.80	-12.90	**-12.72**	-12.77
1-year	-20.75	**-16.69**	-20.78	-20.40	-20.32	-20.27
2-year	-21.66	-21.67	-18.01	**-17.15**	**-17.15**	-17.17
5-year	-17.24	-17.25	-23.13	**-16.94**	-23.46	-23.31
10-year	-41.02	-17.03	-16.49	-17.09	**-15.66**	-18.51
			Maximum Error			
3-month	36.29	11.32	9.15	9.58	**9.02**	9.17
6-month	17.93	22.18	22.03	22.19	15.79	**9.46**
1-year	21.98	22.00	**11.81**	12.01	19.90	11.90
2-year	18.08	18.89	17.32	18.91	18.91	19.20
5-year	19.47	9.32	8.84	9.37	**8.62**	8.66
10-year	20.54	16.97	**7.79**	7.86	7.93	8.32
			$\widehat{\rho}_1$			
3-month	0.748	0.654	0.477	0.505	0.385	0.346
6-month	0.483	0.242	0.270	0.283	0.241	0.245
1-year	0.609	0.510	0.393	0.389	0.399	0.416
2-year	0.688	0.607	0.617	0.603	0.618	0.614
5-year	0.769	0.692	0.644	0.638	0.601	0.608
10-year	0.668	0.551	0.483	0.402	0.389	0.398
			$\widehat{\rho}_{12}$			
3-month	0.305	0.010	0.139	0.146	0.100	0.122
6-month	0.266	0.154	0.216	0.236	0.241	0.230
1-year	0.343	0.339	0.358	0.349	0.354	0.360
2-year	0.058	0.127	0.109	0.090	0.061	0.080
5-year	0.136	-0.133	-0.099	-0.175	-0.181	-0.182
10-year	0.231	0.217	0.264	0.243	0.241	0.215

Notes: The table show in-sample fit error statistics for the full sample 1984:1-2003:12 (240 observations). The statistics are expressed in basis points. Results are shown for models with unrestricted decay parameter(s) [λ(s)]. Error statistics are given for the two, three and four-factor specification [NS2-2-λ, NS2-3-λ, NS-4-λ], the Bliss extension [NS2-B] and the adjusted Svensson model [NS2-(A)S]. The statistics $\widehat{\rho}_1$ and $\widehat{\rho}_{12}$ represent the 1st and 12th autocorrelation of the yield errors. For selected statistics, bold numbers indicate the best performing model.

parameter. For every maturity the Bliss model marginally improves the fit. However, the four-factor model with estimated λ_t is always more accurate than the Bliss model, showing that it is more beneficial to introduce the second slope factor than separate decay parameters.

In fact, the four-factor model with an estimated λ_t fits the term structure only marginally worse than the best fitting model with the maximum overall difference being no larger than 0.6 basis points (for a 3-month maturity based on both the root mean squared error and the mean absolute error). This means that, compared to the three-factor base model, it does not seem to make much difference whether a second curvature factor with a separate decay parameter is added (the Svensson model) or just a second slope factor which has the same λ_t as the first three factors. Comparing the Svensson model with its Adjusted alternative shows that the latter fits marginally better.

To summarize, the best fitting models in an absolute sense are indeed the models which allow for the most amount of flexibility which are the Svensson and Adjusted Svensson models. However, the four-factor model provides a fit which is nearly as accurate and has the benefit of being easier to estimate because the nonlinearities in the model are due to only one decay parameter instead of two. The interesting question now is whether the additional slope factor, in addition to improving the in-sample fit, can also help to improve the out-of-sample performance.

Before I turn to discussing the forecast results I first address the effect of imposing the restrictions on the decay parameters. Restricting these will most likely mean that some of the in-sample fit performance is sacrificed. The question is, however, to what extent this actually is the case. To assess the effect on in-sample fit I report in Table 5.4 the in-sample fit of those specifications that require estimating one or two decay parameters when no restrictions are imposed. Comparing Table 5.3 and Table 5.4 shows that in absolute terms the unrestricted models indeed fit the term structure more accurately. However, the differences are only substantial for the two-factor model which is explainable as with only two factors, having an additional parameter can make quite a difference. For example, for the 3-month maturity, the root mean squared error goes down from 18 to 12 basis points. For all other models, differences are, however, marginal with criteria such as standard deviation and mean absolute error being only 0.5 basis points worse for the restricted models. Furthermore, the bold numbers in each panel, which highlight the best fitting models per maturity, show almost negligible differences.

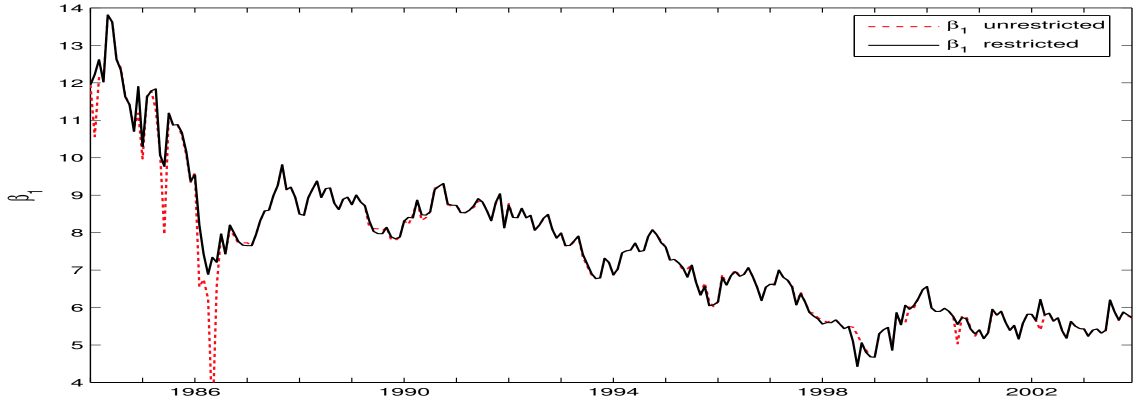
The results indicate that whether or not imposing restrictions does not matter much in terms of in-sample fit. Nevertheless, the reason why the restrictions are useful becomes apparent when we examine the time-series of the estimated factors. As these are modelled in the second step of the two-step procedure it is important that these series are relatively 'well-behaved'. That this not necessarily needs to be the case using the unrestricted estimation procedure is due to potential non-identification issues, as discussed earlier. Not imposing restrictions can result in extreme factor estimates. An example is given in Figure 5.6 in which the solid and dotted lines represent respectively the restricted and unrestricted estimates of the level, slope and curvature factors in the three-factor model. For most of the sample the restricted and unrestricted estimates are all but identical, except for a small number of months. For each of these months the unrestricted λ_t is substantially higher than the upper limit of 33.46. In particular for May 1986 this is the case with $\widehat{\lambda}_t$ equaling 65.61 as a result of which the level and slope factors are estimated at $\widehat{\beta}_{1,t} = 2.94$ and $\widehat{\beta}_{2,t} = 3.26$. Only the sum of these is somewhat close to the true level of the curve of 7.86% (using as proxy the 10-year yield) whereas with the restrictions in place the level estimate is $\widehat{\beta}_{1,t} = 7.34$.

5.6.2 Factor estimates

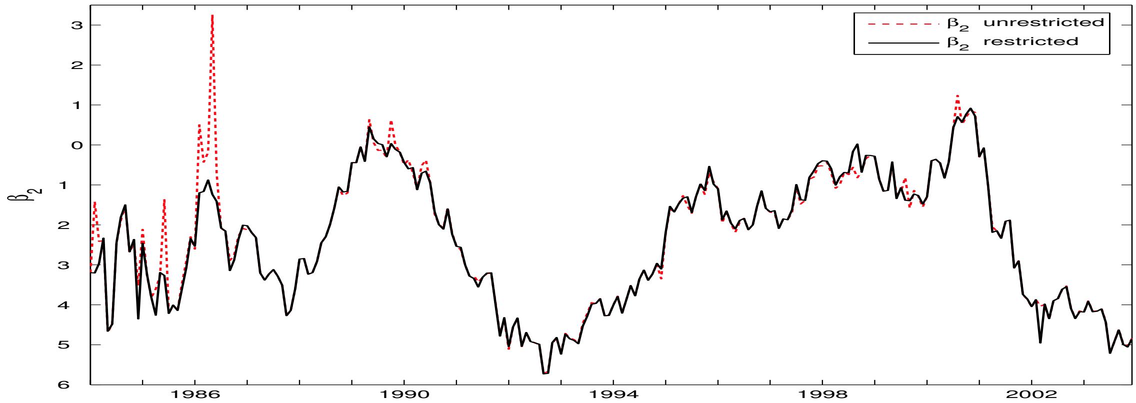
Time-series of the factor estimates, obtained with the two-step procedures are represented by the solid lines in Figures 5.7-5.9. Comparing the subgraphs within each row and across figures shows that the different models all give rather similar estimates for the level, slope and curvature factors. The estimates differ nevertheless in magnitude, mainly for the four-factor model. The time-series for the latter seems to suffer somewhat from outliers, in particular when the decay parameter is fixed (Figure 5.7) with some of the spikes in the slope and curvature factors disappearing when the decay parameter is estimated as well (Figure 5.8). Panels [h]-[k] of Figure 5.9 show that the two curvature factor estimates for the Adjusted Svensson model are more stable than those for the Svensson model. The latter still exhibit severe spikes, despite the restrictions on $\lambda_{1,t}$ and $\lambda_{2,t}$.

As an indication of the differences in the resulting factor estimates between the one and two-step estimation methods, Figures 5.8 and 5.9 also show the Kalman filter factor estimates for the full sample by means of the dotted lines. Whereas in general the time-series are quite close, there are certainly differences, mainly for the more complex models like the Svensson models. Using cross-sectional yield data as well as information concerning the evolution of yields over time smoothes out the factor estimates.

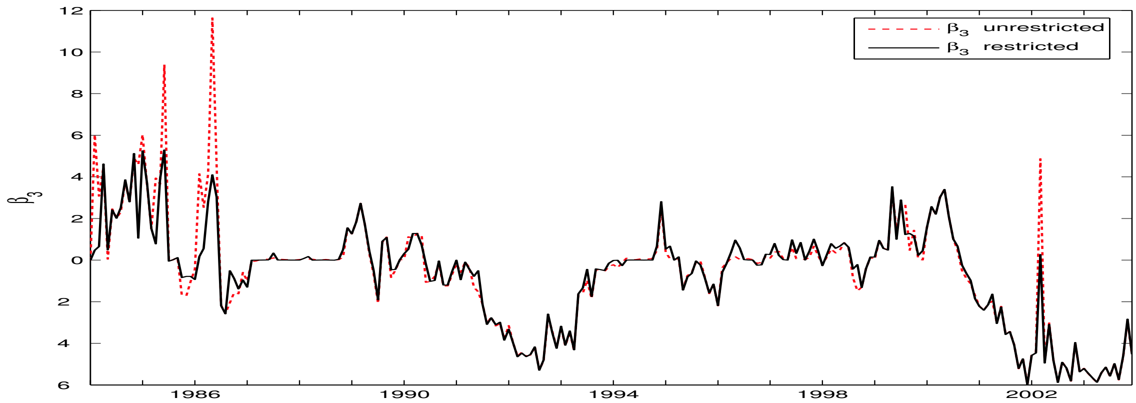
Figure 5.6: Nelson-Siegel factors with and without restrictions on λ_t



[a] β_1



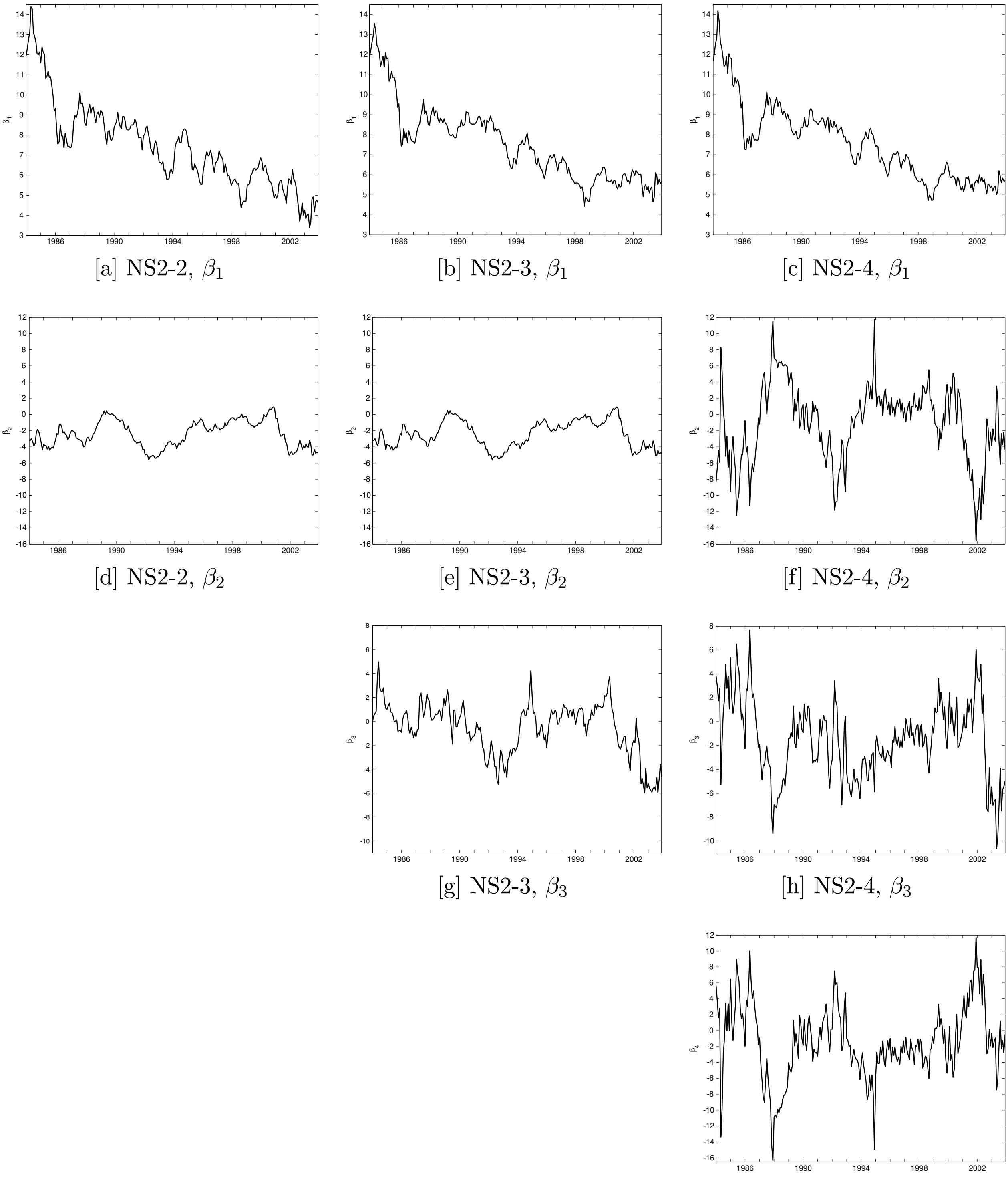
[b] β_2



[c] β_3

Notes: The graph shows time-series plots of the estimated Nelson-Siegel factors for the three-factor base model where λ_t is estimated alongside the factors $(\beta_{1,t}\ \beta_{2,t}\ \beta_{3,t})$. Shown are the estimate of the first factor in Panel [a], the second factor in Panel [b] and the third and last factor in Panel [c]. The solid line is the factor estimate when λ_t is restricted to the domain $[6.69, 33.46]$. The dotted line represented each factor when estimated without the restriction on λ. The sample period is 1984:1 - 2003:12 (240 observations).

Figure 5.7: Time-series of Nelson-Siegel factors with a fixed value for λ_t



[a] NS2-2, β_1 [b] NS2-3, β_1 [c] NS2-4, β_1

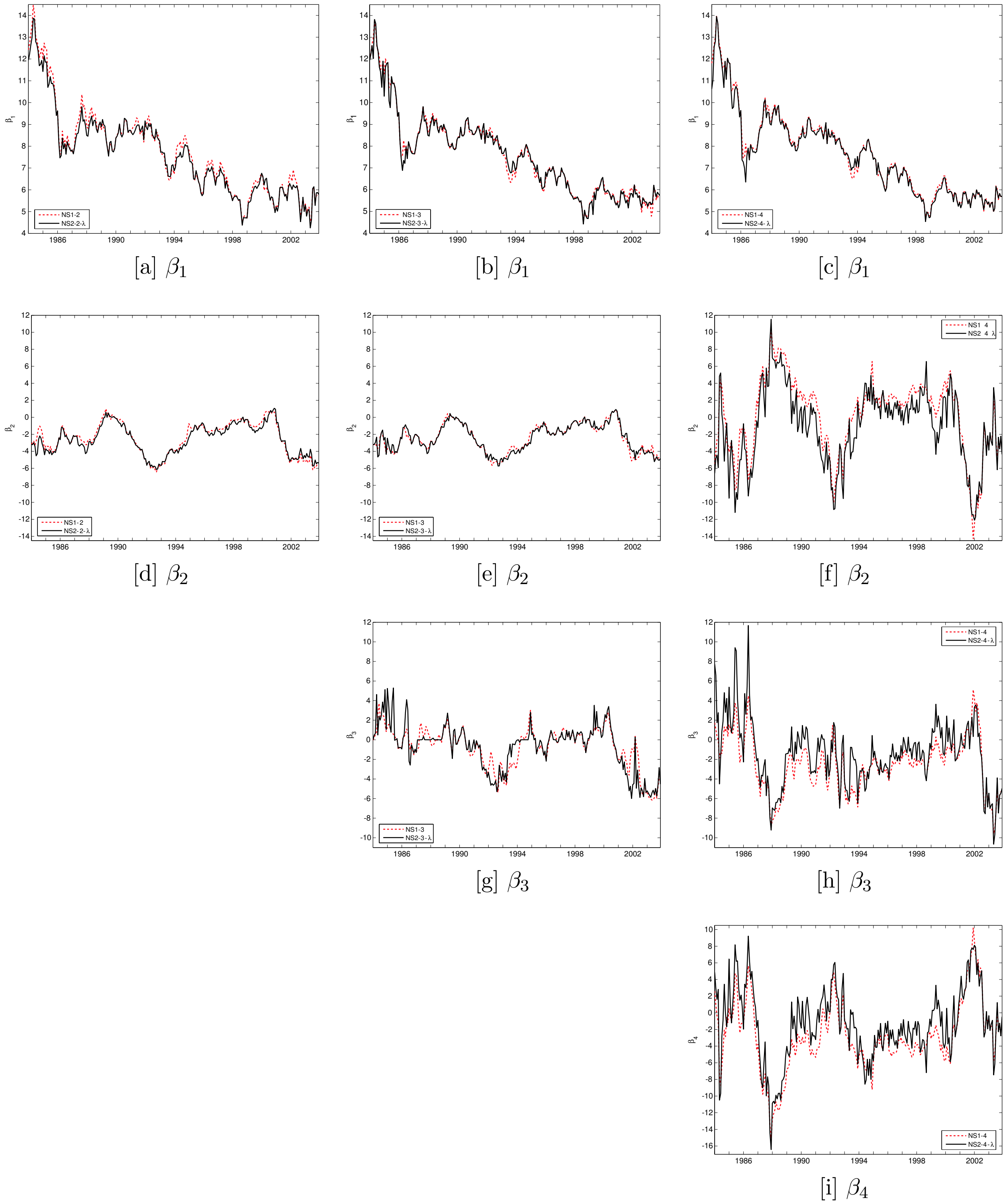
[d] NS2-2, β_2 [e] NS2-3, β_2 [f] NS2-4, β_2

[g] NS2-3, β_3 [h] NS2-4, β_3

[i] NS2-4, β_4

Notes: The graph shows time-series plots of the estimated Nelson-Siegel factors for the two-factor model (first column), the three-factor model (second column) and the four-factor model (fourth column). The factors are estimated using OLS given a fixed λ_t which is set to 16.42. The sample period is 1984:1 - 2003:12 (240 observations).

Figure 5.8: Time-series of Nelson-Siegel factors with estimated λ



[a] β_1 [b] β_1 [c] β_1

[d] β_2 [e] β_2 [f] β_2

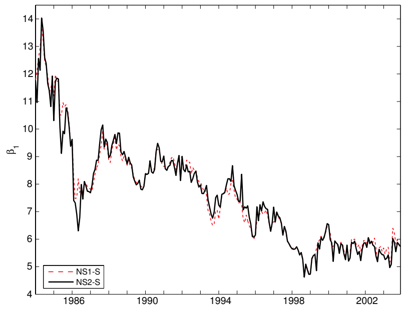
[g] β_3 [h] β_3

[i] β_4

Notes: The graph shows time-series plots of the estimated Nelson-Siegel factors for the two-factor model (first column), the three-factor model (second column) and the four-factor model (fourth column) where λ is estimated alongside (β_1 β_2 β_3 β_4). Shown are the factors estimates from the two-step NLS (solid lines) and the one-step Kalman Filter (dotted lines) estimation methods. The sample period is 1984:1 - 2003:12 (240 observations).

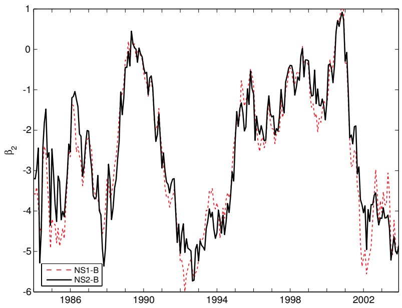
Figure 5.9: Time-series of Nelson-Siegel factors with estimated λs

[a] β_1

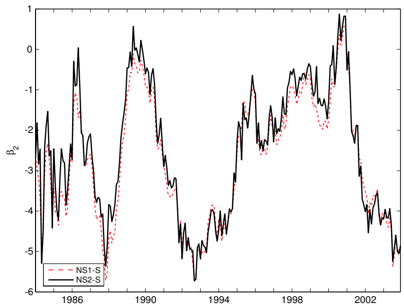


[b] β_1

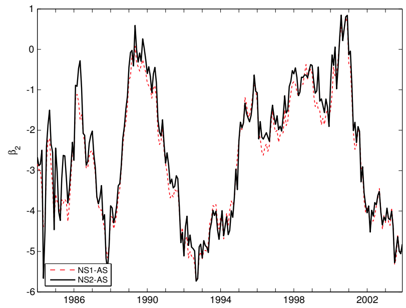
[c] β_1



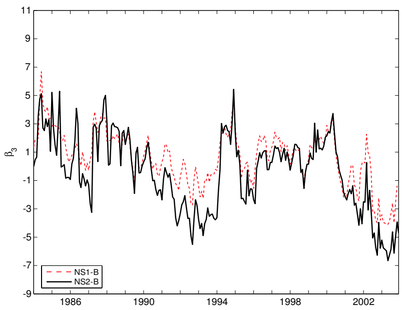
[d] β_2



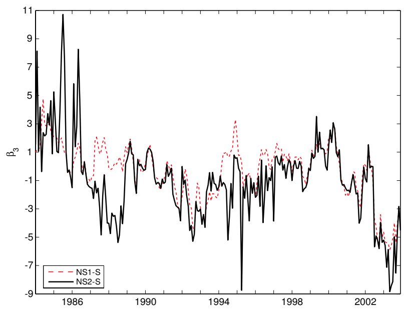
[e] β_2



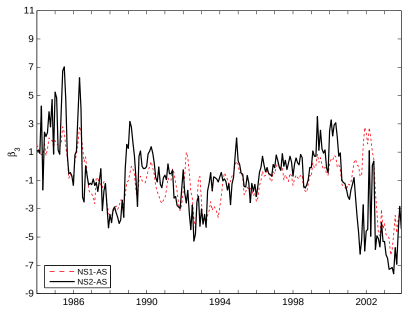
[f] β_2



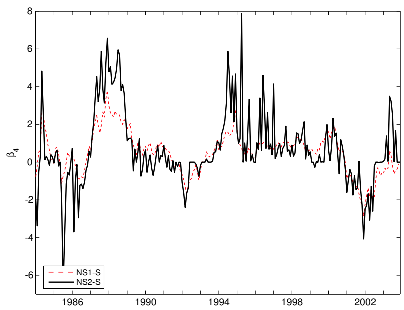
[g] β_3



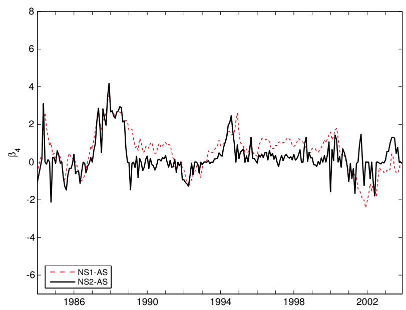
[h] β_3



[i] β_3



[j] β_4



[k] β_4

Notes: The graph shows time-series plots of the estimated Nelson-Siegel factors for the three-factor Bliss model (first column), the four-factor model Svensson extension (second column) and the four-factor model Adjusted Svensson model (fourth column) where λ_1 and λ_2 are estimated alongside (β_1 β_2 β_3 β_4). Shown are the factors resulting from the two-step NLS and the one-step Kalman Filter estimation methods. The sample period is 1984:1 - 2003:12 (240 observations).

Table 5.5: Factor summary statistics

		summary statistics				**correlations estimated factors**				**correlations yield factors**		
	factor	mean	stdev	ρ_1	ρ_{12}	β_1	β_2	β_3	β_4	L	S	C
NS2-2	β_1	7.403	2.115	0.970	0.571	1	-	-	-	*0.996*	-0.195	0.544
	β_2	-2.387	1.625	0.966	0.420	-0.144	1	-	-	-0.089	*0.990*	0.472
NS2-3	β_1	7.531	1.896	0.972	0.635	1	-	-	-	*0.973*	-0.324	0.353
	β_2	-2.394	1.639	0.967	0.426	-0.280	1	-	-	-0.079	*0.988*	0.485
	β_3	-0.571	2.171	0.923	0.376	0.347	0.510	1	-	0.539	0.415	*0.992*
NS2-4	β_1	7.599	1.911	0.971	0.620	1	-	-	-	*0.978*	-0.312	0.390
	β_2	-1.103	4.774	0.857	0.291	0.022	1	-	-	0.063	*0.409*	0.470
	β_3	-1.536	3.199	0.831	0.178	0.136	-0.615	1	-	0.264	0.204	*0.335*
	β_4	-1.426	4.591	0.816	0.202	-0.125	-0.942	0.763	1	-0.104	*-0.080*	-0.348
NS2-2-λ	β_1	7.610	1.944	0.969	0.616	1	-	-	-	*0.974*	-0.326	0.379
	β_2	-2.585	1.766	0.965	0.446	-0.253	1	-	-	-0.050	*0.983*	0.509
NS2-3-λ	β_1	7.534	1.861	0.962	0.632	1	-	-	-	*0.968*	-0.318	0.351
	β_2	-2.414	1.599	0.944	0.445	-0.280	1	-	-	-0.074	*0.981*	0.467
	β_3	-0.686	2.362	0.856	0.454	0.349	0.515	1	-	0.534	0.441	*0.861*
NS2-4-λ	β_1	7.586	1.871	0.969	0.613	1	-	-	-	*0.969*	-0.318	0.380
	β_2	-1.115	4.505	0.868	0.306	0.045	1	-	-	0.080	*0.431*	0.490
	β_3	-1.420	3.284	0.796	0.185	0.140	-0.558	1	-	0.304	0.199	*0.339*
	β_4	-1.412	4.348	0.828	0.197	-0.156	-0.933	0.724	1	-0.115	*-0.082*	-0.358
NS2-B	β_1	7.599	1.888	0.961	0.627	1	-	-	-	*0.967*	-0.326	0.363
	β_2	-2.519	1.649	0.937	0.405	-0.315	1	-	-	-0.103	*0.978*	0.413
	β_3	-0.412	2.685	0.880	0.352	0.394	0.349	1	-	0.571	0.307	*0.937*
NS2-S	β_1	7.596	1.864	0.955	0.605	1	-	-	-	*0.962*	-0.321	0.366
	β_2	-2.530	1.634	0.931	0.382	-0.294	1	-	-	-0.074	*0.966*	0.412
	β_3	-0.961	2.818	0.746	0.288	0.236	0.450	1	-	0.449	0.270	*0.586*
	β_4	0.679	1.968	0.721	0.199	0.172	-0.132	-0.474	1	0.113	0.042	*0.345*
NS2-AS	β_1	7.585	1.883	0.962	0.617	1	-	-	-	*0.963*	-0.321	0.358
	β_2	-2.515	1.643	0.937	0.381	-0.298	1	-	-	-0.083	*0.968*	0.409
	β_3	-0.895	2.593	0.805	0.329	0.284	0.500	1	-	0.487	0.372	*0.703*
	β_4	0.277	0.946	0.715	0.099	0.113	-0.221	-0.330	1	0.077	-0.073	*0.250*
yield factors	L	7.102	1.982	0.973	0.600	-	-	-	-	-	-	-
	S	-1.815	1.217	0.958	0.387	-	-	-	-	-	-	-
	C	-0.222	0.823	0.922	0.417	-	-	-	-	-	-	-

Notes: The table shows summary statistics of estimated factors for different Nelson-Siegel specifications. Statistics are shown for the two, three and four-factor model specification with λ_t fixed to 16.42 [NS2-2, NS2-3, NS-4], with λ_t estimated (but restricted) [NS-2-λ, NS-3-λ, NS-4-λ], the Bliss extension [NS2-B] and the adjusted Svensson model [NS2-(A)S]. Columns 1-4 represent the mean and standard deviation of the factors and their 1$^{\text{st}}$ and 12$^{\text{th}}$ order sample autocorrelation. Columns 5-8 show the correlation matrix of the factors within a given model whereas the final columns give the correlation of each factor with the empirical level (L, [10-year yield]), (negative of the) slope (S, -[10-year yield - 3-month yield]) and curvature (C, [2*2-year yield-10-year yield-3-month yield]). Statistics are calculated over the sample 1984:1 - 2003:12 (408 observations).

Table 5.5 presents detailed full-sample summary statistics for the two-step factor estimates. Statistics for empirical level, slope, curvature estimates, which have been constructed from the yields directly, are shown in the last rows of the table. The estimated factors mimic the empirical factors quite closely which is also clear from the italicized numbers in the last three columns showing the correlations between the estimated and empirical factors. All factors are highly autocorrelated and there is also substantial cross-correlation across factors[21]. The importance of accounting for this cross correlation from a forecasting perspective will be assessed by comparing the results between the AR and VAR specifications for the factor dynamics.

5.7 Out-of-sample forecasting results

For the out-of-sample performance I run a similar horse-race between the different models as for the in-sample fit. However, now there is no clear-cut conjecture how models will perform as there may be a trade-off between in-sample and out-of-sample performance. The models that provide a better in-sample do not necessary have to perform well out-of-sample because of the risk of overfitting. This will especially be the case when the models are estimated with the two-step procedure as the fitting process in the first step does not take into account the dynamics of the factors in the second step, the latter being crucial for the out-of-sample performance.

5.7.1 Forecast procedure

I assess the forecasting performance of the Nelson-Siegel models by dividing the full data sample into the initial estimation period 1984:1 - 1993:12 (120 observations) and the forecasting period 1994:1 - 2003:12 (120 observations). Next to gauging the models' predictive accuracy over the full sample I also consider two subsamples: 1994:1 - 2000:12 (84 observations) and 2001:1 - 2003:12 (36 observations). The first subsample is the out-of-sample period used by Diebold and Li (2006) and allows me to directly compare the performance of the alternative Nelson-Siegel specifications with that of the three-factor factor model results of Diebold and Li (2006). The second subsample starts in 2001 when the Federal

[21] Especially the two slope factors in the four-factor model are very strongly, negatively, correlated. Panels [f] and [i] of Figures 5.7 and 5.8 also indicate that to a certain extent the slope factors seem to offset each other, giving rise to a potential multicolinearity problem. However, the results in this and the following section show that adding the second slope factor helps to improve not only the in-sample fit but also the out-of-sample forecasting accuracy.

Reserve lowered the target rate from 6.5% to 6% in a first of eleven subsequent decreases, resulting in a drop of short-term interest rates by 4% and a strong widening of spreads. Mönch (2006a) and De Pooter *et al.* (2007) both show that predictability is scarce in 2001-2003 and it will be interesting to see how the Nelson-Siegel models perform in this period.

All the models are estimated recursively with an expanding data window. Interest rate forecasting is carried out by constructing factor predictions using the state equations and subsequently substituting these predictions in the measurement equations to obtain the interest rate forecasts. I consider four forecast horizons, $h = 1$ month as well as 3, 6 and 12 months ahead. The h-month ahead factor forecasts, $\widehat{\beta}_{T+h}$, are iterated forecasts which follow from forward iteration of the state equations in (5.18) as follows

$$\widehat{\beta}_{T+h} = \left[I_K - \widehat{\Phi}^h\right]\left[I_K - \widehat{\Phi}\right]^{-1}\widehat{\mu} + \widehat{\Phi}^h\beta_T$$

where I_K is the $(K \times K)$ identity matrix, $\widehat{\mu}$ and $\widehat{\Phi}$ the state equation estimates and β_T the last available factor estimates. With the one-step estimation method I use the in-sample decay parameter estimates to compute the factor loadings. With the two-step method I use the median value of the time-series of decay parameter estimates[22].

5.7.2 Competitor models

Random walk

I consider three competitor models against which to judge the predictive accuracy of the Nelson-Siegel models. The first is the benchmark Random Walk model

$$y_t(\tau_i) = y_{t-1}(\tau_i) + \varepsilon_t(\tau_i), \qquad \varepsilon_t(\tau_i) \sim \mathcal{N}\left(0, \sigma^2(\tau_i)\right) \tag{5.23}$$

Many other studies that consider interest rate forecasting all show that consistently outperforming the random walk is difficult. The Random Walk h-month ahead forecast is equal to the last observed value; $\widehat{y}_{T+h}(\tau_i) = y_T(\tau_i)$.

[22]Experimentation with alternative choices (using the most recent decay parameter estimate and using the mean estimate) revealed that using the median gives more stable results. Note that Nelson and Siegel (1987) who estimate λ_t alongside the factors in the three-factor model also report fit results when imposing the median λ_t estimate. They find that the in-sample fit is not degraded much when doing so.

AR(1) model

The second competitor model is a first-order univariate autoregressive model which allows for mean-reversion

$$y_t(\tau_i) = \mu(\tau_i) + \phi(\tau_i) y_{t-1}(\tau_i) + \varepsilon_t(\tau_i), \qquad \varepsilon_t(\tau_i) \sim \mathcal{N}\left(0, \sigma^2(\tau_i)\right) \tag{5.24}$$

VAR(1) model

The third and final competitor model is an unrestricted VAR(1) model for yield levels. A well-known shortcoming of using VAR models for yield forecasting is that only maturities that are included in the model can be forecasted. To keep down the number of parameters I therefore estimate a VAR(1) model in which the lagged yields are replaced by their first three common factors. The reason is that these factors explain over 99% of the total variation and also because of their Litterman and Scheinkman (1991) interpretation as level, slope and curvature factors. I extract the factor matrix, denoted by F_{t-1}, by applying static principal component analysis on the panel of lagged yields (using data up until month $t-1$) which consist of 13 maturities: $\tau = 1$, 3 and 6 months and 1, 2, ..., 10 years[23]. The VAR(1) model is then given by

$$Y_t = \mu + \Phi F_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \Sigma) \tag{5.25}$$

with $Y_t = [y_t^{(1m)}, ..., y_t^{(10y)}]'$, μ a (13×1) vector, Φ a (13×3) matrix and Σ a full (13×13) matrix.

One important difference between the Nelson-Siegel and VAR models is that the latter does not impose a specific parametric form on the right hand side of the measurement equations. The VAR(1) model can therefore be used to determine whether the exponential factor loading structure of the Nelson-Siegel class of model is beneficial for forecasting yields. For the AR and VAR model I similarly construct iterated forecasts.

5.7.3 Forecast evaluation

I use a number of standard forecast error evaluation criteria to assess the quality of the out-of-sample forecasts. In particular, I report the Root Mean Squared Prediction Error (RMSPE) for the individual maturities as well as the Trace Root Mean Squared Prediction Error (TRMSPE). The latter combines the forecast errors of all maturities and summarizes

[23] Note that similar to Diebold and Li (2006), I do not use the 1-month maturity in the Nelson-Siegel models. I do include it here in order to also assess the forecasts for this short maturity.

the performance per model, thereby allowing for a direct comparison between models[24]. Significant differences between the forecast performance of the random walk and each of the models are tested for using the White (2000) 'reality check' test which I implement using the stationary bootstrap method of Politis and Romano (1994) with 1000 block-bootstraps of the forecast error series and an average block-length of 12 months.

5.7.4 Forecast results

The results for the full sample period 1994:1 - 2003:12 are presented in Tables 5.6-5.9. The first line in each table show the (T)RMSPEs for the random walk. All other entries are relative (T)RMSPEs with respect to the random walk, including those in lines two and three which show the results for the competitor AR(1) and VAR(1) models. Bold numbers indicate outperformance with respect to the random walk. The results for this sample are directly comparable to those of Mönch (2006a) as he uses an almost identical forecasting sample (1994:1 - 2003:9) and also reports results for the NS2-3-AR and NS2-3-VAR specifications. Although the random walk statistics are all but identical, for the Nelson-Siegel three factor model I find somewhat better statistics for longer forecast horizons than Mönch. This is most likely caused by my use of iterated forecasts whereas Mönch uses direct forecasts, the different sets of maturities used and by the small differences in estimation and forecast periods.

For a 1-month horizon there only seems to be a certain degree of predictability for maturities of less than one year. With the two-step estimation method it is mainly the four-factor model that does well for the one and three-month maturities. The 19% outperformance for the one-month maturity relative to the random walk is significant at the 5% level according to the White reality check test. A VAR(1) specification for the factor dynamics clearly works better than separate AR(1) models per factor. For the one-step models, it is in terms of TRMSPE less clear which specification for the factor dynamics is to be preferred. However, allowing for a full coefficient matrix clearly produces the most accurate short-maturity forecasts, although the best model is NS1-4-AR with a TRMSPE of 0.98.

[24]Results of other evaluation criteria such as the Mean Prediction Error (MPE), Mean Absolute Prediction Error (MAPE) and forecast regression R^2-s are not reported here but are available upon request. For more details regarding the TRMSPE, see Christoffersen and Diebold (1998). I compute the TRMSPE over the following maturities for which I compute forecasts: $\tau = 1$, 3, 6 and 12 months and 2,...,10 years. Tables 5.6-5.13 show results for individual maturities for only a subset of these thirteen maturities.

Table 5.6: Forecast results for the sample 1994:1 - 2003:12, 1-month horizon

	TRMSPE	**RMSPE**							
Maturity	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y
RW	101.59	30.12	21.18	21.82	25.71	29.12	30.48	29.30	27.95
AR	1.01	1.01	1.01	1.01	1.01	1.01	1.00	1.00	1.00
VAR	1.01	**0.81**	**0.95**^{10}	**0.99**	1.03	1.07	1.01	1.02	1.07
Panel A: two-step models									
NS2-2-AR	1.18	1.55	1.44	1.05	1.07	1.27	1.04	1.05	1.26
NS2-2-VAR	1.19	1.62	1.55	1.12	1.05	1.24	1.03	1.07	1.28
NS2-3-AR	1.02	**0.97**	**0.98**	1.07	1.07	1.07	1.02	1.01	1.03
NS2-3-VAR	1.01	**0.95**^{5}	**0.87**^{5}	**0.97**^{10}	1.02	1.06	1.03	1.03	1.03
NS2-4-AR	1.07	**0.97**	1.26	1.23	1.18	1.13	1.04	1.03	1.01
NS2-4-VAR	**0.99**	**0.81**^{5}	**0.85**	**0.94**	**0.98**	1.06	1.02	1.03	1.03
NS2-2-λ-AR	1.68	1.00	1.01	1.38	1.88	2.17	1.85	1.58	1.45
NS2-2-λ-VAR	1.66	1.05	1.05	1.34	1.83	2.12	1.82	1.57	1.44
NS2-3-λ-AR	1.23	1.01	1.26	1.41	1.36	1.33	1.25	1.19	1.15
NS2-3-λ-VAR	1.21	**0.87**^{10}	1.05	1.26	1.30	1.33	1.26	1.21	1.19
NS2-4-λ-AR	1.10	**0.97**	1.26	1.26	1.20	1.18	1.09	1.06	1.06
NS2-4-λ-VAR	1.00	**0.83**^{5}	**0.90**	**0.97**	1.00	1.08	1.03	1.02	1.06
NS2-B-AR	1.18	1.04	1.30	1.36	1.23	1.16	1.22	1.20	1.19
NS2-B-VAR	1.14	**0.90**	1.06	1.16	1.12	1.12	1.19	1.20	1.21
NS2-S-AR	1.30	1.04	1.36	1.53	1.47	1.40	1.32	1.27	1.30
NS2-S-VAR	1.11	**0.84**^{5}	**0.98**	1.12	1.12	1.13	1.14	1.16	1.22
NS2-AS-AR	1.29	1.08	1.42	1.52	1.45	1.40	1.31	1.25	1.20
NS2-AS-VAR	1.24	**0.92**	1.18	1.30	1.30	1.33	1.28	1.25	1.22
Panel B: one-step models									
NS1-2-RW	1.29	2.22	2.31	1.67	1.06	1.05	1.00	1.02	1.09
NS1-2-AR	1.30	2.23	2.31	1.68	1.09	1.09	1.04	1.03	1.11
NS1-2-VAR	1.32	2.25	2.35	1.72	1.10	1.08	1.05	1.04	1.13
NS1-3-RW	1.01	1.07	1.00	**0.99**	1.01	1.04	1.01	1.01	1.01
NS1-3-AR	1.02	1.09	1.04	1.05	1.06	1.07	1.02	1.01	1.02
NS1-3-VAR	1.04	1.07	**0.97**^{5}	**0.98**^{5}	1.01	1.07	1.07	1.05	1.07
NS1-4-RW	1.00	**0.93**	1.01	1.01	**0.97**	1.04	1.01	1.01	1.02
NS1-4-AR	**0.98**	**0.89**	**0.97**	**0.97**	**0.96**	1.03	1.00	1.00	1.02
NS1-4-VAR	**0.99**	**0.82**^{10}	**0.86**	**0.90**^{10}	**0.94**^{10}	1.04	1.02	1.03	1.08
NS1-B-RW	1.01	1.05	1.05	1.01	**0.98**	1.03	1.01	1.01	1.04
NS1-B-AR	1.02	1.05	1.07	1.04	1.01	1.05	1.02	1.01	1.04
NS1-B-VAR	1.02	**0.92**^{5}	**0.88**^{5}	**0.96**^{5}	1.00	1.06	1.06	1.05	1.10
NS1-S-RW	1.01	1.01	1.06	1.01	**0.97**	1.04	1.01	1.02	1.01
NS1-S-AR	1.02	1.02	1.10	1.04	1.00	1.07	1.02	1.02	1.04
NS1-S-VAR	**0.99**	**0.84**	**0.87**	**0.90**^{10}	**0.94**^{10}	1.03	1.02	1.03	1.07
NS1-AS-RW	1.00	1.00	1.06	1.01	**0.97**	1.04	1.01	1.02	1.01
NS1-AS-AR	1.02	1.01	1.09	1.03	**0.99**	1.06	1.02	1.01	1.03
NS1-AS-VAR	**0.99**	**0.84**	**0.87**	**0.90**^{10}	**0.94**^{10}	1.03	1.02	1.03	1.07

Notes: Bold numbers indicate outperformance relative to the random walk (RW) whereas ()^{10}, ()^{5} and ()^{1} 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 5.7: Forecast results for the sample 1994:1 - 2003:12, 3-month horizon

	TRMSPE	**RMSPE**							
Maturity	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y
RW	197.03	54.05	48.59	51.06	55.68	60.20	57.56	53.78	50.08
AR	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
VAR	1.01	**0.74**^{10}	**0.89**	**0.98**	1.05	1.07	1.03	1.03	1.08
Panel A: two-step models									
NS2-2-AR	1.03	**0.90**^{1}	**0.89**^{1}	**0.97**^{5}	1.05	1.10	1.02	1.03	1.13
NS2-2-VAR	1.04	**0.97**^{5}	**0.96**^{5}	**0.99**^{10}	1.03	1.08	1.03	1.05	1.16
NS2-3-AR	1.01	**0.93**	1.01	1.07	1.08	1.06	1.02	1.00	1.00
NS2-3-VAR	1.00	**0.80**^{1}	**0.85**^{1}	**0.96**^{10}	1.03	1.06	1.04	1.03	1.03
NS2-4-AR	1.08	1.05	1.22	1.20	1.18	1.13	1.04	1.01	1.00
NS2-4-VAR	**0.99**	**0.72**^{1}	**0.81**^{5}	**0.92**^{10}	**0.99**	1.04	1.03	1.03	1.05
NS2-2-λ-AR	1.20	**0.82**^{1}	**0.95**	1.13	1.29	1.38	1.27	1.17	1.15
NS2-2-λ-VAR	1.18	**0.86**^{1}	**0.96**^{10}	1.11	1.25	1.33	1.25	1.17	1.16
NS2-3-λ-AR	1.07	1.04	1.17	1.22	1.20	1.13	1.05	1.01	1.02
NS2-3-λ-VAR	**0.99**	**0.82**^{1}	**0.93**	1.03	1.06	1.03	**0.99**	1.00	1.05
NS2-4-λ-AR	1.08	1.01	1.16	1.17	1.17	1.14	1.06	1.03	1.02
NS2-4-λ-VAR	**0.95**	**0.71**^{5}	**0.79**	**0.89**	**0.96**	1.01	**0.99**	**0.99**	1.04
NS2-B-AR	1.11	1.06	1.19	1.23	1.19	1.13	1.11	1.09	1.10
NS2-B-VAR	1.03	**0.82**^{10}	**0.91**	1.00	1.03	1.03	1.06	1.08	1.14
NS2-S-AR	1.27	1.14	1.32	1.41	1.43	1.37	1.25	1.20	1.22
NS2-S-VAR	1.00	**0.77**^{1}	**0.85**^{10}	**0.95**	**0.99**	1.02	1.02	1.05	1.14
NS2-AS-AR	1.16	1.11	1.26	1.31	1.29	1.22	1.13	1.08	1.08
NS2-AS-VAR	1.00	**0.84**	**0.94**	1.01	1.04	1.03	1.01	1.02	1.08
Panel B: one-step models									
NS1-2-RW	1.03	1.20	1.11	1.04	**0.99**	1.02	1.00	1.01	1.06
NS1-2-AR	1.07	1.25	1.15	1.10	1.04	1.06	1.04	1.04	1.10
NS1-2-VAR	1.09	1.26	1.18	1.11	1.04	1.06	1.06	1.06	1.13
NS1-3-RW	**0.99**	**0.91**^{5}	**0.94**^{10}	**0.99**	1.00	1.02	1.01	1.00	1.00
NS1-3-AR	1.00	**0.95**	**0.99**	1.05	1.06	1.05	1.01	**0.99**	**0.99**
NS1-3-VAR	1.05	**0.85**^{1}	**0.89**^{1}	**0.99**^{5}	1.04	1.09	1.11	1.10	1.12
NS1-4-RW	1.00	**0.94**	**0.99**	1.00	**0.99**	1.02	1.01	1.01	1.01
NS1-4-AR	**0.96**	**0.81**	**0.88**	**0.93**	**0.95**	**0.99**	**0.99**	**0.98**	**0.99**
NS1-4-VAR	**0.99**	**0.68**^{1}	**0.76**^{5}	**0.87**^{5}	**0.94**^{10}	1.02	1.05	1.06	1.11
NS1-B-RW	1.00	**0.98**	1.00	1.00	**0.99**	1.02	1.00	1.00	1.02
NS1-B-AR	1.01	**0.99**	1.03	1.05	1.03	1.04	1.00	1.00	1.02
NS1-B-VAR	1.04	**0.78**^{1}	**0.84**^{1}	**0.96**^{5}	1.03	1.09	1.10	1.10	1.14
NS1-S-RW	1.00	**0.99**	1.01	1.00	**0.99**	1.02	1.01	1.01	1.00
NS1-S-AR	1.01	1.01	1.04	1.04	1.02	1.05	1.02	1.00	1.01
NS1-S-VAR	**0.99**	**0.69**^{5}	**0.77**^{5}	**0.87**^{5}	**0.94**^{10}	1.02	1.05	1.06	1.10
NS1-AS-RW	1.00	**0.99**	1.01	1.00	**0.99**	1.02	1.01	1.01	1.00
NS1-AS-AR	1.01	1.00	1.04	1.03	1.01	1.04	1.01	**0.99**	1.00
NS1-AS-VAR	**0.99**	**0.70**	**0.77**	**0.87**	**0.94**	1.02	1.05	1.06	1.11

Notes: Bold numbers indicate outperformance relative to the random walk (RW) whereas $()^{10}$, $()^{5}$ and $()^{1}$ 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 5.8: Forecast results for the sample 1994:1 - 2003:12, 6-month horizon

	TRMSPE	**RMSPE**							
Maturity	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y
RW	295.47	84.53	82.84	84.95	87.81	90.76	84.03	77.10	70.68
AR	1.01	1.02	1.00	1.00	1.01	1.01	1.01	1.01	1.02
VAR	1.12	**0.92**	1.03	1.10	1.17	1.16	1.11	1.12	1.19
Panel A: two-step models									
NS2-2-AR	1.02	**0.89**^{1}	**0.92**^{1}	1.00	1.05	1.06	1.03	1.03	1.11
NS2-2-VAR	1.03	**0.91**^{5}	**0.93**^{5}	**0.99**	1.03	1.05	1.05	1.06	1.15
NS2-3-AR	1.02	**0.99**	1.02	1.07	1.08	1.05	1.01	**0.99**	1.01
NS2-3-VAR	1.02	**0.86**^{1}	**0.89**^{1}	**0.98**^{10}	1.05	1.07	1.05	1.04	1.06
NS2-4-AR	1.04	1.02	1.09	1.10	1.13	1.09	1.02	**0.99**	1.00
NS2-4-VAR	**0.99**	**0.78**^{1}	**0.84**^{5}	**0.93**^{10}	1.00	1.04	1.04	1.04	1.08
NS2-2-λ-AR	1.09	**0.92**^{1}	**0.99**	1.08	1.16	1.18	1.12	1.08	1.10
NS2-2-λ-VAR	1.08	**0.93**^{1}	**0.98**	1.06	1.13	1.15	1.11	1.09	1.12
NS2-3-λ-AR	1.09	1.06	1.12	1.17	1.18	1.13	1.06	1.04	1.07
NS2-3-λ-VAR	**0.97**	**0.85**^{1}	**0.92**	**0.99**	1.02	**0.99**	**0.96**	**0.99**	1.08
NS2-4-λ-AR	1.05	**0.98**	1.05	1.08	1.12	1.10	1.05	1.03	1.06
NS2-4-λ-VAR	**0.96**	**0.75**^{1}	**0.81**^{10}	**0.89**	**0.95**	**0.98**	**0.99**	1.02	1.09
NS2-B-AR	1.13	1.07	1.13	1.17	1.19	1.15	1.13	1.12	1.16
NS2-B-VAR	1.02	**0.82**^{10}	**0.89**	**0.97**	1.02	1.02	1.05	1.09	1.18
NS2-S-AR	1.27	1.14	1.24	1.32	1.38	1.34	1.26	1.23	1.28
NS2-S-VAR	1.02	**0.80**^{1}	**0.86**^{10}	**0.95**	1.01	1.02	1.04	1.09	1.20
NS2-AS-AR	1.15	1.10	1.18	1.24	1.27	1.21	1.12	1.09	1.12
NS2-AS-VAR	**0.97**	**0.85**	**0.91**	**0.97**	1.00	**0.98**	**0.96**	**0.99**	1.09
Panel B: one-step models									
NS1-2-RW	1.00	**0.97**^{1}	**0.95**^{1}	**0.98**^{1}	1.00	1.00	**0.99**	1.01	1.06
NS1-2-AR	1.06	1.07	1.04	1.07	1.08	1.07	1.05	1.05	1.12
NS1-2-VAR	1.08	1.06	1.02	1.06	1.06	1.07	1.08	1.10	1.19
NS1-3-RW	1.00	**0.96**^{5}	**0.96**^{5}	**0.99**^{10}	1.01	1.01	1.01	1.00	1.01
NS1-3-AR	1.00	1.00	1.01	1.05	1.06	1.03	**0.98**	**0.97**	**0.98**
NS1-3-VAR	1.10	**0.92**^{1}	**0.94**^{1}	1.03	1.09	1.13	1.16	1.16	1.21
NS1-4-RW	1.00	**0.99**	**0.99**	1.00	1.00	1.01	1.01	1.00	1.02
NS1-4-AR	**0.93**	**0.83**^{10}	**0.86**	**0.91**	**0.93**	**0.95**	**0.95**	**0.95**	**0.98**
NS1-4-VAR	1.01	**0.73**^{1}	**0.79**^{1}	**0.88**^{5}	**0.97**	1.03	1.08	1.10	1.18
NS1-B-RW	1.00	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.03
NS1-B-AR	1.01	1.03	1.03	1.05	1.05	1.03	**0.99**	**0.99**	1.02
NS1-B-VAR	1.10	**0.86**^{1}	**0.90**^{1}	1.01	1.09	1.14	1.16	1.17	1.24
NS1-S-RW	1.01	1.02	1.00	1.00	1.00	1.01	1.01	1.01	1.01
NS1-S-AR	1.00	1.04	1.03	1.03	1.03	1.02	**0.98**	**0.97**	**0.99**
NS1-S-VAR	1.01	**0.74**^{1}	**0.79**^{1}	**0.88**^{5}	**0.96**	1.03	1.08	1.10	1.18
NS1-AS-RW	1.00	1.02	1.00	1.00	1.00	1.01	1.01	1.01	1.01
NS1-AS-AR	**0.99**	1.03	1.02	1.02	1.01	1.01	**0.97**	**0.96**	**0.99**
NS1-AS-VAR	1.01	**0.74**^{1}	**0.79**^{1}	**0.88**^{5}	**0.96**	1.03	1.08	1.10	1.18

Notes: Bold numbers indicate outperformance relative to the random walk (RW) whereas $()^{10}$, $()^{5}$ and $()^{1}$ 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 5.9: Forecast results for the sample 1994:1 - 2003:12, 12-month horizon

	TRMSPE	**RMSPE**							
Maturity	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y
RW	435.08	138.26	140.53	142.35	140.73	133.83	114.30	103.36	94.12
AR	1.01	1.00	**0.99**	**0.99**	**0.99**	1.01	1.03	1.03	1.05
VAR	1.37	1.17	1.23	1.26	1.32	1.36	1.41	1.47	1.60
Panel A: two-step models									
NS2-2-AR	1.04	**0.99**^{5}	**0.99**	1.02	1.05	1.06	1.06	1.06	1.11
NS2-2-VAR	1.04	**0.97**^{5}	**0.97**^{5}	**0.99**	1.02	1.04	1.07	1.09	1.15
NS2-3-AR	1.03	1.01	1.01	1.03	1.05	1.06	1.04	1.02	1.04
NS2-3-VAR	1.05	**0.96**^{5}	**0.97**	1.01	1.05	1.08	1.08	1.07	1.10
NS2-4-AR	1.00	**0.88**	**0.92**	**0.95**	1.01	1.04	1.03	1.02	1.06
NS2-4-VAR	1.02	**0.89**^{1}	**0.92**^{5}	**0.97**	1.01	1.04	1.05	1.07	1.13
NS2-2-λ-AR	1.08	**0.99**	1.00	1.03	1.07	1.11	1.11	1.10	1.14
NS2-2-λ-VAR	1.06	**0.99**^{5}	**0.99**	1.02	1.05	1.07	1.10	1.10	1.15
NS2-3-λ-AR	1.09	1.01	1.02	1.05	1.09	1.11	1.12	1.12	1.17
NS2-3-λ-VAR	**0.96**	**0.91**^{10}	**0.93**	**0.95**	**0.97**	**0.95**	**0.94**^{5}	**0.98**^{5}	1.08
NS2-4-λ-AR	1.03	**0.86**	**0.90**	**0.94**	1.00	1.05	1.09	1.10	1.17
NS2-4-λ-VAR	**0.98**	**0.85**^{1}	**0.88**^{5}	**0.92**	**0.96**	**0.99**	1.02	1.06	1.16
NS2-B-AR	1.14	1.00	1.03	1.06	1.11	1.15	1.20	1.21	1.29
NS2-B-VAR	1.05	**0.90**	**0.94**	**0.97**	1.02	1.05	1.10	1.14	1.25
NS2-S-AR	1.24	1.06	1.10	1.15	1.22	1.26	1.30	1.32	1.41
NS2-S-VAR	1.04	**0.90**^{5}	**0.92**	**0.97**	1.01	1.03	1.08	1.13	1.25
NS2-AS-AR	1.13	1.02	1.05	1.09	1.13	1.15	1.16	1.16	1.22
NS2-AS-VAR	**0.97**	**0.90**	**0.92**	**0.95**	**0.98**	**0.97**	**0.97**^{10}	1.01	1.12
Panel B: one-step models									
NS1-2-RW	1.00	1.01	**0.99**	1.01	1.01	1.00	**0.99**	1.01	1.06
NS1-2-AR	1.10	1.12	1.09	1.10	1.10	1.10	1.08	1.09	1.17
NS1-2-VAR	1.11	1.08	1.05	1.07	1.07	1.08	1.13	1.16	1.27
NS1-3-RW	1.00	1.00	**0.99**	1.00	1.00	1.01	1.00	1.00	1.01
NS1-3-AR	1.00	1.02	1.01	1.02	1.03	1.02	**0.98**	**0.96**	**0.98**
NS1-3-VAR	1.16	1.03	1.03	1.07	1.12	1.17	1.22	1.24	1.32
NS1-4-RW	1.00	1.02	1.00	1.00	1.00	1.01	1.00	1.00	1.02
NS1-4-AR	**0.90**	**0.85**^{5}	**0.86**^{5}	**0.88**^{5}	**0.90**^{10}	**0.90**^{10}	**0.91**^{5}	**0.92**^{5}	**0.98**
NS1-4-VAR	1.06	**0.86**^{1}	**0.88**^{1}	**0.94**^{5}	1.00	1.06	1.14	1.19	1.31
NS1-B-RW	1.00	1.03	1.00	1.00	1.00	1.01	**0.99**	1.00	1.03
NS1-B-AR	1.02	1.03	1.02	1.03	1.04	1.04	1.00	**0.99**	1.02
NS1-B-VAR	1.18	1.00	1.02	1.08	1.14	1.21	1.26	1.28	1.37
NS1-S-RW	1.01	1.03	1.00	1.00	1.00	1.01	1.01	1.00	1.01
NS1-S-AR	**0.98**	1.02	1.00	**0.99**	1.00	**0.99**	**0.95**	**0.94**	**0.98**
NS1-S-VAR	1.05	**0.85**^{1}	**0.87**^{1}	**0.93**^{5}	**0.99**	1.05	1.13	1.18	1.30
NS1-AS-RW	1.01	1.03	1.00	1.00	1.00	1.01	1.00	1.00	1.01
NS1-AS-AR	**0.96**	1.01	**0.99**	**0.98**	**0.98**	**0.97**	**0.94**	**0.93**^{10}	**0.97**
NS1-AS-VAR	1.05	**0.85**^{1}	**0.87**^{1}	**0.93**^{5}	**0.99**	1.05	1.13	1.18	1.30

Notes: Bold numbers indicate outperformance relative to the random walk (RW) whereas $()^{10}$, $()^{5}$ and $()^{1}$ 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 5.7 shows for the 3-month horizon that for all models the best results are obtained for short maturities using a VAR specification for the factors, instead of using an AR or RW specification. The second and third line of the table shows this to hold also for the VAR model using yield levels directly. From Panel B is it clear that the one-step estimation procedure yields the most accurate results. It is interesting that the four-factor and (Adjusted) Svensson models not only fit the term structure very well, but that they also produce accurate short-maturity forecasts. For the one, three, six and twelve-month maturity they all outperform the random walk by 30%, 23%, 13% and 6% respectively. The four-factor model yields the most accurate results with the one-step procedure. The four-factor model is also the most accurate using the two-step procedure, with NS2-4-λ-VAR doing marginally better than NS2-4-VAR due to being more accurate for long maturities. The TRMSPEs of NS2-4-λ-VAR and NS1-4-AR are very similar but the latter model outperforms the random walk also for long maturities. In fact, from Panel B it seems that whereas the VAR specification works well for short maturities, the AR specification in general produces better forecasts for long maturities.

This pattern becomes more evident for the 6-month horizon. For the one-step procedure the VAR specification outperforms the random walk for short maturities, with the outperformance being strongly statistically significant, but for long maturities the performance is poor whereas the exact opposite pattern is visible for the AR specification. The NS1-4-AR again is the only model that forecasts well across the entire maturity spectrum, clearly giving it the lowest TRMSPE of 0.93. The closest competitor with the two-step procedure is still the NS2-4-λ-VAR model although the NS2-3-λ-VAR is a close second with a relative TRMSPE with is only 1% lower.

Finally, for the 12-month horizon, shown in Table 5.9, all the two-step VAR specifications outperform the random walk by up to 15% for short maturities but at the same time produce very poor forecasts for long maturities. Only the NS2-3-λ-VAR model is more accurate than the random walk for all maturities except the 10-year maturity. The VAR model for yield levels has an increasingly worse performance for longer maturities and the relative RMSPEs I report in the third row of the table are higher than those given in Mönch (2006a). This is most likely due to my use of a larger set of maturities and because I construct iterated forecasts. The Svensson and Adjusted Svensson models are again able to forecast both short and long maturities with the one-step estimation procedure, although not consistently with either AR or VAR factor dynamics. The NS1-4-AR model is still the most accurate model with a relative TRMSPE of 0.90 and significant outperformance for individual maturities up to ten years.

Overall the full-sample results can be summarized as follows. With both the one-step and two-step estimation procedures, using VAR factor dynamics is typically optimal for constructing short-maturity forecasts, irrespective of the forecast horizon. It is, however, only the one-step estimation procedure that also produces increasingly accurate forecasts when the forecast horizon lengthens, more specifically with the assumption of AR factor dynamics. With the two-step procedure such an improvement is lacking. This strongly suggests to simultaneously use cross-sectional and time-series information when the purpose of using the Nelson-Siegel model is that of forecasting the term structure. The best overall performing model is the four-factor model which is the only model that accurately forecasts the entire maturity spectrum, especially for the 6 and 12-month horizons. It is interesting that adding a second slope factor not only improves the in-sample fit but also the out-of-sample performance. In fact, adding factors in general seems to benefit the forecasting performance as the (Adjusted) Svensson model also predicts reasonably well compared to for the example the three-factor model. Whether freeing up decay parameters is helpful is somewhat ambiguous for the three and four-factor model. The Bliss model, however, does not forecast well.

Although in general it holds that imposing the Nelson-Siegel exponential structure on the factor loadings certainly helps compared to the AR(1) and VAR(1) yield level models, the three-factor model, and in particular the two-factor model, forecasts rather disappointingly. As Diebold and Li (2006) report very accurate forecasts for the NS2-AR model it will be interesting to see how well the performance of the four-factor model holds up for the 1994-2000 period.

5.7.5 Subsample results

Sample 1994:1 - 2000:12

Except from an initial surge during 1994, interest rates are fairly stable during this period which is characterized by a substantial amount of predictability as shown in Tables 5.10 and 5.11. Although for a one-month horizon this is not case, it is certainly true for longer horizons. The results I find for the NS2-AR and NS2-VAR are very similar to those reported in Diebold and Li (2006) and Mönch (2006b). For the 3-month horizon, the two-step NS2-3-AR model produces accurate forecasts and performs better than the NS2-3-VAR model. Although NS2-2-λ-AR, NS2-3-λ-AR and NS2-AS-AR have a lower TRMSPEs (0.91 vs

Table 5.10: Forecast results for the sample 1994:1 - 2000:12, 1 & 3-month horizons

	***h* = 1**									***h* = 3**								
	TRMSPE	**RMSPE**								**TRMSPE**	**RMSPE**							
Maturity	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y
RW	92.65	29.82	17.87	19.30	23.95	26.84	27.48	26.40	25.31	184.76	45.82	36.70	41.99	50.42	57.46	55.79	53.25	49.22
AR	1.00	**0.99**	1.01	1.00	1.00	1.00	1.00	1.00	1.00	**0.99**	**0.97**	1.02	1.01	**0.99**	**0.99**	**0.99**	**0.99**	1.00
VAR	1.00	**0.82**	1.01	**0.93**	**0.97**	1.07	1.04	1.03	1.06	**0.97**	**0.69**	**0.86**	**0.90**	**0.95**	**0.99**	1.00	1.00	1.06
Panel A: two-step models																		
NS2-2-AR	1.08	1.30	1.25	1.00	**0.98**	1.04	1.02	1.06	1.20	**0.98**	**0.85**^{5}	**0.92**^{10}	**0.97**	**0.95**	**0.94**	**0.99**	1.03	1.13
NS2-2-VAR	1.08	1.28	1.25	1.03	1.01	1.05	1.03	1.07	1.21	1.01	**0.86**^{5}	**0.99**	1.03	1.00	**0.97**	1.01	1.05	1.15
NS2-3-AR	**0.98**	**0.90**	**0.91**	1.00	**0.99**	1.02	1.02	1.02	1.00	**0.94**	**0.73**^{5}	**0.90**	**0.97**	**0.96**	**0.96**	**0.96**	**0.96**	**0.97**
NS2-3-VAR	1.01	**0.99**	**0.93**	**0.96**	**0.98**	1.06	1.05	1.04	1.02	1.00	**0.80**^{10}	**0.88**	**0.98**	1.01	1.04	1.04	1.02	1.02
NS2-4-AR	1.01	**0.88**	1.33	1.20	1.07	1.02	**0.97**	1.03	**0.99**	1.00	1.07	1.41	1.23	1.07	**0.96**	**0.92**^{5}	**0.95**^{5}	**0.95**^{10}
NS2-4-VAR	**0.99**	**0.80**	**0.91**	**0.96**	**0.99**	1.07	1.02	1.04	1.04	1.00	**0.68**^{1}	**0.89**	**0.99**	1.02	1.05	1.02	1.02	1.04
NS2-2-λ-AR	1.16	**0.93**	1.11	1.30	1.41	1.37	1.16	1.12	1.10	**0.91**	**0.73**^{5}	**0.93**	1.00	**0.98**	**0.93**	**0.91**	**0.92**	**0.97**
NS2-2-λ-VAR	1.17	**0.93**	1.14	1.33	1.43	1.38	1.17	1.12	1.10	**0.95**	**0.78**^{1}	1.02	1.08	1.04	**0.97**	**0.93**	**0.94**	**0.99**
NS2-3-λ-AR	1.15	**0.87**	1.22	1.41	1.35	1.25	1.17	1.16	1.15	**0.91**	**0.75**	**0.96**	1.00	**0.95**	**0.88**	**0.91**	**0.93**	**0.97**
NS2-3-λ-VAR	1.16	**0.85**	1.17	1.41	1.40	1.30	1.16	1.14	1.19	**0.99**	**0.75**	1.00	1.09	1.09	1.02	**0.97**	**0.99**	1.07
NS2-4-λ-AR	1.02	**0.90**	1.34	1.23	1.06	1.02	1.01	1.04	1.05	**0.96**	1.00	1.29	1.14	**0.99**	**0.91**	**0.92**	**0.95**	**0.97**
NS2-4-λ-VAR	**0.99**	**0.85**	1.01	1.01	**0.98**	1.04	1.01	1.02	1.08	**0.96**	**0.69**	**0.86**	**0.93**	**0.95**	**0.98**	**0.98**	1.00	1.06
NS2-B-AR	1.17	**0.87**	1.17	1.15	1.01	1.05	1.28	1.30	1.28	**0.96**	**0.77**	**0.97**	**0.94**	**0.86**	**0.85**	1.00	1.04	1.09
NS2-B-VAR	1.20	**0.87**	1.08	1.11	1.07	1.12	1.30	1.34	1.35	1.05	**0.75**	**0.92**	**0.98**	**0.99**	**0.98**	1.08	1.13	1.21
NS2-S-AR	1.18	**0.81**	1.11	1.25	1.15	1.14	1.24	1.30	1.36	**0.97**	**0.78**	**0.92**	**0.93**	**0.88**	**0.89**	1.00	1.05	1.14
NS2-S-VAR	1.12	**0.81**	1.08	1.21	1.14	1.10	1.13	1.19	1.31	**0.99**	**0.74**^{10}	**0.90**	**0.97**	**0.97**	**0.97**	1.00	1.04	1.17
NS2-AS-AR	1.20	**0.90**	1.37	1.40	1.31	1.25	1.22	1.25	1.23	**0.91**	**0.84**	1.02	**0.97**	**0.89**	**0.86**	**0.91**	**0.94**	**0.98**
NS2-AS-VAR	1.19	**0.90**	1.35	1.39	1.34	1.28	1.18	1.20	1.23	**0.94**	**0.80**	**0.99**	**0.98**	**0.98**	**0.95**	**0.93**	**0.96**	1.04
Panel B: one-step models																		
NS1-2-RW	1.28	2.14	2.40	1.61	1.01	1.05	1.01	1.02	1.08	1.04	1.28	1.23	1.09	**0.99**	1.00	1.00	1.01	1.05
NS1-2-AR	1.33	2.21	2.51	1.71	1.05	1.07	1.06	1.04	1.13	1.08	1.42	1.37	1.18	1.02	1.02	1.03	1.03	1.11
NS1-2-VAR	1.32	2.18	2.47	1.68	1.05	1.08	1.07	1.05	1.14	1.09	1.36	1.33	1.17	1.03	1.03	1.05	1.05	1.13
NS1-3-RW	1.00	1.01	**0.99**	**0.98**	**0.99**	1.03	1.02	1.02	1.01	**0.99**	**0.76**^{1}	**0.92**	**0.99**	1.00	1.01	1.01	1.01	1.01
NS1-3-AR	1.01	1.04	1.01	1.01	1.01	1.04	1.02	1.02	1.01	**0.96**	**0.78**^{10}	**0.92**	**0.99**	**0.98**	**0.98**	**0.97**	**0.97**	**0.98**
NS1-3-VAR	1.06	1.15	1.11	1.00	**0.96**	1.05	1.10	1.08	1.08	1.06	**0.90**	**0.97**	1.02	1.01	1.05	1.11	1.09	1.13
NS1-4-RW	**0.99**	**0.87**	1.05	**0.99**	**0.96**	1.04	1.00	1.02	1.04	**0.99**	**0.81**^{1}	1.00	1.00	**0.98**^{10}	1.01	1.00	1.01	1.02
NS1-4-AR	**0.98**	**0.84**	**0.99**	**0.95**	**0.94**	1.02	1.00	1.03	1.05	**0.95**	**0.70**^{10}	**0.90**	**0.93**	**0.93**	**0.96**	**0.97**	**0.98**	1.01
NS1-4-VAR	1.01	**0.84**	**0.94**	**0.93**	**0.95**	1.05	1.04	1.06	1.13	1.02	**0.68**^{5}	**0.86**	**0.94**	**0.97**	1.03	1.05	1.07	1.15
NS1-B-RW	1.00	**0.91**	1.03	1.00	**0.95**	1.04	1.03	1.02	1.02	**0.99**	**0.82**^{1}	1.00	1.00	**0.98**^{10}	1.01	1.01	1.00	1.02
NS1-B-AR	**0.99**	**0.90**	1.02	1.00	**0.95**	1.03	1.03	1.02	1.02	**0.97**	**0.79**^{1}	**0.97**	**0.98**	**0.96**	**0.98**	**0.98**	**0.99**	1.00
NS1-B-VAR	1.03	**0.92**	**0.96**	**0.96**	**0.94**	1.07	1.11	1.08	1.10	1.06	**0.79**^{10}	**0.92**	1.00	1.01	1.07	1.11	1.10	1.14
NS1-S-RW	1.00	**0.91**	1.09	**0.99**	**0.95**	1.04	1.00	1.02	1.03	**0.99**	**0.85**^{1}	1.02	1.00	**0.98**^{5}	1.01	1.01	1.01	1.01
NS1-S-AR	1.01	**0.89**	1.08	1.00	**0.97**	1.05	1.02	1.04	1.06	**0.98**	**0.82**^{5}	1.00	**0.98**	**0.97**	**0.99**	**0.99**	**0.99**	1.01
NS1-S-VAR	1.01	**0.83**	**0.95**	**0.93**	**0.95**	1.05	1.04	1.07	1.12	1.02	**0.68**^{5}	**0.87**	**0.95**	**0.98**	1.02	1.05	1.07	1.14
NS1-AS-RW	1.00	**0.91**	1.09	**0.99**	**0.96**	1.04	1.00	1.02	1.03	**0.99**	**0.85**^{1}	1.02	1.00	**0.98**^{5}	1.01	1.01	1.01	1.01
NS1-AS-AR	1.00	**0.89**	1.08	**0.99**	**0.96**	1.05	1.02	1.04	1.05	**0.98**	**0.82**^{5}	1.00	**0.98**	**0.96**	**0.99**	**0.99**	**0.99**	1.01
NS1-AS-VAR	1.01	**0.83**	**0.95**	**0.93**	**0.95**	1.05	1.04	1.07	1.12	1.02	**0.68**^{5}	**0.87**	**0.95**	**0.98**	1.02	1.05	1.07	1.14

Notes: . Bold numbers indicate outperformance relative to the random walk (RW) whereas $()^{10}$, $()^{5}$ and $()^{1}$ 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 5.11: Forecast results for the sample 1994:1 - 2000:12, 6 & 12-month horizons

	$h=6$									$h=12$								
	TRMSPE	**RMSPE**								**TRMSPE**	**RMSPE**							
Maturity	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y
RW	271.12	63.55	59.67	65.57	74.29	83.88	82.10	77.99	73.00	366.08	94.51	93.83	97.71	101.96	108.91	107.22	102.70	98.50
AR	**0.98**	**0.95**	1.03	1.01	**0.98**	**0.97**	**0.97**	**0.98**	1.00	**0.97**	**0.92**^{1}	1.02	1.00	**0.96**^{1}	**0.94**^{1}	**0.95**^{5}	**0.97**	1.00
VAR	**0.97**	**0.68**^{5}	**0.79**	**0.85**^{10}	**0.91**	**0.92**	**0.99**	1.02	1.12	1.08	**0.70**^{1}	**0.73**^{1}	**0.77**^{1}	**0.81**^{1}	**0.88**	1.13	1.24	1.42
Panel A: two-step models																		
NS2-2-AR	**0.96**	**0.81**^{1}	**0.89**^{1}	**0.94**^{5}	**0.93**	**0.90**^{10}	**0.97**	1.01	1.09	**0.95**	**0.87**^{5}	**0.90**^{1}	**0.91**^{1}	**0.89**^{1}	**0.88**^{1}	**0.96**^{10}	**0.99**	1.05
NS2-2-VAR	1.01	**0.88**^{1}	1.00	1.04	1.01	**0.96**	1.01	1.04	1.12	1.01	1.01	1.06	1.05	1.01	**0.94**^{1}	**0.99**	1.02	1.08
NS2-3-AR	**0.92**	**0.80**^{5}	**0.90**	**0.94**	**0.93**	**0.90**^{5}	**0.93**^{5}	**0.93**^{1}	**0.94**	**0.90**	**0.85**^{5}	**0.88**^{5}	**0.90**^{1}	**0.89**^{1}	**0.86**^{1}	**0.91**^{1}	**0.92**^{1}	**0.95**^{5}
NS2-3-VAR	1.01	**0.83**^{1}	**0.91**^{5}	1.00	1.04	1.04	1.03	1.02	1.03	1.03	**0.95**	1.00	1.04	1.06	1.04	1.03	1.03	1.05
NS2-4-AR	**0.98**	1.15	1.28	1.17	1.06	**0.93**^{5}	**0.90**^{1}	**0.93**^{1}	**0.95**^{5}	**0.99**	1.06	1.11	1.07	1.02	**0.94**^{5}	**0.94**^{1}	**0.96**^{1}	1.01
NS2-4-VAR	1.01	**0.78**^{1}	**0.94**^{5}	1.01	1.05	1.03	1.02	1.02	1.05	1.03	**0.96**	1.02	1.05	1.06	1.03	1.01	1.02	1.06
NS2-2-λ-AR	**0.90**	**0.78**^{5}	**0.90**	**0.94**	**0.91**	**0.86**	**0.90**^{5}	**0.93**^{10}	**0.98**	**0.89**	**0.83**^{5}	**0.86**^{5}	**0.86**^{1}	**0.83**^{1}	**0.80**^{1}	**0.90**^{1}	**0.93**^{1}	**0.99**
NS2-2-λ-VAR	**0.95**	**0.88**^{1}	1.02	1.04	**0.99**	**0.91**	**0.93**	**0.96**	1.01	**0.94**	**0.98**	1.01	**0.99**	**0.93**^{1}	**0.85**^{1}	**0.91**^{1}	**0.94**^{5}	1.00
NS2-3-λ-AR	**0.87**	**0.73**	**0.86**	**0.88**	**0.85**	**0.80**^{5}	**0.88**^{5}	**0.92**^{10}	**0.98**	**0.88**	**0.72**^{1}	**0.77**^{1}	**0.79**^{1}	**0.78**^{1}	**0.77**^{1}	**0.92**^{1}	**0.97**^{5}	1.06
NS2-3-λ-VAR	**0.98**	**0.78**^{1}	**0.97**	1.03	1.03	**0.96**	**0.96**	1.00	1.10	**0.95**	**0.90**^{1}	**0.98**^{10}	**0.99**^{10}	**0.98**	**0.89**^{1}	**0.90**^{1}	**0.95**	1.06
NS2-4-λ-AR	**0.95**	1.03	1.14	1.04	**0.96**	**0.87**^{5}	**0.92**^{5}	**0.96**^{10}	1.01	**0.98**	**0.92**	**0.97**	**0.95**	**0.92**^{5}	**0.88**^{1}	**0.97**^{5}	1.02	1.10
NS2-4-λ-VAR	**0.98**	**0.72**^{1}	**0.86**	**0.93**	**0.96**	**0.96**	**0.99**	1.02	1.10	**0.99**	**0.85**^{5}	**0.92**	**0.94**	**0.95**^{10}	**0.94**^{1}	**0.99**	1.03	1.12
NS2-B-AR	**0.93**	**0.76**	**0.86**	**0.85**	**0.81**^{10}	**0.80**^{1}	**0.96**	1.01	1.09	**0.96**	**0.72**^{1}	**0.77**^{1}	**0.78**^{1}	**0.78**^{1}	**0.81**^{1}	1.02	1.08	1.18
NS2-B-VAR	1.03	**0.72**^{1}	**0.89**^{5}	**0.95**^{5}	**0.98**^{10}	**0.96**^{1}	1.05	1.11	1.22	1.07	**0.90**^{5}	1.00	1.02	1.03	**0.99**	1.06	1.12	1.24
NS2-S-AR	**0.96**	**0.76**	**0.83**	**0.84**	**0.83**	**0.84**^{5}	1.00	1.06	1.17	1.01	**0.70**^{1}	**0.73**^{1}	**0.77**^{1}	**0.79**^{1}	**0.86**^{1}	1.08	1.15	1.29
NS2-S-VAR	1.04	**0.77**^{1}	**0.89**^{5}	**0.96**	**0.99**	**0.99**	1.05	1.10	1.22	1.06	**0.91**	**0.96**	**0.99**	1.00	**0.98**	1.06	1.12	1.24
NS2-AS-AR	**0.88**	**0.81**	**0.90**	**0.87**	**0.83**	**0.79**^{1}	**0.88**^{5}	**0.93**^{10}	1.00	**0.91**	**0.74**^{1}	**0.78**^{5}	**0.78**^{1}	**0.78**^{1}	**0.78**^{1}	**0.95**^{1}	1.01	1.11
NS2-AS-VAR	**0.94**	**0.79**^{1}	**0.92**	**0.93**	**0.93**	**0.90**^{10}	**0.93**	**0.98**	1.08	**0.95**	**0.89**^{5}	**0.94**^{5}	**0.93**^{5}	**0.92**^{1}	**0.87**^{1}	**0.92**^{1}	**0.98**	1.09
Panel B: one-step models																		
NS1-2-RW	1.01	1.00	1.00	1.02	1.01	**0.99**	**0.99**	1.01	1.04	1.01	1.00	1.01	1.02	1.02	**0.98**^{5}	**0.99**^{10}	1.01	1.03
NS1-2-AR	1.05	1.19	1.13	1.10	1.03	1.00	1.03	1.04	1.10	1.04	1.13	1.08	1.06	1.01	**0.97**	1.01	1.03	1.11
NS1-2-VAR	1.06	1.12	1.10	1.09	1.05	1.02	1.05	1.06	1.14	1.06	1.09	1.07	1.07	1.03	**0.99**	1.04	1.07	1.16
NS1-3-RW	**0.99**	**0.84**^{1}	**0.95**^{5}	1.00	1.01	1.00	1.01	1.00	1.00	1.00	**0.95**	**0.99**	1.00	1.01	1.00	1.01	1.00	**0.99**
NS1-3-AR	**0.93**	**0.86**	**0.94**	**0.98**	**0.97**	**0.93**	**0.93**	**0.94**	**0.95**	**0.91**	**0.93**	**0.95**	**0.96**	**0.93**^{5}	**0.88**^{1}	**0.89**^{1}	**0.90**^{1}	**0.93**^{5}
NS1-3-VAR	1.10	**0.94**^{10}	**0.99**	1.05	1.06	1.07	1.13	1.13	1.18	1.13	1.04	1.06	1.08	1.08	1.07	1.15	1.17	1.25
NS1-4-RW	1.00	**0.92**^{1}	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	**0.99**	1.01	1.00	1.00	1.00	1.00	1.00	1.00
NS1-4-AR	**0.95**	**0.79**^{5}	**0.91**	**0.94**	**0.96**	**0.95**	**0.95**	**0.96**	**0.99**	**0.94**	**0.91**	**0.95**	**0.96**	**0.96**	**0.94**^{1}	**0.93**^{1}	**0.94**^{1}	**0.97**
NS1-4-VAR	1.07	**0.77**^{1}	**0.89**^{10}	**0.98**	1.03	1.05	1.09	1.12	1.20	1.13	**0.95**	**0.99**	1.04	1.07	1.08	1.14	1.19	1.29
NS1-B-RW	1.00	**0.93**^{5}	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
NS1-B-AR	**0.95**	**0.89**^{10}	**0.97**	**0.97**	**0.96**	**0.95**^{5}	**0.96**^{10}	**0.96**	**0.98**	**0.93**	**0.92**^{10}	**0.95**^{10}	**0.94**^{5}	**0.94**^{1}	**0.92**^{1}	**0.92**^{1}	**0.93**^{1}	**0.94**
NS1-B-VAR	1.11	**0.88**^{5}	**0.96**	1.05	1.08	1.10	1.15	1.15	1.21	1.18	1.04	1.06	1.11	1.13	1.15	1.21	1.23	1.29
NS1-S-RW	1.00	**0.95**^{5}	1.02	1.00	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
NS1-S-AR	**0.95**	**0.90**^{10}	**0.98**	**0.97**	**0.97**	**0.96**	**0.95**	**0.95**	**0.97**	**0.92**	**0.93**^{10}	**0.96**	**0.95**^{10}	**0.95**^{5}	**0.92**^{1}	**0.90**^{1}	**0.91**^{1}	**0.94**
NS1-S-VAR	1.07	**0.78**^{1}	**0.90**^{10}	**0.99**	1.04	1.05	1.09	1.12	1.19	1.13	**0.95**	1.00	1.05	1.08	1.08	1.14	1.19	1.28
NS1-AS-RW	1.00	**0.95**^{5}	1.01	1.00	1.00	1.00	1.01	1.01	1.01	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
NS1-AS-AR	**0.96**	**0.90**	**0.98**	**0.97**	**0.97**	**0.96**	**0.95**	**0.96**	**0.98**	**0.93**	**0.93**^{10}	**0.96**	**0.95**^{10}	**0.95**^{5}	**0.92**^{1}	**0.91**^{1}	**0.91**^{1}	**0.94**
NS1-AS-VAR	1.07	**0.78**^{1}	**0.90**^{10}	**0.99**	1.04	1.05	1.09	1.12	1.19	1.13	**0.95**	1.00	1.05	1.08	1.09	1.15	1.19	1.29

Notes: . Bold numbers indicate outperformance relative to the random walk (RW) whereas $()^{10}$, $()^{5}$ and $()^{1}$ 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.

0.94) all these models forecast well mainly for medium and long maturities whereas the NS2-3-AR model outperforms the random walk for all maturities. The same holds for the NS1-3-AR model which for short maturities is outperformed by the NS1-4-AR model which is the best one-step model. The VAR specifications still deliver the more accurate short-maturity forecasts but the AR specifications now also do well for short maturities as well as reasonably well for longer maturities.

The relative TRMSPE numbers for the 6 and 12-month horizons in Table 5.11 show that AR dynamics, either in the two-step or one-step estimation approach, clearly outperform VAR dynamics. The two-step three-factor model with the Diebold and Li (2006) approach of fixing λ_t indeed forecasts well. However, estimating the decay parameter alongside the factors seems worthwhile. The results for the NS2-3-λ-AR show that doing so improves forecasts for short and medium maturities, with strong statistical outperformance relative to the random walk, but that is also leads to a decrease in accuracy for long maturities. With the one-step approach, however, the random walk is outperformed fairly evenly for each maturity. The results for the one-step four-factor, Bliss and (Adjusted) Svensson models are all very similar.

Sample 2001:1 - 2003:12

Mönch (2006b) examines the performance of the NS2-AR and NS2-VAR models for the sample 2000:1 - 2003:9 and finds that it is much worse than for the Diebold and Li (2006) out-of-sample period 1994:1 - 2000:12. This leads him to conclude that "...some of the strong forecast performance of the Nelson-Siegel model documented by Diebold and Li may be due to their choice of forecast period". The results for the second subsample that I examine shed more light on this claim. In fact, the results in Tables 5.12 and 5.13 support Mönch's conclusion as the NS2-AR and NS2-VAR models are shown to forecast poorly for all horizons[25]. For the 1 and 3-month horizon there is some predictability again for short maturities using the one-step estimation methods with VAR dynamics for the factors. The four-factor performance is reasonable for the 3-month horizon with RMSPEs below one for short and long maturities although medium maturities are predicted poorly. For the 6 and 12-month horizon the VAR specification again beats the AR specification due to accurate short-maturity forecasts.

[25] See also the subsample analysis in De Pooter *et al.* (2007) who arrive at the same conclusion that the forecasting performance of the three-factor Nelson-Siegel varies substantially across subperiods.

Table 5.12: Forecast results for the sample 2001:1 - 2003:12, 1 & 3-month horizons

	***h* = 1**									***h* = 3**								
	TRMSPE	**RMSPE**								**TRMSPE**	**RMSPE**							
Maturity	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y
RW	119.88	30.79	27.39	26.81	29.40	33.84	36.55	35.16	33.31	222.46	69.23	68.35	67.30	66.12	66.01	61.41	54.99	51.99
AR	1.02	1.04	1.01	1.01	1.03	1.02	1.01	1.01	1.01	1.03	1.06	1.00	1.01	1.03	1.05	1.04	1.04	1.02
VAR	1.02	**0.76**	**0.90**	1.06	1.12	1.07	**0.98**	1.01	1.08	1.08	**0.78**	**0.92**	1.04	1.17	1.20	1.09	1.09	1.11
Panel A: two-step models																		
NS2-2-AR	1.31	1.98	1.60	1.11	1.20	1.55	1.06	1.05	1.34	1.09	0.94	**0.88**	**0.97**	1.17	1.33	1.09	1.04	1.13
NS2-2-VAR	1.33	2.18	1.80	1.22	1.12	1.48	1.04	1.07	1.37	1.09	1.07	**0.94**	**0.95**	1.07	1.24	1.08	1.06	1.17
NS2-3-AR	1.06	1.10	1.03	1.15	1.19	1.14	1.03	1.01	1.06	1.12	1.10	1.07	1.16	1.22	1.21	1.11	1.08	1.08
NS2-3-VAR	1.00	**0.85**	**0.81**	**0.98**	1.08	1.06	1.00	1.01	1.06	1.00	**0.80**	**0.82**	**0.95**	1.05	1.09	1.05	1.05	1.06
NS2-4-AR	1.14	1.13	1.19	1.27	1.32	1.28	1.13	1.04	1.04	1.19	1.03	1.06	1.17	1.31	1.38	1.24	1.14	1.09
NS2-4-VAR	**0.98**	**0.83**	**0.80**	**0.90**	**0.96**	1.03	1.02	1.01	1.02	**0.96**	**0.75**	**0.75**	**0.86**	**0.95**	1.04	1.05	1.04	1.05
NS2-2-λ-AR	2.21	1.15	**0.91**	1.46	2.44	2.98	2.47	2.04	1.81	1.54	**0.90**	**0.97**	1.23	1.61	1.91	1.75	1.58	1.45
NS2-2-λ-VAR	2.17	1.28	**0.95**	1.35	2.31	2.89	2.43	2.01	1.79	1.48	**0.93**	**0.93**	1.13	1.49	1.80	1.69	1.54	1.43
NS2-3-λ-AR	1.32	1.28	1.30	1.41	1.37	1.44	1.36	1.22	1.16	1.30	1.27	1.29	1.40	1.47	1.45	1.27	1.17	1.11
NS2-3-λ-VAR	1.28	**0.91**	**0.91**	1.06	1.12	1.37	1.38	1.30	1.20	1.00	**0.88**	**0.88**	0.98	1.03	1.07	1.03	1.02	**0.99**
NS2-4-λ-AR	1.19	1.10	1.17	1.29	1.39	1.39	1.19	1.08	1.08	1.23	1.02	1.07	1.20	1.37	1.44	1.28	1.17	1.11
NS2-4-λ-VAR	1.02	**0.79**	**0.78**	**0.92**	1.03	1.15	1.05	1.02	1.03	**0.94**	**0.72**	**0.74**	**0.86**	**0.97**	1.06	1.01	**0.98**	**0.99**
NS2-B-AR	1.21	1.34	1.42	1.58	1.51	1.31	1.13	1.04	1.04	1.33	1.29	1.32	1.44	1.53	1.49	1.29	1.19	1.14
NS2-B-VAR	1.05	**0.97**	1.03	1.22	1.20	1.11	1.03	1.00	**0.98**	1.00	**0.88**	**0.91**	1.02	1.09	1.10	**1.02**	**0.98**	**0.96**
NS2-S-AR	1.45	1.43	1.58	1.81	1.85	1.71	1.41	1.24	1.22	1.63	1.41	1.53	1.73	1.93	1.93	1.62	1.47	1.37
NS2-S-VAR	1.10	**0.88**	**0.87**	1.02	1.09	1.18	1.14	1.11	1.08	1.01	**0.80**	**0.82**	**0.94**	1.03	1.09	1.06	1.06	1.08
NS2-AS-AR	1.41	1.39	1.48	1.65	1.64	1.59	1.42	1.25	1.16	1.46	1.32	1.40	1.56	1.68	1.67	1.46	1.34	1.26
NS2-AS-VAR	1.30	**0.98**	**0.98**	1.16	1.23	1.41	1.39	1.31	1.19	1.09	**0.88**	**0.91**	1.03	1.11	1.16	1.15	1.14	1.14
Panel B: one-step models																		
NS1-2-RW	1.30	2.39	2.22	1.74	1.13	1.06	**0.98**	1.01	1.10	1.02	1.11	1.02	1.00	**0.99**	1.04	**0.99**	1.01	1.06
NS1-2-AR	1.27	2.26	2.09	1.65	1.13	1.11	1.01	1.01	1.10	1.05	1.04	**0.98**	1.02	1.07	1.13	1.05	1.04	1.08
NS1-2-VAR	1.32	2.40	2.23	1.76	1.16	1.10	1.02	1.03	1.13	1.09	1.15	1.06	1.06	1.06	1.12	1.08	1.09	1.14
NS1-3-RW	1.02	1.19	1.01	1.01	1.03	1.04	**0.99**	**0.99**	1.01	1.00	1.04	**0.96**	**0.99**	1.01	1.04	1.01	1.00	**0.99**^{5}
NS1-3-AR	1.05	1.20	1.06	1.10	1.13	1.11	1.02	1.00	1.02	1.07	1.08	1.04	1.10	1.15	1.15	1.07	1.02	1.01
NS1-3-VAR	1.02	**0.89**	**0.80**	**0.96**	1.09	1.10	1.03	1.02	1.06	1.04	**0.80**	**0.83**	**0.96**	1.08	1.15	1.12	1.10	1.10
NS1-4-RW	1.00	1.04	**0.97**	1.02	**0.99**	1.03	1.01	1.00	**0.98**^{5}	1.01	1.04	**0.98**	1.01	1.00	1.03	1.02	1.00	**0.98**^{5}
NS1-4-AR	**0.99**	1.00	**0.95**	**0.99**	**0.99**	1.04	1.00	**0.97**	**0.98**	**0.97**	**0.90**	**0.87**	**0.93**	**0.97**	1.05	1.02	**0.98**	**0.97**
NS1-4-VAR	**0.97**	**0.79**	**0.77**	**0.86**	**0.93**	1.02	1.01	**0.99**	1.00	**0.94**	**0.69**	**0.68**	**0.79**	**0.90**	1.02	1.05	1.03	1.03
NS1-B-RW	1.03	1.30	1.07	1.02	1.02	1.03	**0.98**	**0.99**	1.06	1.01	1.12	1.00	1.01	1.00	1.03	**0.99**	1.00	1.03
NS1-B-AR	1.05	1.31	1.11	1.10	1.10	1.07	1.00	**0.99**	1.06	1.07	1.15	1.07	1.10	1.12	1.12	1.04	1.02	1.04
NS1-B-VAR	1.01	**0.92**	**0.80**	**0.95**	1.08	1.06	1.01	1.02	1.10	1.03	**0.77**	**0.79**	**0.93**	1.06	1.12	1.09	1.10	1.15
NS1-S-RW	1.02	1.19	1.03	1.02	**0.98**	1.04	1.01	1.00	**0.99**	1.01	1.11	1.00	1.01	**0.99**	1.03	1.02	1.01	**0.97**^{10}
NS1-S-AR	1.04	1.26	1.11	1.09	1.05	1.09	1.02	**0.98**	1.01	1.07	1.17	1.07	1.08	1.09	1.13	1.06	1.01	1.01
NS1-S-VAR	**0.97**	**0.86**	**0.79**	**0.86**	**0.92**	1.02	1.00	**0.99**	1.00	**0.94**	**0.71**	**0.69**	**0.79**	**0.89**	1.02	1.04	1.03	1.04
NS1-AS-RW	1.02	1.18	1.03	1.02	**0.98**	1.04	1.01	1.00	**0.99**	1.01	1.11	1.00	1.01	**0.99**	1.03	1.02	1.01	**0.97**^{10}
NS1-AS-AR	1.04	1.25	1.10	1.08	1.04	1.08	1.01	**0.98**	1.00	1.05	1.16	1.06	1.07	1.07	1.11	1.04	1.00	**0.99**
NS1-AS-VAR	**0.97**	**0.85**	**0.79**	**0.86**	**0.92**	1.02	1.00	**0.99**	1.00	**0.94**	**0.71**	**0.69**	**0.79**	**0.89**	1.02	1.04	1.03	1.04

Notes: . Bold numbers indicate outperformance relative to the random walk (RW) whereas $()^{10}$, $()^{5}$ and $()^{1}$ 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 5.13: Forecast results for the sample 2001:1 - 2003:12, 6 & 12-month horizons

	$h = 6$									$h = 12$								
	TRMSPE	**RMSPE**								**TRMSPE**	**RMSPE**							
Maturity	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y	all	1-m	3-m	6-m	1-y	2-y	5-y	7-y	10-y
RW	342.90	118.17	118.78	116.69	111.89	104.27	88.12	75.10	65.31	548.98	199.40	204.79	204.91	197.20	173.72	127.44	104.69	84.53
AR	1.05	1.05	0.99	1.00	1.03	1.06	1.08	1.07	1.06	1.05	1.04	0.97	0.98	1.01	1.06	1.13	1.15	1.18
VAR	1.29	1.04	1.15	1.24	1.38	1.43	1.32	1.33	1.38	1.59	1.33	1.39	1.43	1.53	1.64	1.73	1.84	2.00
Panel A: two-step models																		
NS2-2-AR	1.09	**0.93**	**0.94**	1.03	1.16	1.25	1.12	1.07	1.16	1.12	1.03	1.03	1.06	1.12	1.19	1.19	1.18	1.27
NS2-2-VAR	1.06	**0.93**	**0.89**	**0.95**	1.05	1.16	1.12	1.11	1.23	1.07	**0.94**^{1}	**0.93**^{1}	**0.96**	1.02	1.11	1.18	1.21	1.34
NS2-3-AR	1.15	1.10	1.09	1.15	1.21	1.24	1.16	1.12	1.16	1.14	1.07	1.06	1.08	1.13	1.19	1.20	1.19	1.27
NS2-3-VAR	1.03	**0.87**	**0.88**	**0.97**	1.06	1.12	1.10	1.08	1.14	1.06	**0.97**	**0.96**	1.00	1.05	1.12	1.14	1.15	1.25
NS2-4-AR	1.13	**0.92**	**0.96**	1.05	1.19	1.29	1.20	1.13	1.12	1.01	**0.79**	**0.82**	**0.89**	1.00	1.11	1.15	1.14	1.20
NS2-4-VAR	**0.97**	**0.77**	**0.79**	**0.87**	**0.96**	1.04	1.07	1.08	1.16	1.01	**0.86**^{10}	**0.87**	**0.92**	**0.98**	1.05	1.11	1.15	1.31
NS2-2-λ-AR	1.31	1.00	1.03	1.17	1.36	1.53	1.45	1.38	1.37	1.22	1.05	1.05	1.10	1.18	1.30	1.37	1.38	1.48
NS2-2-λ-VAR	1.24	**0.96**^{10}	**0.97**	1.07	1.24	1.42	1.39	1.34	1.36	1.17	**0.99**	**0.99**	1.03	1.10	1.22	1.33	1.36	1.49
NS2-3-λ-AR	1.33	1.22	1.24	1.33	1.43	1.48	1.34	1.27	1.27	1.25	1.12	1.12	1.16	1.23	1.32	1.36	1.36	1.45
NS2-3-λ-VAR	**0.96**	**0.89**	**0.89**	**0.96**	1.00	1.02	**0.96**	**0.96**	1.03	**0.96**	**0.91**	**0.90**	**0.93**	**0.96**	1.00	**0.99**^{5}	1.02	1.14
NS2-4-λ-AR	1.18	**0.95**	1.00	1.10	1.25	1.36	1.26	1.19	1.19	1.07	**0.84**	**0.87**	**0.94**	1.05	1.17	1.23	1.24	1.33
NS2-4-λ-VAR	**0.93**	**0.77**	**0.78**	**0.87**	**0.95**	1.01	1.00	1.00	1.09	**0.98**	**0.85**	**0.86**	**0.90**	**0.96**	1.02	1.06	1.11	1.26
NS2-B-AR	1.36	1.22	1.25	1.35	1.46	1.52	1.39	1.33	1.33	1.29	1.11	1.12	1.17	1.25	1.37	1.43	1.44	1.54
NS2-B-VAR	1.01	**0.88**	**0.90**	**0.98**	1.06	1.10	1.05	1.03	1.08	1.04	**0.91**	**0.91**	**0.95**	1.01	1.09	1.15	1.18	1.30
NS2-S-AR	1.60	1.33	1.41	1.57	1.76	1.84	1.64	1.56	1.54	1.42	1.19	1.22	1.29	1.40	1.51	1.56	1.59	1.71
NS2-S-VAR	**0.99**	**0.82**	**0.85**	**0.94**	1.02	1.06	1.04	1.06	1.16	1.03	**0.89**	**0.91**	**0.96**	1.02	1.07	1.10	1.15	1.30
NS2-AS-AR	1.45	1.25	1.32	1.44	1.58	1.64	1.47	1.39	1.38	1.29	1.12	1.14	1.20	1.29	1.38	1.40	1.41	1.49
NS2-AS-VAR	1.00	**0.89**	**0.91**	1.00	1.06	1.08	1.02	1.02	1.10	**0.99**	**0.90**	**0.92**	**0.96**	1.01	1.04	1.03	1.05	1.17
Panel B: one-step models																		
NS1-2-RW	**0.98**	**0.95**	**0.92**	**0.96**	**0.99**	1.02	**0.99**	1.01	1.12	1.00	1.02	**0.99**	1.00	1.00	1.02	**0.99**	1.01	1.13
NS1-2-AR	1.08	**0.99**	**0.98**	1.05	1.11	1.16	1.09	1.09	1.18	1.16	1.12	1.10	1.12	1.15	1.19	1.18	1.21	1.34
NS1-2-VAR	1.10	1.01	**0.98**	1.03	1.08	1.14	1.14	1.18	1.31	1.16	1.07	1.05	1.07	1.10	1.15	1.24	1.32	1.54
NS1-3-RW	1.00	1.02	**0.97**	**0.99**	1.00	1.03	1.01	1.00	1.04	1.00	1.03	**0.99**	**0.99**	1.00	1.02	1.00	**0.99**	1.06
NS1-3-AR	1.08	1.07	1.05	1.10	1.14	1.16	1.07	1.03	1.06	1.07	1.06	1.04	1.05	1.08	1.11	1.09	1.07	1.12
NS1-3-VAR	1.11	**0.90**	**0.92**	1.02	1.12	1.21	1.21	1.21	1.27	1.19	1.03	1.03	1.07	1.14	1.24	1.32	1.36	1.50
NS1-4-RW	1.01	1.03	**0.98**	1.00	1.00	1.02	1.02	1.00	1.02	1.01	1.03	1.00	1.00	1.00	1.02	1.01	1.00	1.05
NS1-4-AR	**0.91**	**0.85**	**0.84**	**0.88**	**0.91**	**0.96**	**0.94**	**0.92**	**0.97**	**0.85**	**0.82**	**0.81**	**0.84**	**0.86**	**0.87**	**0.87**	**0.89**	1.00
NS1-4-VAR	**0.93**	**0.71**	**0.72**	**0.81**	**0.90**	1.01	1.06	1.06	1.14	1.00	**0.81**^{10}	**0.83**	**0.89**	**0.96**	1.04	1.12	1.19	1.36
NS1-B-RW	1.01	1.06	**0.99**	1.00	1.00	1.02	**0.99**	**0.99**	1.09	1.01	1.04	1.00	1.00	1.00	1.02	**0.98**^{5}	**0.99**	1.10
NS1-B-AR	1.09	1.11	1.07	1.10	1.13	1.14	1.06	1.05	1.12	1.09	1.07	1.05	1.06	1.09	1.13	1.09	1.09	1.20
NS1-B-VAR	1.09	**0.84**	**0.87**	**0.98**	1.10	1.19	1.19	1.22	1.33	1.19	**0.98**	1.00	1.06	1.15	1.25	1.33	1.39	1.57
NS1-S-RW	1.01	1.06	1.00	1.00	1.00	1.02	1.02	1.01	1.02	1.01	1.04	1.00	1.00	1.00	1.02	1.01	1.00	1.04
NS1-S-AR	1.06	1.11	1.05	1.07	1.08	1.10	1.04	1.00	1.04	1.03	1.05	1.01	1.01	1.02	1.04	1.02	1.01	1.08
NS1-S-VAR	**0.93**	**0.71**	**0.72**	**0.80**	**0.89**	1.00	1.05	1.05	1.13	**0.98**	**0.80**^{10}	**0.81**	**0.87**	**0.94**	1.02	1.10	1.17	1.34
NS1-AS-RW	1.01	1.06	1.00	1.00	1.00	1.02	1.02	1.01	1.02	1.01	1.04	1.00	1.00	1.00	1.02	1.01	1.00	1.04
NS1-AS-AR	1.03	1.10	1.04	1.05	1.06	1.08	1.01	**0.97**	1.01	1.00	1.04	1.00	**0.99**	1.00	1.01	**0.97**	**0.97**^{10}	1.04
NS1-AS-VAR	**0.93**	**0.71**	**0.72**	**0.80**	**0.89**	1.00	1.05	1.05	1.14	**0.98**	**0.80**^{10}	**0.81**	**0.87**	**0.94**	1.02	1.10	1.17	1.34

Notes: . Bold numbers indicate outperformance relative to the random walk (RW) whereas $()^{10}$, $()^{5}$ and $()^{1}$ 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.

The only model that forecasts well in this period with its downward trend in short term interest rates and strong increase in interest spreads is the NS1-4-AR model. For the 1 and 3-month horizon the model has difficulties forecasting the two and five-year maturities. However, for longer horizons, the model produces increasingly accurate forecasts and does so consistently across all maturities. Especially for the 12-month horizon NS1-4-AR reduces RMSPEs relative to the random walk by at least 11% for maturities up to seven years resulting in an overall relative TRMSPE of 0.85. Mönch (2006a) compares the forecasting performance of a number of competing models, among which are the NS2-3-AR and NS2-3-VAR specifications, for a very similar period (2000:1 - 2003:9) and finds that his proposed Factor Augmented VAR model is the only model capable of accurately forecasting the term structure for longer forecast horizons. However, Table 5.13 shows that the four-factor model is a strong competitor for Mönch's FAVAR model.

5.8 Concluding remarks

In this chapter I compare the in-sample fit and out-of-sample performance of a range of different Nelson-Siegel specifications. The in-sample results show that more elaborate models which incorporate multiple decay parameters and additional slope or curvature factors improve the fit of the original Nelson and Siegel (1987) three-factor functional form. The four-factor model performs qualitatively similar as the popular Svensson (1994) model but has the advantage that it is easier to estimate as is it less affected by potential multicolinearity problems.

Besides an improved in-sample fit relative to the three-factor model I also document a better out-of-sample performance with the four-factor model. More specifically, using the Kalman filter to estimate the latent factors over time and assuming AR(1) factor dynamics produces forecasts which consistently outperform those of benchmark models such as the random walk and unrestricted VAR models. This outperformance holds across the maturity spectrum and is most prominent for longer forecast horizons.

The analysis of this chapter can be extended in a number of ways. Firstly, I have only judged the forecast performance of the various Nelson-Siegel models by considering statistical accuracy by means of the (T)RMSPE. Fabozzi *et al.* (2005) use the slope and curvature forecasts of the three-factor model to implement systematic trading strategies and assess the returns of these strategies. It will be interesting to conduct a similar type of analysis for the models used here to evaluate forecasts from an economic point of view. Secondly, the use of Bayesian inference techniques will be interesting to examine. Mönch (2006a) and De Pooter *et al.* (2007) both use MCMC methods to draw inference on the parameters and latent factors in the three-factor model. Explicitly taking into account parameter uncertainty may further improve the predictive accuracy of especially the more complex models. Finally, the use of macroeconomic variables and/or macroeconomic factors as in Diebold, Rudebusch, and Aruoba (2006b) can potentially further improve forecasts compared to the yields-only approach that I have used here. All these topics are part of ongoing research.

5A Appendix A: Estimation details

General specification

The general specification which captures all the different Nelson-Siegel specifications is given by

$$Y_t = X_t\beta_t + \varepsilon_t \tag{5A.1}$$
$$\beta_t = \mu + \Phi\beta_{t-1} + \nu_t \tag{5A.2}$$

with Y_t the $(N \times 1)$ vector of yields, $X_t\beta_t$ the Nelson-Siegel spot rate curve, β_t the $(K \times 1)$ vector of factors and X_t the $(N \times K)$ matrix of factor loadings. The latter are time-varying if the decay parameter(s) are estimated alongside the factors in the two-step procedure

Two-step procedure

I estimate the parameters in step one of the two-step estimation procedure by minimizing the sum of squared yield errors, $\sum_{i=1}^{N} [y_t(\tau_i) - \widehat{y}_t(\tau_i)]^2$. When the decay parameter is fixed in the two, three and four-factor models I apply OLS. When decay parameters are estimated alongside the factors I use NLS to find optimal parameter estimates. In the latter case, the parameters of the two, three and four-factor models are initialized at the Diebold and Li (2006) value for λ_t and the OLS estimates for the factors. For the Bliss model both $\lambda_{1,t}$ and $\lambda_{2,t}$ are initialized at 16.42. Determining starting values for the (Adjusted) Svensson model is somewhat more complex as there is the additional restriction on $\lambda_{1,t}$ and $\lambda_{2,t}$. As starting values for the level, slope, curvature factor and $\lambda_{1,t}$ I use the optimal factor and λ_t estimates from the three-factor model. The fourth factor, $\beta_{4,t}$, is initialized to zero. If $\widehat{\lambda}_{1,t}$ is larger than twice the minimum allowed value of 6.69 then $\lambda_{2,t}$ is initialized to $0.5\widehat{\lambda}_{1,t}$. When $\widehat{\lambda}_{1,t}$ is smaller than 13.38 then $\lambda_{1,t}$ and $\lambda_{2,t}$ are initialized to 13.38 and 6.69 respectively. By doing so all the restrictions on $\lambda_{1,t}$ and $\lambda_{2,t}$ are satisfied. Because the minimum distance restriction is only imposed for the Svensson model, I initialize $\lambda_{2,t}$ in the Adjusted Svensson model to $\frac{1}{2}(6.69 + \widehat{\lambda}_{1,t})$. As the two-step estimation procedure is numerically challenging because of the nonlinearity in the factor loadings, whenever possible I use the analytical gradient and hessian which are given in the Appendix to this paper[26].

One-step procedure

For the one-step state-space estimation method, which is only used when constructing forecasts, I maximize the likelihood given in (5.22). It is of particular importance to start the optimization procedure with accurate starting values because of the large number of parameters. For the two, three and four-factor and Bliss models I initialize the parameters as follows. The decay parameters are set to 16.42 and the factors to their two-step OLS estimates. The equation parameters μ and Φ in the state equations are initialized with the estimates from either a VAR model or AR models for the factors. The variance parameters in H and Q are initialized to one and the optimization is performed using standard deviations to ensure positive variance estimates. The covariance parameters in Q are initially set to zero. The Kalman filter is started with the unconditional mean and variance of the factor estimates and the first twelve observations are discarded when computing the likelihood in (5.22). The approach for the (Adjusted) Svensson model is the same except for the fact that I use the optimal factor estimates from the two-step procedure as starting values. Furthermore, $\lambda_{1,t}$ and $\lambda_{2,t}$ are initialized by using the median of the two-step estimates.

[26] The Appendix is available on `http://www.depooter.net`

5B Appendix: Derivations for gradient and hessian

5B.1 Two-factor Nelson-Siegel model

In this appendix I specify the gradient and Hessian for the two-factor Nelson-Siegel model where λ is estimated alongside with β_1 and β_2 (when λ is fixed, only the terms related to βs remain). First, define the objective function as

$$
\begin{aligned}
F(\beta_{1,t},\beta_{2,t},\lambda) &\equiv \sum_{i=1}^{N}\{y(\tau_i)-\hat{y}(\tau_i)\}^2=\sum_{i=1}^{N}\{E(\tau_i,\beta_{1,t},\beta_{2,t},\lambda)\}^2 \\
&= \sum_{i=1}^{N}\left\{y(\tau_i)-\beta_{1,t}-\beta_{2,t}\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_t}\right)}{\left(\frac{\tau_i}{\lambda_t}\right)}\right]\right\}^2
\end{aligned}
$$

In the remainder of this appendix I drop the time subscript t from the β's and λ. Also, define

$$
\begin{aligned}
K(\tau_i) &\equiv \frac{\partial E(\tau_i)}{\partial\lambda}=\beta_2\left[\frac{1-\exp(-\frac{\tau_i}{\lambda})}{\tau_i}-\frac{\exp(-\frac{\tau_i}{\lambda})}{\lambda}\right] \\
\frac{\partial K(\tau_i)}{\partial\lambda} &= \beta_2\exp\left(-\frac{\tau_i}{\lambda}\right)\left(\frac{\tau_i}{\lambda^3}\right)
\end{aligned}
$$

Gradient

The elements of the gradient $G\equiv\begin{pmatrix}\frac{\partial F(\cdot)}{\partial\beta_1} & \frac{\partial F(\cdot)}{\partial\beta_2} & \frac{\partial F(\cdot)}{\partial\lambda}\end{pmatrix}$ are defined as follows

$$
\begin{aligned}
\frac{\partial F(\cdot)}{\partial\beta_1} &= -2\textstyle\sum_{i=1}^{N}\{E(\tau_i)\} \\
\frac{\partial F(\cdot)}{\partial\beta_2} &= -2\textstyle\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]\right\} \\
\frac{\partial F(\cdot)}{\partial\lambda} &= -2\textstyle\sum_{i=1}^{N}\{E(\tau_i)K(\tau_i)\}
\end{aligned}
$$

Hessian

The elements of the Hessian $H\equiv\begin{bmatrix}\frac{\partial^2F(\cdot)}{(\partial\beta_1)^2} & \frac{\partial^2F(\cdot)}{\partial\beta_2\partial\beta_1} & \frac{\partial^2F(\cdot)}{\partial\lambda\partial\beta_1} \\ & \frac{\partial^2F(\cdot)}{(\partial\beta_2)^2} & \frac{\partial^2F(\cdot)}{\partial\lambda\partial\beta_2} \\ & & \frac{\partial^2F(\cdot)}{(\partial\lambda)^2}\end{bmatrix}$ are given by

$$
\begin{aligned}
\frac{\partial^2F(\cdot)}{(\partial\beta_1)^2} &= 2N \\
\frac{\partial^2F(\cdot)}{\partial\beta_2\partial\beta_1} &= 2\textstyle\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right\} \\
\frac{\partial^2F(\cdot)}{\partial\lambda\partial\beta_1} &= -2\textstyle\sum_{i=1}^{N}K(\tau_i) \\
\frac{\partial^2F(\cdot)}{(\partial\beta_2)^2} &= 2\textstyle\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]^2\right\} \\
\frac{\partial^2F(\cdot)}{\partial\lambda\partial\beta_2} &= -2\textstyle\sum_{i=1}^{N}\left\{K(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]+E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\tau_i}-\frac{\exp\left(-\frac{\tau_i}{\lambda}\right)}{\lambda}\right]\right\} \\
\frac{\partial^2F(\cdot)}{(\partial\lambda)^2} &= 2\textstyle\sum_{i=1}^{N}\left\{[K(\tau_i)]^2+E(\tau_i)\frac{\partial K(\cdot)}{\partial\lambda}\right\}
\end{aligned}
$$

5B.2 Three-factor Nelson-Siegel model

I specify the gradient and Hessian for the three-factor Nelson-Siegel model where λ is estimated alongside with β_1, β_2 and β_3. The objective function is defined as

$$\begin{aligned} F(\beta_1,\beta_2,\beta_3,\lambda) &\equiv \sum_{i=1}^{N}\{y(\tau_i)-\hat{y}(\tau_i)\}^2 = \sum_{i=1}^{N}\{E(\tau_i,\beta_1,\beta_2,\beta_3,\lambda)\}^2 \\ &= \sum_{i=1}^{N}\left\{y(\tau_i)-\beta_1-\beta_2\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]-\beta_3\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}-\exp\left(-\frac{\tau_i}{\lambda}\right)\right]\right\}^2 \end{aligned}$$

Also, define

$$\begin{aligned} K(\tau_i) &\equiv \frac{\partial E(\tau_i)}{\partial\lambda} = -(\beta_2+\beta_3)\left[\frac{1-\exp(-\frac{\tau_i}{\lambda})}{\tau_i}-\frac{\exp(-\frac{\tau_i}{\lambda})}{\lambda}\right]+\beta_3\exp\left(-\frac{\tau_i}{\lambda}\right)\left(\frac{\tau_i}{\lambda^2}\right) \\ \frac{\partial K(\tau_i)}{\partial\lambda} &= \left[\beta_2+\left(\frac{\tau_i}{\lambda}-1\right)\beta_3\right]\exp\left(-\frac{\tau_i}{\lambda}\right)\left(\frac{\tau_i}{\lambda^3}\right) \end{aligned}$$

Gradient

The elements of the gradient $G \equiv \begin{pmatrix} \frac{\partial F(\cdot)}{\partial\beta_1} & \frac{\partial F(\cdot)}{\partial\beta_2} & \frac{\partial F(\cdot)}{\partial\beta_3} & \frac{\partial F(\cdot)}{\partial\lambda} \end{pmatrix}$ are defined as follows

$$\begin{aligned} \frac{\partial F(\cdot)}{\partial\beta_1} &= -2\sum_{i=1}^{N}\{E(\tau_i)\} \\ \frac{\partial F(\cdot)}{\partial\beta_2} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]\right\} \\ \frac{\partial F(\cdot)}{\partial\beta_3} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}-\exp\left(-\frac{\tau_i}{\lambda}\right)\right]\right\} \\ \frac{\partial F(\cdot)}{\partial\lambda} &= -2\sum_{i=1}^{N}\{E(\tau_i)K(\tau_i)\} \end{aligned}$$

Hessian

The elements of the Hessian $H \equiv \begin{bmatrix} \frac{\partial^2 F(\cdot)}{(\partial\beta_1)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_2\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_1} \\ & \frac{\partial^2 F(\cdot)}{(\partial\beta_2)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_2} \\ & & \frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} & \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_3} \\ & & & \frac{\partial^2 F(\cdot)}{(\partial\lambda)^2} \end{bmatrix}$ are given by

$$\begin{aligned} \frac{\partial^2 F(\cdot)}{(\partial\beta_1)^2} &= 2N \\ \frac{\partial^2 F(\cdot)}{\partial\beta_2\partial\beta_1} &= 2\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right\} \\ \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_1} &= 2\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}-\exp\left(-\frac{\tau_i}{\lambda}\right)\right\} \\ \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_1} &= -2\sum_{i=1}^{N}K(\tau_i) \\ \frac{\partial^2 F(\cdot)}{(\partial\beta_2)^2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]^2\right\} \\ \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}-\exp\left(-\frac{\tau_i}{\lambda}\right)\right]\right\} \\ \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_2} &= -2\sum_{i=1}^{N}\left\{K(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]+E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\tau_i}-\frac{\exp\left(-\frac{\tau_i}{\lambda}\right)}{\lambda}\right]\right\} \end{aligned}$$

$$\frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} = 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right]^2\right\}$$

$$\frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_3} = -2\sum_{i=1}^N \left\{K(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right] + E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda}\right)}{\lambda} - \exp\left(-\frac{\tau_i}{\lambda}\right)\left(\frac{\tau_i}{\lambda^2}\right)\right]\right\}$$

$$\frac{\partial^2 F(\cdot)}{(\partial\lambda)^2} = 2\sum_{i=1}^N \left\{[K(\tau_i)]^2 + E(\tau_i)\frac{\partial K(\tau_i)}{\partial\lambda}\right\}$$

5B.3 Four-factor Nelson-Siegel model

I specify the gradient and Hessian for the four-factor Nelson-Siegel model where λ is estimated alongside with $\beta_1, \beta_2, \beta_3$ and β_4. The objective function is defined as

$$\begin{aligned} F(\beta_1, \beta_2, \beta_3, \beta_4, \lambda) &\equiv \sum_{i=1}^N \{y(\tau_i) - \hat{y}(\tau_i)\}^2 = \sum_{i=1}^N \{E(\tau_i, \beta_1, \beta_2, \beta_3, \beta_4, \lambda)\}^2 \\ &= \sum_{i=1}^N \left\{y(\tau_i) - \beta_1 - \beta_2\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right] - \beta_3\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right] - \beta_4\left[\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{\left(\frac{2\tau_i}{\lambda}\right)}\right]\right\}^2 \end{aligned}$$

Also, define

$$K(\tau_i) \equiv \frac{\partial E(\tau_i)}{\partial\lambda} = -(\beta_2+\beta_3)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda}\right)}{\lambda}\right] + \beta_3\exp\left(-\frac{\tau_i}{\lambda}\right)\left(\frac{\tau_i}{\lambda^2}\right) - \beta_4\left[\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{2\tau_i} - \frac{\exp\left(\frac{-2\tau_i}{\lambda}\right)}{\lambda}\right]$$

$$\frac{\partial K(\tau_i)}{\partial\lambda} = \left[\beta_2 + \left(\frac{\tau_i}{\lambda} - 1\right)\beta_3\right]\exp\left(-\frac{\tau_i}{\lambda}\right)\left(\frac{\tau_i}{\lambda^3}\right) + \beta_4\exp\left(-\frac{2\tau_i}{\lambda}\right)\left(\frac{2\tau_i}{\lambda^3}\right)$$

Gradient

The elements of the gradient $G \equiv \left(\frac{\partial F(\cdot)}{\partial\beta_1} \quad \frac{\partial F(\cdot)}{\partial\beta_2} \quad \frac{\partial F(\cdot)}{\partial\beta_3} \quad \frac{\partial F(\cdot)}{\partial\beta_4} \quad \frac{\partial F(\cdot}{\partial\lambda}\right)$ are defined as follows

$$\begin{aligned} \frac{\partial F(\cdot)}{\partial\beta_1} &= -2\sum_{i=1}^N \{E(\tau_i)\} \\ \frac{\partial F(\cdot)}{\partial\beta_2} &= -2\sum_{i=1}^N \left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]\right\} \\ \frac{\partial F(\cdot)}{\partial\beta_3} &= -2\sum_{i=1}^N \left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right]\right\} \\ \frac{\partial F(\cdot)}{\partial\beta_4} &= -2\sum_{i=1}^N \left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{\left(\frac{2\tau_i}{\lambda}\right)}\right]\right\} \\ \frac{\partial F(\cdot)}{\partial\lambda} &= -2\sum_{i=1}^N \{E(\tau_i)K(\tau_i)\} \end{aligned}$$

Hessian

The elements of the Hessian $H \equiv \begin{bmatrix} \frac{\partial^2 F(\cdot)}{(\partial\beta_1)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_2\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_1} \\ & \frac{\partial^2 F(\cdot)}{(\partial\beta_2)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_2} \\ & & \frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_3} & \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_3} \\ & & & \frac{\partial^2 F(\cdot)}{(\partial\beta_4)^2} & \frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_4} \\ & & & & \frac{\partial^2 F(\cdot)}{(\partial\lambda)^2} \end{bmatrix}$ are given by

$$\begin{aligned}
\frac{\partial^2 F(\cdot)}{(\partial\beta_1)^2} &= 2N \\
\frac{\partial^2 F(\cdot)}{\partial\beta_2\partial\beta_1} &= 2\sum_{i=1}^N \left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_1} &= 2\sum_{i=1}^N \left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_1} &= 2\sum_{i=1}^N \left\{\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{\left(\frac{2\tau_i}{\lambda}\right)}\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_1} &= -2\sum_{i=1}^N K(\tau_i) \\
\frac{\partial^2 F(\cdot)}{(\partial\beta_2)^2} &= 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]^2\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} &= 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_2} &= 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right]\left[\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{\left(\frac{2\tau_i}{\lambda}\right)}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_2} &= -2\sum_{i=1}^N \left\{K(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)}\right] + E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda}\right)}{\lambda}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} &= 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right]^2\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_3} &= 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{\left(\frac{2\tau_i}{\lambda}\right)}\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_3} &= -2\sum_{i=1}^N \left\{K(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda}\right)} - \exp\left(-\frac{\tau_i}{\lambda}\right)\right] + E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda}\right)}{\lambda} - \exp\left(-\frac{\tau_i}{\lambda}\right)\left(\frac{\tau_i}{\lambda^2}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\beta_4)^2} &= 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{\left(\frac{2\tau_i}{\lambda}\right)}\right]^2\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda\partial\beta_4} &= -2\sum_{i=1}^N \left\{K(\tau_i)\left[\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{\left(\frac{2\tau_i}{\lambda}\right)}\right] + E(\tau_i)\left[\frac{1-\exp\left(-\frac{2\tau_i}{\lambda}\right)}{2\tau_i} - \frac{\exp\left(-\frac{2\tau_i}{\lambda}\right)}{\lambda}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\lambda)^2} &= 2\sum_{i=1}^N \left\{[K(\tau_i)]^2 + E(\tau_i)\frac{\partial K(\tau_i)}{\partial\lambda}\right\}
\end{aligned}$$

5B.4 Bliss (1997) Nelson-Siegel model

I specify the gradient and the Hessian for the three-factor Nelson-Siegel model with the Bliss (1997) extension where λ_1 and λ_2 are estimated alongside with β_1, β_2 and β_3. The objective function is defined as

$$\begin{aligned}
F(\beta_1,\beta_2,\beta_3,\lambda_1,\lambda_2) &\equiv \sum_{i=1}^N \{y(\tau_i)-\hat{y}(\tau_i)\}^2 = \sum_{i=1}^N \{E(\tau_i,\beta_1,\beta_2,\beta_3,\lambda_1,\lambda_2)\}^2 \\
&= \sum_{i=1}^N \left\{y(\tau_i) - \beta_1 - \beta_2\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right] - \beta_3\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\right\}^2
\end{aligned}$$

Define

$$K_1(\tau_i) \equiv \frac{\partial E(\tau_i)}{\partial\lambda_1} \equiv -\beta_2\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\lambda_1}\right]$$

$$\begin{aligned}
\frac{\partial K_1(\tau_i)}{\partial \lambda_1} &= \beta_2 \exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i}{\lambda_1^3}\right) \\
K_2(\tau_i) &\equiv \frac{\partial E(\tau_i)}{\partial \lambda_2} \equiv -\beta_3 \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\lambda_2} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\left(\frac{\tau_i}{\lambda_2^2}\right)\right] \\
\frac{\partial K_2(\tau_i)}{\partial \lambda_2} &= \beta_3 \left[\exp\left(-\frac{\tau_i}{\lambda_2}\right)\left(\frac{\tau_i(\tau_i-\lambda_2)}{\lambda_2^4}\right)\right]
\end{aligned}$$

Gradient

The elements of the gradient

$$G \equiv \begin{pmatrix} \frac{\partial F(\cdot)}{\partial \beta_1} & \frac{\partial F(\cdot)}{\partial \beta_2} & \frac{\partial F(\cdot)}{\partial \beta_3} & \frac{\partial F(\cdot)}{\partial \lambda_1} & \frac{\partial F(\cdot)}{\partial \lambda_2} \end{pmatrix}$$

are defined as follows

$$\begin{aligned}
\tfrac{\partial F(\cdot)}{\partial \beta_1} &= -2\textstyle\sum_{i=1}^{N}\{E(\tau_i)\} \\
\tfrac{\partial F(\cdot)}{\partial \beta_2} &= -2\textstyle\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\} \\
\tfrac{\partial F(\cdot)}{\partial \beta_3} &= -2\textstyle\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\right\} \\
\tfrac{\partial F(\cdot)}{\partial \lambda_1} &= -2\textstyle\sum_{i=1}^{N}\{E(\tau_i)K_1(\tau_i)\} \\
\tfrac{\partial F(\cdot)}{\partial \lambda_2} &= -2\textstyle\sum_{i=1}^{N}\{E(\tau_i)K_2(\tau_i)\}
\end{aligned}$$

Hessian

The elements of the Hessian

$$H \equiv \begin{bmatrix}
\frac{\partial^2 F(\cdot)}{(\partial \beta_1)^2} & \frac{\partial^2 F(\cdot)}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 F(\cdot)}{\partial \beta_3 \partial \beta_1} & \frac{\partial^2 F(\cdot)}{\partial \lambda_1 \partial \beta_1} & \frac{\partial^2 F(\cdot)}{\partial \lambda_2 \partial \beta_1} \\
 & \frac{\partial^2 F(\cdot)}{(\partial \beta_2)^2} & \frac{\partial^2 F(\cdot)}{\partial \beta_3 \partial \beta_2} & \frac{\partial^2 F(\cdot)}{\partial \lambda_1 \partial \beta_2} & \frac{\partial^2 F(\cdot)}{\partial \lambda_2 \partial \beta_2} \\
 & & \frac{\partial^2 F(\cdot)}{(\partial \beta_3)^2} & \frac{\partial^2 F(\cdot)}{\partial \lambda_1 \partial \beta_3} & \frac{\partial^2 F(\cdot)}{\partial \lambda_2 \partial \beta_3} \\
 & & & \frac{\partial^2 F(\cdot)}{(\partial \lambda_1)^2} & \frac{\partial^2 F(\cdot)}{\partial \lambda_2 \partial \lambda_1} \\
 & & & & \frac{\partial^2 F(\cdot)}{(\partial \lambda_2)^2}
\end{bmatrix}$$

are given by

$$\begin{aligned}
\tfrac{\partial^2 F(\cdot)}{(\partial \beta_1)^2} &= 2N \\
\tfrac{\partial^2 F(\cdot)}{\partial \beta_2 \partial \beta_1} &= 2\textstyle\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right\} \\
\tfrac{\partial^2 F(\cdot)}{\partial \beta_3 \partial \beta_1} &= 2\textstyle\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right\} \\
\tfrac{\partial^2 F(\cdot)}{\partial \lambda_1 \partial \beta_1} &= -2\textstyle\sum_{i=1}^{N} K_1(\tau_i) \\
\tfrac{\partial^2 F(\cdot)}{\partial \lambda_2 \partial \beta_1} &= -2\textstyle\sum_{i=1}^{N} K_2(\tau_i) \\
\tfrac{\partial^2 F(\cdot)}{(\partial \beta_2)^2} &= 2\textstyle\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]^2\right\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)}-\exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_2} &= -2\sum_{i=1}^{N}\left\{K_1(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]+E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\tau_i}-\frac{\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\lambda_1}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_2} &= -2\sum_{i=1}^{N}\left\{K_2(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)}-\exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]^2\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_3} &= -2\sum_{i=1}^{N}\left\{K_1(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)}-\exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_3} &= -2\sum_{i=1}^{N}\left\{K_2(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)}-\exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]+E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\tau_i}-\frac{\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\lambda_2}-\exp\left(-\frac{\tau_i}{\lambda_2}\right)\left(\frac{\tau_i}{\lambda_2^2}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\lambda_1)^2} &= 2\sum_{i=1}^{N}\left\{[K_1(\tau_i)]^2+E(\tau_i)\frac{\partial K_1(\tau_i)}{\partial\lambda_1}\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\lambda_1} &= 2\sum_{i=1}^{N}\{K_1(\tau_i)K_2(\tau_i)\} \\
\frac{\partial^2 F(\cdot)}{(\partial\lambda_2)^2} &= 2\sum_{i=1}^{N}\left\{[K_2(\tau_i)]^2+E(\tau_i)\frac{\partial K_2(\tau_i)}{\partial\lambda_2}\right\}
\end{aligned}$$

5B.5 Svensson (1994) Nelson-Siegel model

I specify the gradient and Hessian for the four-factor Nelson-Siegel model with the Svensson (1994) extension where λ_1 and λ_2 are estimated alongside with $\beta_1, \beta_2, \beta_3$ and β_4. The objective function is defined as

$$F(\beta_1,\beta_2,\beta_3,\beta_4,\lambda_1,\lambda_2) \equiv \sum_{i=1}^{N}\{y(\tau_i)-\hat{y}(\tau_i)\}^2 = \sum_{i=1}^{N}\{E(\tau_i,\beta_1,\beta_2,\beta_3,\beta_4,\lambda_1,\lambda_2)\}^2$$

$$= \sum_{i=1}^{N}\left\{y(\tau_i)-\beta_1-\beta_2\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]-\beta_3\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}-\exp\left(-\frac{\tau_i}{\lambda_1}\right)\right]-\beta_4\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)}-\exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\right\}^2$$

Also, define

$$\begin{aligned}
K_1(\tau_i) &\equiv \frac{\partial E(\tau_i)}{\partial\lambda_1} = -(\beta_2+\beta_3)\left[\frac{1-\exp(-\frac{\tau_i}{\lambda_1})}{\tau_i}-\frac{\exp(-\frac{\tau_i}{\lambda_1})}{\lambda_1}\right]-\beta_3\exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i}{\lambda_1^2}\right) \\
\frac{\partial K_1(\tau_i)}{\partial\lambda_1} &= (\beta_2-\beta_3)\exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i}{\lambda_1^3}\right)-\beta_3\exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i^2}{\lambda_1^4}\right) \\
K_2(\tau_i) &\equiv \frac{\partial E(\tau_i)}{\partial\lambda_2} = -\beta_4\left[\frac{1-\exp(-\frac{\tau_i}{\lambda_2})}{\tau_i}-\frac{\exp(-\frac{\tau_i}{\lambda_2})}{\lambda_2}-\exp\left(-\frac{\tau_i}{\lambda_2}\right)\left(\frac{\tau_i}{\lambda_2^2}\right)\right] \\
\frac{\partial K_2(\tau_i)}{\partial\lambda_2} &= -\beta_4\left[\exp\left(-\frac{\tau_i}{\lambda_2}\right)\left(\frac{\tau_i}{\lambda_2^3}\right)-\exp\left(-\frac{\tau_i}{\lambda_2}\right)\left(\frac{\tau_i^2}{\lambda_2^4}\right)\right]
\end{aligned}$$

Gradient

The elements of the gradient $G \equiv \left(\frac{\partial F(\cdot)}{\partial\beta_1} \quad \frac{\partial F(\cdot)}{\partial\beta_2} \quad \frac{\partial F(\cdot)}{\partial\beta_3} \quad \frac{\partial F(\cdot)}{\partial\beta_4} \quad \frac{\partial F(\cdot)}{\partial\lambda_1} \quad \frac{\partial F(\cdot)}{\partial\lambda_2}\right)$

are defined as follows

$$
\begin{aligned}
\frac{\partial F(\cdot)}{\partial \beta_1} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\right\} \\
\frac{\partial F(\cdot)}{\partial \beta_2} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\} \\
\frac{\partial F(\cdot)}{\partial \beta_3} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\right]\right\} \\
\frac{\partial F(\cdot)}{\partial \beta_4} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\right\} \\
\frac{\partial F(\cdot)}{\partial \lambda_1} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)K_1(\tau_i)\right\} \\
\frac{\partial F(\cdot)}{\partial \lambda_2} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)K_2(\tau_i)\right\}
\end{aligned}
$$

Hessian

The elements of the Hessian $H \equiv \begin{bmatrix} \frac{\partial^2 F(\cdot)}{(\partial\beta_1)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_2\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_1} \\ & \frac{\partial^2 F(\cdot)}{(\partial\beta_2)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_2} \\ & & \frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_3} & \frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_3} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_3} \\ & & & \frac{\partial^2 F(\cdot)}{(\partial\beta_4)^2} & \frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_4} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_4} \\ & & & & \frac{\partial^2 F(\cdot)}{(\partial\lambda_1)^2} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\lambda_1} \\ & & & & & \frac{\partial^2 F(\cdot)}{(\partial\lambda_2)^2} \end{bmatrix}$ are given by

$$
\begin{aligned}
\frac{\partial^2 F(\cdot)}{(\partial\beta_1)^2} &= 2N \\
\frac{\partial^2 F(\cdot)}{\partial\beta_2\partial\beta_1} &= 2\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_1} &= 2\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_1} &= 2\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_1} &= -2\sum_{i=1}^{N}K_1(\tau_i) \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_1} &= -2\sum_{i=1}^{N}K_2(\tau_i) \\
\frac{\partial^2 F(\cdot)}{(\partial\beta_2)^2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]^2\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_2} &= -2\sum_{i=1}^{N}\left\{K_1(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right] + E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\lambda_1}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_2} &= -2\sum_{i=1}^{N}\left\{K_2(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\right]^2\right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_3} &= 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_3} &= -2\sum_{i=1}^N \left\{K(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\right] + E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\lambda_1} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i}{\lambda_1^2}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_3} &= -2\sum_{i=1}^N \left\{K_2(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\beta_4)^2} &= 2\sum_{i=1}^N \left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]^2\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_4} &= -2\sum_{i=1}^N \left\{K_1(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_4} &= -2\sum_{i=1}^N \left\{K_2(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\right] + E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\lambda_2} - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\left(\frac{\tau_i}{\lambda_2^2}\right)\right]\right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\lambda_1)^2} &= 2\sum_{i=1}^N \left\{[K_1(\tau_i)]^2 + E(\tau_i)\frac{\partial K_1(\tau_i)}{\partial\lambda_1}\right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\lambda_1} &= 2\sum_{i=1}^N \{K_1(\tau_i)K_2(\tau_i)\} \\
\frac{\partial^2 F(\cdot)}{(\partial\lambda_1)^2} &= 2\sum_{i=1}^N \left\{[K_2(\tau_i)]^2 + E(\tau_i)\frac{\partial K_2(\tau_i)}{\partial\lambda_1}\right\}
\end{aligned}
$$

5B.6 Adjusted Svensson (1994) Nelson-Siegel model

I specify the gradient and Hessian for the four-factor Nelson-Siegel model with the adjusted Svensson (1994) extension where λ_1 and λ_2 are estimated alongside with $\beta_1, \beta_2, \beta_3$ and β_4. The objective function is defined as

$$
F(\beta_1,\beta_2,\beta_3,\beta_4,\lambda_1,\lambda_2) \equiv \sum_{i=1}^N \{y(\tau_i) - \hat{y}(\tau_i)\}^2 = \sum_{i=1}^N \{E(\tau_i,\beta_1,\beta_2,\beta_3,\beta_4,\lambda_1,\lambda_2)\}^2
$$

$$
= \sum_{i=1}^N \left\{y(\tau_i) - \beta_1 - \beta_2\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right] - \beta_3\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\right] - \beta_4\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{2\tau_i}{\lambda_2}\right)\right]\right\}^2
$$

Also, define

$$
\begin{aligned}
K_1(\tau_i) &\equiv \frac{\partial E(\tau_i)}{\partial\lambda_1} = -(\beta_2+\beta_3)\left[\frac{1-\exp(-\frac{\tau_i}{\lambda_1})}{\tau_i} - \frac{\exp(-\frac{\tau_i}{\lambda_1})}{\lambda_1}\right] - \beta_3\exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i}{\lambda_1^2}\right) \\
\frac{\partial K_1(\tau_i)}{\partial\lambda_1} &= (\beta_2-\beta_3)\exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i}{\lambda_1^3}\right) - \beta_3\exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i}{\lambda_1^4}\right) \\
K_2(\tau_i) &\equiv \frac{\partial E(\tau_i)}{\partial\lambda_2} = -\beta_4\left[\frac{1-\exp(-\frac{\tau_i}{\lambda_2})}{\tau_i} - \frac{\exp(-\frac{\tau_i}{\lambda_2})}{\lambda_2} - \exp\left(-\frac{2\tau_i}{\lambda_2}\right)\left(\frac{2\tau_i}{\lambda_2^2}\right)\right] \\
\frac{\partial K_2(\tau_i)}{\partial\lambda_2} &= -\beta_4\left[\exp\left(-\frac{2\tau_i}{\lambda_2}\right)\left(-\frac{4\tau_i^2}{\lambda_2^4}\right) + \exp\left(-\frac{2\tau_i}{\lambda_2}\right)\left(\frac{4\tau_i}{\lambda_2^3}\right) - \exp\left(-\frac{\tau_i}{\lambda_2}\right)\left(\frac{\tau_i}{\lambda_2^3}\right)\right]
\end{aligned}
$$

Gradient

The elements of the gradient $G \equiv \begin{pmatrix} \frac{\partial F(\cdot)}{\partial\beta_1} & \frac{\partial F(\cdot)}{\partial\beta_2} & \frac{\partial F(\cdot)}{\partial\beta_3} & \frac{\partial F(\cdot)}{\partial\beta_4} & \frac{\partial F(\cdot)}{\partial\lambda_1} & \frac{\partial F(\cdot)}{\partial\lambda_2}\end{pmatrix}$

are defined as follows

$$
\begin{aligned}
\frac{\partial F(\cdot)}{\partial \beta_1} &= -2\sum_{i=1}^{N}\{E(\tau_i)\}\\
\frac{\partial F(\cdot)}{\partial \beta_2} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\}\\
\frac{\partial F(\cdot)}{\partial \beta_3} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}-\exp\left(-\frac{\tau_i}{\lambda_1}\right)\right]\right\}\\
\frac{\partial F(\cdot)}{\partial \beta_4} &= -2\sum_{i=1}^{N}\left\{E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)}-\exp\left(-\frac{2\tau_i}{\lambda_2}\right)\right]\right\}\\
\frac{\partial F(\cdot)}{\partial \lambda_1} &= -2\sum_{i=1}^{N}\{E(\tau_i)K_1(\tau_i)\}\\
\frac{\partial F(\cdot)}{\partial \lambda_2} &= -2\sum_{i=1}^{N}\{E(\tau_i)K_2(\tau_i)\}
\end{aligned}
$$

Hessian

The elements of the Hessian $H \equiv \begin{bmatrix} \frac{\partial^2 F(\cdot)}{(\partial\beta_1)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_2\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_1} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_1}\\ & \frac{\partial^2 F(\cdot)}{(\partial\beta_2)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_2} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_2}\\ & & \frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} & \frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_3} & \frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_3} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_3}\\ & & & \frac{\partial^2 F(\cdot)}{(\partial\beta_4)^2} & \frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_4} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_4}\\ & & & & \frac{\partial^2 F(\cdot)}{(\partial\lambda_1)^2} & \frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\lambda_1}\\ & & & & & \frac{\partial^2 F(\cdot)}{(\partial\lambda_2)^2} \end{bmatrix}$ are given by

$$
\begin{aligned}
\frac{\partial^2 F(\cdot)}{(\partial\beta_1)^2} &= 2N\\
\frac{\partial^2 F(\cdot)}{\partial\beta_2\partial\beta_1} &= 2\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right\}\\
\frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_1} &= 2\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}-\exp\left(-\frac{\tau_i}{\lambda_1}\right)\right\}\\
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_1} &= 2\sum_{i=1}^{N}\left\{\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)}-\exp\left(-\frac{2\tau_i}{\lambda_2}\right)\right\}\\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_1} &= -2\sum_{i=1}^{N}K_1(\tau_i)\\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_1} &= -2\sum_{i=1}^{N}K_2(\tau_i)\\
\frac{\partial^2 F(\cdot)}{(\partial\beta_2)^2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]^2\right\}\\
\frac{\partial^2 F(\cdot)}{\partial\beta_3\partial\beta_2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}-\exp\left(-\frac{\tau_i}{\lambda_1}\right)\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\}\\
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_2} &= 2\sum_{i=1}^{N}\left\{\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)}-\exp\left(-\frac{2\tau_i}{\lambda_2}\right)\right]\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\}\\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_2} &= -2\sum_{i=1}^{N}\left\{K_1(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]+E(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\tau_i}-\frac{\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\lambda_1}\right]\right\}\\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_2} &= -2\sum_{i=1}^{N}\left\{K_2(\tau_i)\left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)}\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 F(\cdot)}{(\partial\beta_3)^2} &= 2\sum_{i=1}^N \left\{ \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right) \right]^2 \right\} \\
\frac{\partial^2 F(\cdot)}{\partial\beta_4\partial\beta_3} &= 2\sum_{i=1}^N \left\{ \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right) \right] \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{2\tau_i}{\lambda_2}\right) \right] \right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_3} &= -2\sum_{i=1}^N \left\{ K(\tau_i) \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right) \right] + E(\tau_i) \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\lambda_1} - \exp\left(-\frac{\tau_i}{\lambda_1}\right)\left(\frac{\tau_i}{\lambda_1^2}\right) \right] \right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_3} &= -2\sum_{i=1}^N \left\{ K_2(\tau_i) \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_1}\right)}{\left(\frac{\tau_i}{\lambda_1}\right)} - \exp\left(-\frac{\tau_i}{\lambda_1}\right) \right] \right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\beta_4)^2} &= 2\sum_{i=1}^N \left\{ \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{2\tau_i}{\lambda_2}\right) \right]^2 \right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_1\partial\beta_4} &= -2\sum_{i=1}^N \left\{ K_1(\tau_i) \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{2\tau_i}{\lambda_2}\right) \right] \right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\beta_4} &= -2\sum_{i=1}^N \left\{ K_2(\tau_i) \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\left(\frac{\tau_i}{\lambda_2}\right)} - \exp\left(-\frac{2\tau_i}{\lambda_2}\right) \right] + E(\tau_i) \left[\frac{1-\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\tau_i} - \frac{\exp\left(-\frac{\tau_i}{\lambda_2}\right)}{\lambda_2} - \exp\left(-\frac{2\tau_i}{\lambda_2}\right)\left(\frac{2\tau_i}{\lambda_2^2}\right) \right] \right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\lambda_1)^2} &= 2\sum_{i=1}^N \left\{ [K_1(\tau_i)]^2 + E(\tau_i)\frac{\partial K_1(\tau_i)}{\partial\lambda_1} \right\} \\
\frac{\partial^2 F(\cdot)}{\partial\lambda_2\partial\lambda_1} &= 2\sum_{i=1}^N \left\{ K_1(\tau_i)K_2(\tau_i) \right\} \\
\frac{\partial^2 F(\cdot)}{(\partial\lambda_1)^2} &= 2\sum_{i=1}^N \left\{ [K_2(\tau_i)]^2 + E(\tau_i)\frac{\partial K_2(\tau_i)}{\partial\lambda_1} \right\}
\end{aligned}
$$

Chapter 6

Predicting the Term Structure of Interest Rates

Incorporating parameter uncertainty, model uncertainty and macroeconomic information

6.1 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 originated 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

[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.* (1999).

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 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.* (2006b) 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, 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

[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.

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 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) 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 evidence 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 strong support 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 6.2 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 6.3 to present the different models we use to construct forecasts. In Section 6.4 we discuss results of the individual models whereas in Section 6.5 we outline and discuss results of several forecasting combination schemes. Finally, in Section 6.6 we conclude. The Appendix provides details on the frequentist and Bayesian techniques that we use for estimating model parameters and for constructing forecasts.

6.2 Data

6.2.1 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 (2006a), our dataset consists of unsmoothed Fama-Bliss yields. These unsmoothed yields exactly price 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.

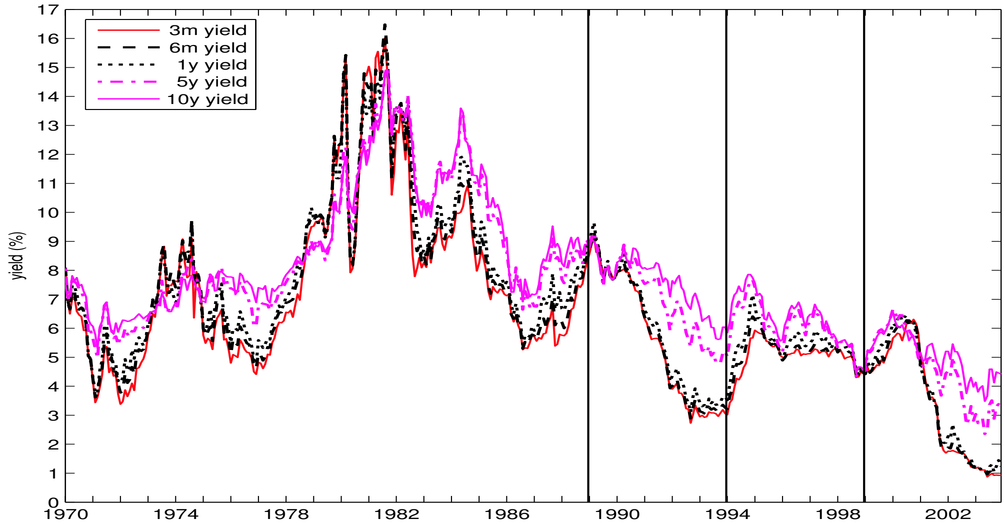
Throughout our analysis we use yields with $N = 13$ different maturities of $\tau = 1$, 3 and 6 months and 1, 2,..., 10 years. We denote yields by $y^{(\tau_i)}$ for $i = 1, \ldots, N$. To estimate the Nelson-Siegel models we follow Diebold and Li (2006) and Diebold *et al.* (2006b) by including additional 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 6.1 shows time-series plots for a subsample of the 13 maturities whereas Table 6.1 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

[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 6.1: U.S. zero-coupon yields



Notes: 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 indicate the three forecasting subsamples (1989:1 - 1993:12, 1994:1 - 1998:12 and 1999:1 - 2003:12).

Table 6.1: 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

Notes: 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$^{\text{st}}$, 12$^{\text{th}}$ and 24$^{\text{th}}$ sample autocorrelation. The results shown are for annualized yields (in %). The sample period is January 1970 - December 2003 (408 monthly observations).

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%.

6.2.2 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 6.2 lists the series included in 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 6.2 lists the applied transformations. We follow Ang and Piazzesi (2003), Mönch (2006a) and Diebold *et al.* (2006b) 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 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[6].

[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 fed funds rate because it closely follows the fed 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.

[6] Note that Ang and Piazzesi (2003) and Mönch (2006a) use *contemporaneous* macro information to construct their term structure forecasts. Using contemporaneous information may exaggerate the benefits from using macroeconomic series when forecasting yields. Note, however, that we are only able to fully mimic the information available to the econometrician at the time of making forecasts by using 'vintage data'. Croushore (2006) provides a discussion of the use of vintage data and shows that data revisions can lead to an improvement in perceived forecastability. Here we use only revised macroeconomic series meaning that this may effect out results as well.

Table 6.2: Macroeconomic 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)

Notes: 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).

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 6.2 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 *monetary aggregates* or *money stock* factor. Figure 6.3 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.

6.3 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

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

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.

6.3.1 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 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:

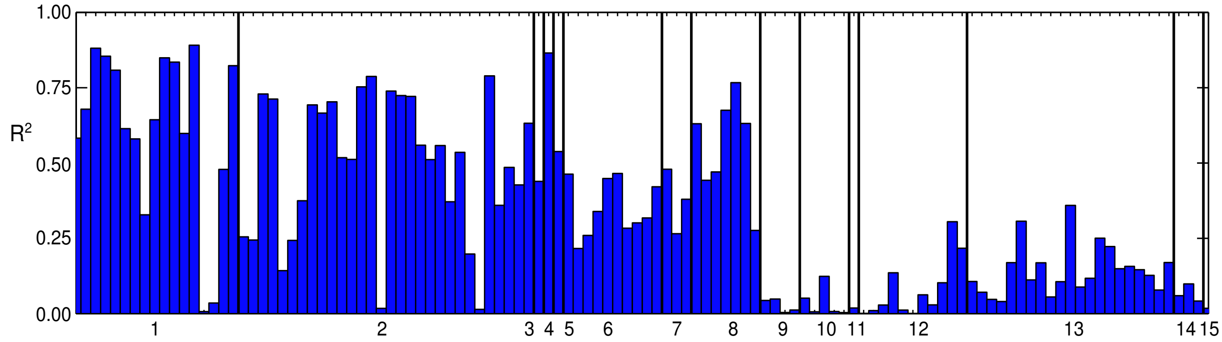
$$M_t = c + \Phi_1 M_{t-1} + \Phi_2 M_{t-2} + \Phi_3 M_{t-3} + H\xi_t, \qquad \varepsilon_t \sim \mathcal{N}(0, I) \tag{6.1}$$

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.

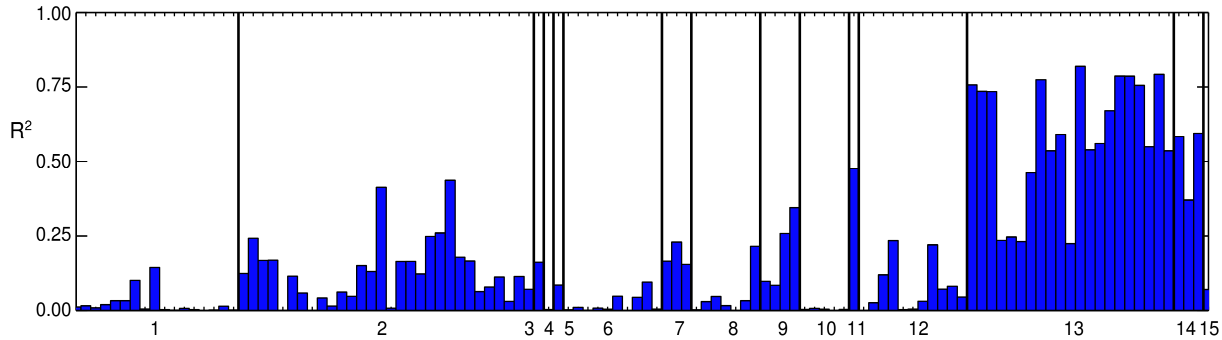
[8] Contemporaneous in the sense of same-month values for stock prices, exchange rates and the federal funds rate but one-month lagged values for the remaining macro series, see Section 2.2 for further details.

[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.* (2006b) 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.

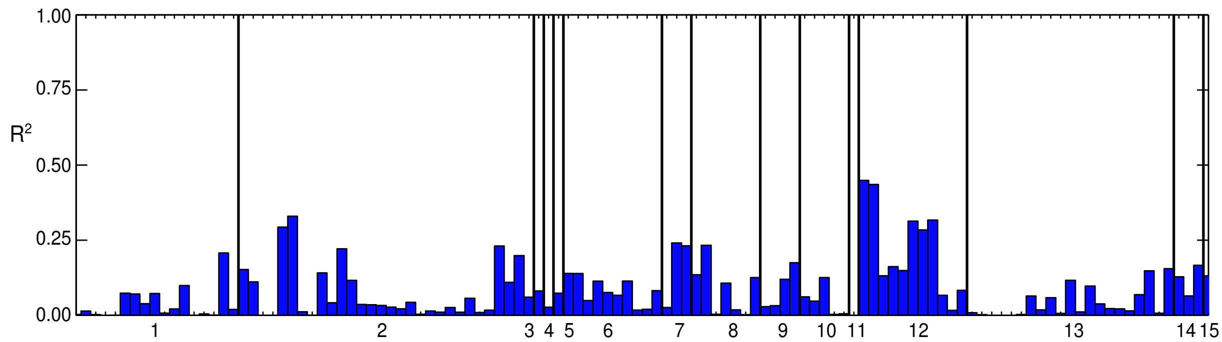
Figure 6.2: R^2 in regressions of individual macro series on PCA factors



(a) PCA factor #1



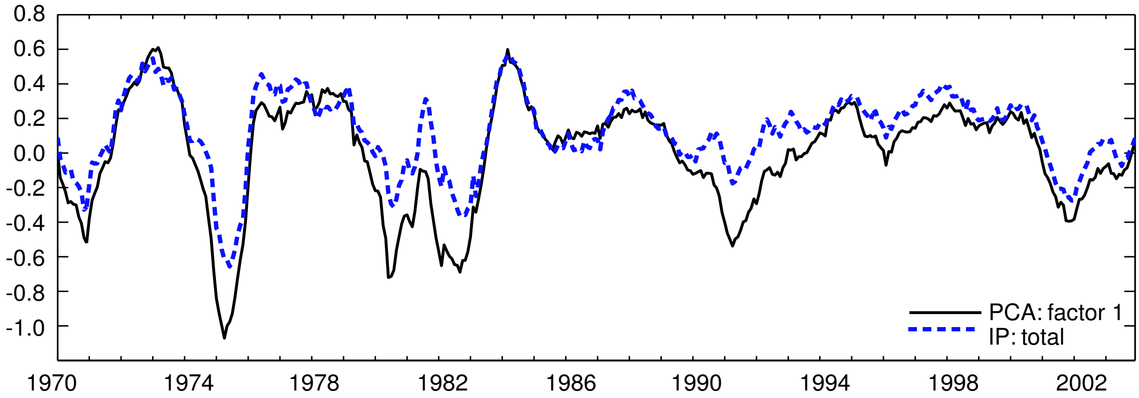
(b) PCA factor #2



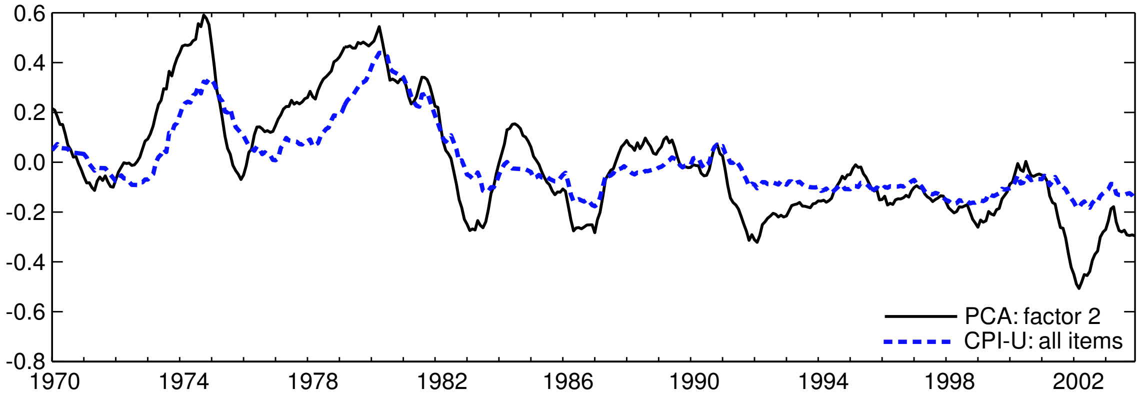
(c) PCA factor #3

Notes: 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.

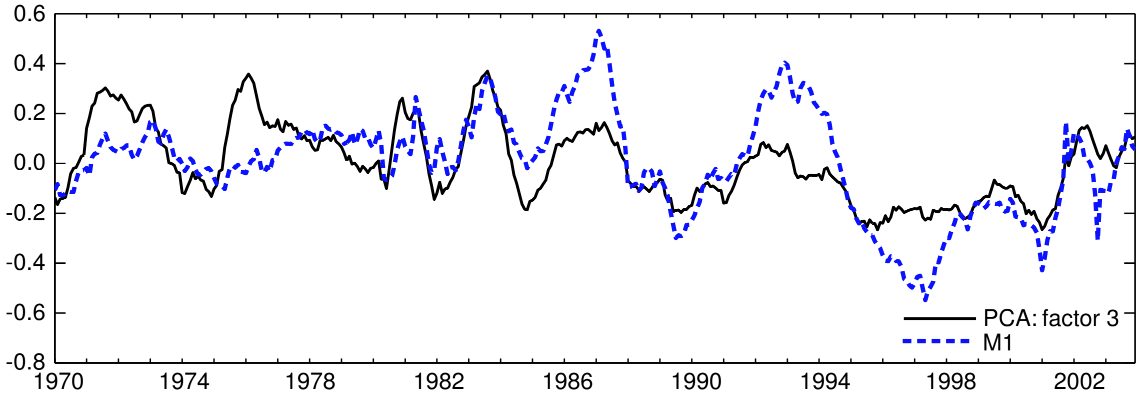
Figure 6.3: Macro factors compared to individual macro series



(a) PCA factor #1 - IP:total



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



(c) PCA factor #3 - M1

Notes: 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.

6.3.2 Models

Random walk

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

$$y_t^{(\tau_i)} = y_{t-1}^{(\tau_i)} + \sigma^{(\tau_i)}\varepsilon_t^{(\tau_i)}, \qquad \varepsilon_t^{(\tau_i)} \sim \mathcal{N}(0,1) \tag{6.2}$$

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 (2006a) 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 6.1 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

$$y_t^{(\tau_i)} = c^{(\tau_i)} + \phi^{(\tau_i)} y_{t-1}^{(\tau_i)} + \psi^{(\tau_i)\prime} X_t + \sigma^{(\tau_i)}\varepsilon_t^{(\tau_i)}, \qquad \varepsilon_t^{(\tau_i)} \sim \mathcal{N}(0,1) \tag{6.3}$$

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

[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).

first-order VAR specification[11],

$$Y_t = c + \Phi Y_{t-1} + \Psi X_t + H\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, I) \tag{6.4}$$

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 (6.4) accordingly with the (13×3) factor matrix F_{t-1}[12]:

$$Y_t = c + \Phi F_{t-1} + \Psi X_t + H\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, I) \tag{6.5}$$

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[13] The Nelson-Siegel model differs from the unrestricted VAR model in (6.5) by imposing a parametric

[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.

[13]De Pooter (2007) examines several extensions of the Nelson and Siegel (1987) three-factor model and shows that adding a second slope factor to the model improves its forecasting performance. However, here we only consider the original three-factor model as this is more commonly used.

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.* (2006b), 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

$$y_t^{(\tau_i)} = \beta_{1,t} + \beta_{2,t}\left[\frac{1-\exp(-\tau_i/\lambda)}{\tau_i/\lambda}\right] + \beta_{3,t}\left[\frac{1-\exp(-\tau_i/\lambda)}{\tau_i/\lambda} - \exp(-\tau_i/\lambda)\right] + \varepsilon_t^{(\tau_i)} \quad (6.6)$$

$$\beta_t = a + \Gamma\beta_{t-1} + u_t \quad (6.7)$$

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 (6.6) and (6.7) are normally distributed and mutually uncorrelated,

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0_{18\times1} \\ 0_{3\times1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix}\right) \quad (6.8)$$

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 ($\tau = 9$, 15, 18, 21 and 30 months) to the short end of the yield curve to estimate the Nelson-Siegel model in (6.6)-(6.8). 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 (6.7) 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.* (2006b) and involves jointly estimating (6.6)-(6.8) 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.* (2006b) 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:

$$
\begin{aligned}
y_t^{(\tau_i)} &= \beta_{1,t} + \beta_{2,t}\left[\frac{1-\exp(-\tau_i/\lambda)}{\tau_i/\lambda}\right] + \beta_{3,t}\left[\frac{1-\exp(-\tau_i/\lambda)}{\tau_i/\lambda} - \exp(-\tau_i/\lambda)\right] + \varepsilon_t^{(\tau_i)} && (6.9)\\
f_t &= a + \Gamma f_{t-1} + \eta_t && (6.10)\\
\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} 0_{18\times 1} \\ 0_{12\times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) && (6.11)
\end{aligned}
$$

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[14]. 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 chapter 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

$$Z_t = \mu + \Psi Z_{t-1} + u_t \tag{6.12}$$

[14]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 (6.9). Diebold *et al.* (2006b) 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

$$r_t = \delta_0 + \delta_1' Z_t \tag{6.13}$$

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

$$\lambda_t = \lambda_0 + \lambda_1 Z_t \tag{6.14}$$

with λ_0 a $(K \times 1)$ vector and λ_1 a $(K \times K)$ matrix[15]. Under the assumption that bond prices are an exponentially-affine function of the state variables,

$$P_t^{(\tau)} = \exp[A^{(\tau)} + B^{(\tau)\prime} Z_t] \tag{6.15}$$

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

$$P_t^{(\tau)} = \mathbb{E}_t[m_{t+1} P_{t+1}^{(\tau-1)}] \tag{6.16}$$

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 $B^{(\tau)}$:

$$A^{(\tau+1)} = A^{(\tau)} + B^{(\tau)\prime}[\mu - \Sigma\lambda_0] + \frac{1}{2}B^{(\tau)\prime}\Sigma\Sigma' B^{(\tau)} - \delta_0 \tag{6.17}$$
$$B^{(\tau+1)\prime} = B^{(\tau)\prime}[\Psi - \Sigma\lambda_1] - \delta_1' \tag{6.18}$$

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 $y_t^{(\tau)} = a^{(\tau)} + b^{(\tau)\prime} Z_t$ 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.

[15]Risk premia are constant over time if λ_1 equals zero. With λ_0 also equal to zero, risk premia are absent.

To summarize, we specify the following affine model

$$
\begin{aligned}
y_t^{(\tau_i)} &= a^{(\tau_i)} + b^{(\tau_i)} Z_t + \varepsilon_t^{(\tau_i)} && (6.19)\\
Z_t &= \mu + \Psi Z_{t-1} + u_t && (6.20)\\
\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} 0_{13\times 1} \\ 0_{3\times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) && (6.21)
\end{aligned}
$$

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

$$
\begin{aligned}
y_t^{(\tau_i)} &= a^{(\tau_i)} + b^{(\tau_i)} f_t + \varepsilon_t^{(\tau_i)} && (6.22)\\
f_t &= \mu + \Psi f_{t-1} + \eta_t && (6.23)\\
\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} 0_{13\times 1} \\ 0_{12\times 1} \end{bmatrix}, \begin{bmatrix} H & 0 \\ 0 & Q \end{bmatrix} \right) && (6.24)
\end{aligned}
$$

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 (6.23) 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[16]. 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.

[16]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.

6.4 Forecasting

6.4.1 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 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 6.1 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.

6.4.2 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)[17]. 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 (2006a), 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

[17]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.

error series with an average block-length of 12 months.

6.4.3 Forecasting results: individual models

Tables 6.3-6.6 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 6.7-6.10 for the period 1994:1-1999:12 and Tables 6.11-6.14 for the period 1999:1-2003:12.

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

Table 6.3: Forecast results for the sample 1994:1 - 2003:12, 1-month 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

Notes: 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 6.4: Forecast results for the sample 1994:1 - 2003:12, 3-month 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**

Notes: The table reports forecast results for a 3-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 6.3 for further details.

Table 6.5: Forecast results for the sample 1994:1 - 2003:12, 6-month 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	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**

Notes: The table reports forecast results for a 6-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 6.3 for further details.

Table 6.6: Forecast results for the sample 1994:1 - 2003:12, 12-month 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

Notes: The table reports forecast results for a 12-month horizon for the out-of-sample period 1994:1 - 2003:12. See Table 6.3 for further details.

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 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[18]. 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

[18] 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).

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 (2006a) 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 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 (2006a) 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 associated with period. Figure 6.1 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 (2006a) 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-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[19]. These studies suggest that there should be a high degree of predictability for this subperiod. Tables 6.7-6.10 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

[19] 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.

Table 6.7: Forecast results for the sample 1994:1 - 1998:12, 1-month 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

Notes: The table reports forecast results for a 1-month horizon for the out-of-sample period 1994:1 - 1998:12. See Table 6.3 for further details.

Table 6.8: Forecast results for the sample 1994:1 - 1998:12, 3-month 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

Notes: The table reports forecast results for a 3-month horizon for the out-of-sample period 1994:1 - 1998:12. See Table 6.3 for further details.

Table 6.9: Forecast results for the sample 1994:1 - 1998:12, 6-month 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

Notes: The table reports forecast results for a 6-month horizon for the out-of-sample period 1994:1 - 1998:12. See Table 6.3 for further details.

Table 6.10: Forecast results for the sample 1994:1 - 1998:12, 12-month 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**

Notes: The table reports forecast results for a 12-month horizon for the out-of-sample period 1994:1 - 1998:12. See Table 6.3 for further details.

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[20]. 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 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.

[20]Note that the forecast period of Diebold and Li (2006) does contains 24 months more.

Table 6.11: Forecast results for the sample 1999:1 - 2003:12, 1-month 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

Notes: The table reports forecast results for a 1-month horizon for the out-of-sample period 1999:1 - 2003:12. See Table 6.3 for further details.

Table 6.12: Forecast results for the sample 1999:1 - 2003:12, 3-month 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**

Notes: The table reports forecast results for a 3-month horizon for the out-of-sample period 1999:1 - 2003:12. See Table 6.3 for further details.

Table 6.13: Forecast results for the sample 1999:1 - 2003:12, 6-month 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**

Notes: The table reports forecast results for a 6-month horizon for the out-of-sample period 1999:1 - 2003:12. See Table 6.3 for further details.

Table 6.14: Forecast results for the sample 1999:1 - 2003:12, 12-month 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

Notes: The table reports forecast results for a 12-month horizon for the out-of-sample period 1999:1 - 2003:12. See Table 6.3 for further details.

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 6.11 - 6.14. 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.

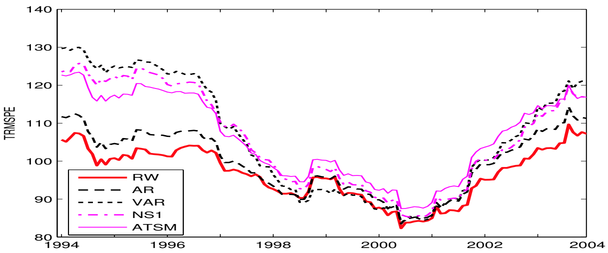
Rolling TRMSPE

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 6.4-6.7 show results for all forecast horizons considered and for a subset of models[21]. 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)

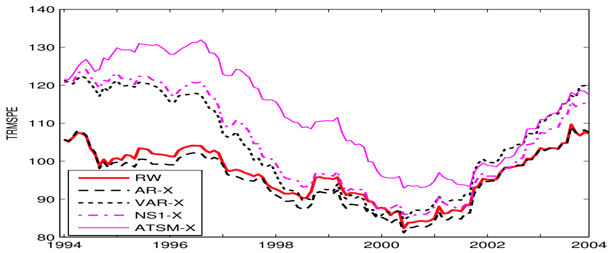
The patterns for the two five year subsamples reappear. TRMSPEs are fairly stable until 1997 after which a decreasing trend sets in lasting until mid 2000. The high degree of interest rate predictability during the 1994-1998 subperiod is the cause of the decreasingly low TRMSPEs for the period 1998-2001. 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

[21]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.

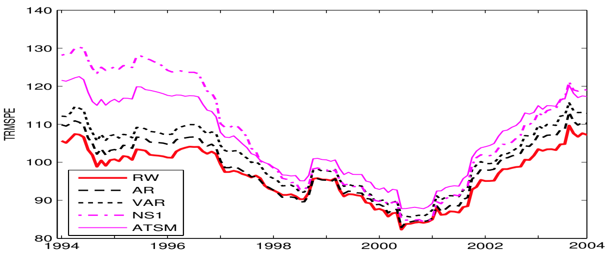
Figure 6.4: 60-month moving TRMSPE: 1-month horizon



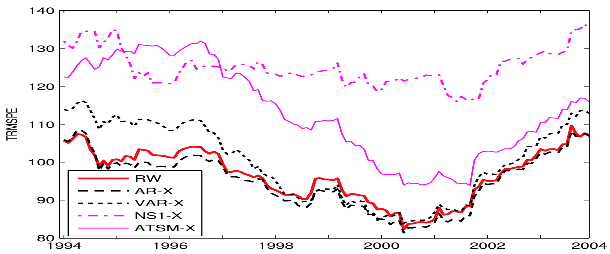
(a) classical inference



(b) classical inference

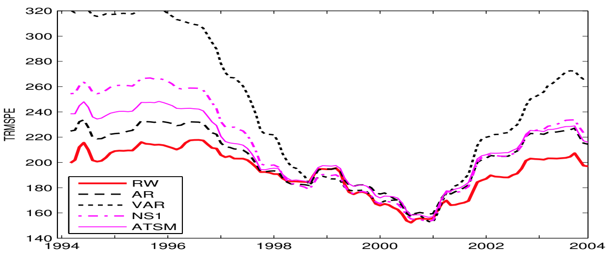


(c) Bayesian inference

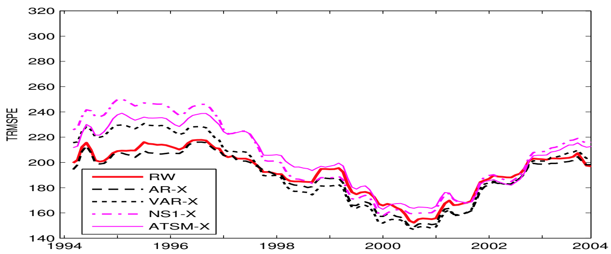


(d) Bayesian inference

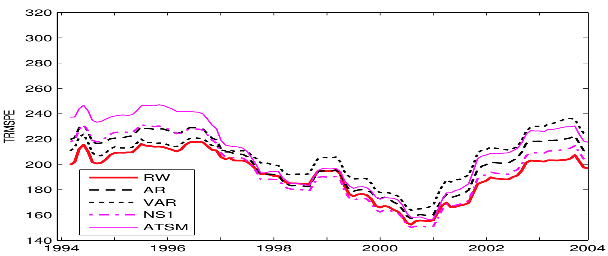
Figure 6.5: 60-month moving TRMSPE: 3-month horizon



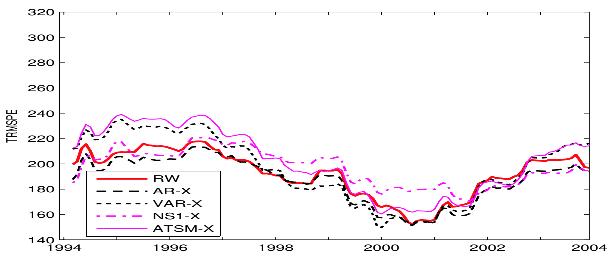
(a) classical inference



(b) classical inference



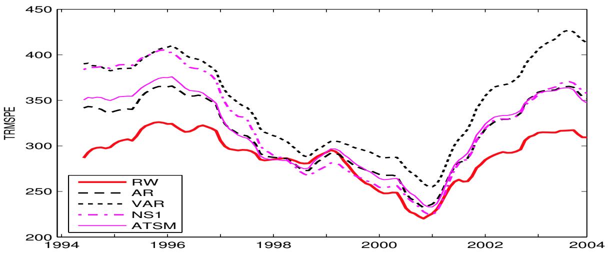
(c) Bayesian inference



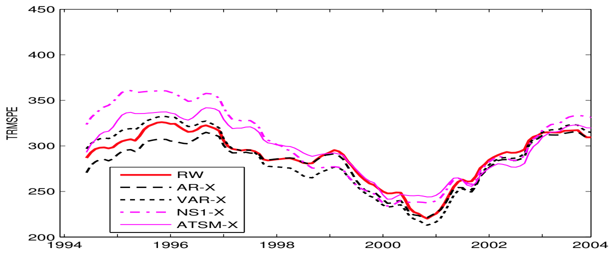
(d) Bayesian inference

Notes: 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 6.4 and a 3-month horizon in Figure 6.5. 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.

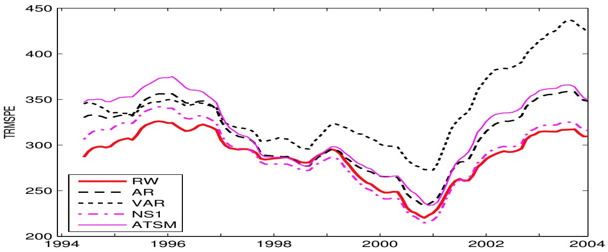
Figure 6.6: 60-month moving TRMSPE: 6-month horizon



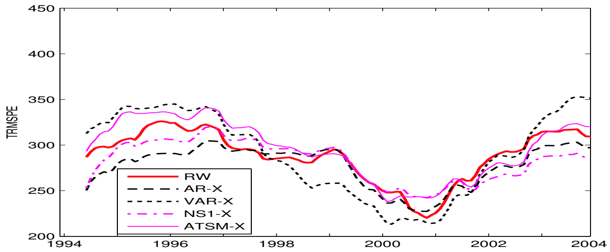
(a) classical inference



(b) classical inference

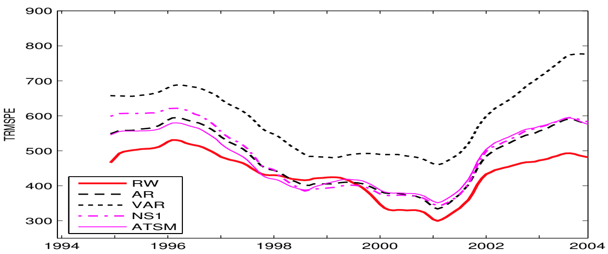


(c) Bayesian inference

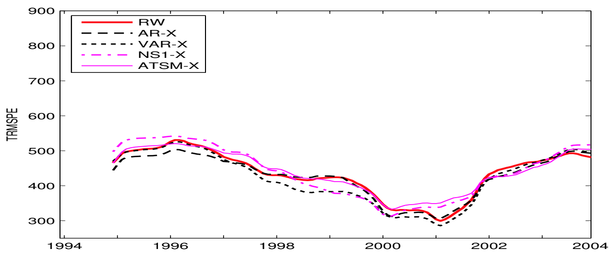


(d) Bayesian inference

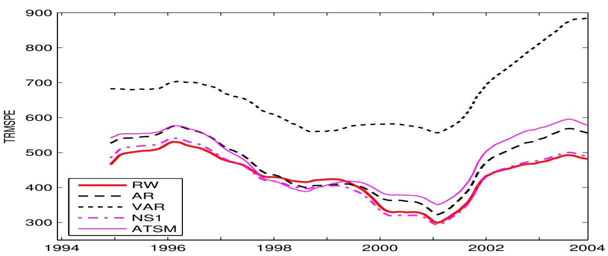
Figure 6.7: 60-month moving TRMSPE: 12-month horizon



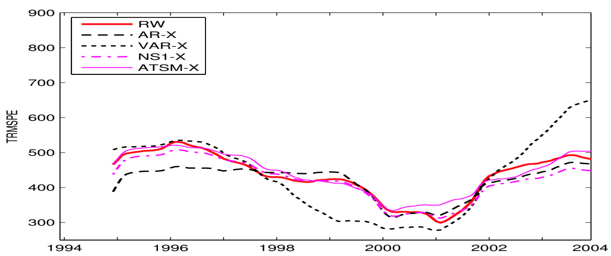
(a) classical inference



(b) classical inference



(c) Bayesian inference



(d) Bayesian inference

Notes: 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 6.6 and a 12-month horizon in Figure 6.7. 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.

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.

6.5 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 portfolio construction. 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.

6.5.1 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**)[22]. The weight for model m is computed as $w^{(\tau_i)}_{T+h,m} = \frac{1/\text{MSPE}^{(\tau_i)}_{T+h,m}}{\sum_{m=1}^{M}(1/\text{MSPE}^{(\tau_i)}_{T+h,m})}$ where $\text{MSPE}^{(\tau_i)}_{T+h,m} = \frac{1}{\upsilon}\sum_{r=1}^{\upsilon}(\widehat{y}^{(\tau_i)}_{T+h-r|T-r,m} - y^{(\tau_i)}_{T+h-r})^2$. 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[23]. 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**[24], and is based on the predictive likelihood approach 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

[22]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.

[23]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.

[24]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.

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) 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.

6.5.2 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)[25]. 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 6.3-6.6. 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 provide complementary information definitely seems to hold for longer maturities.

[25] 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.

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 6.7-6.10. 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.

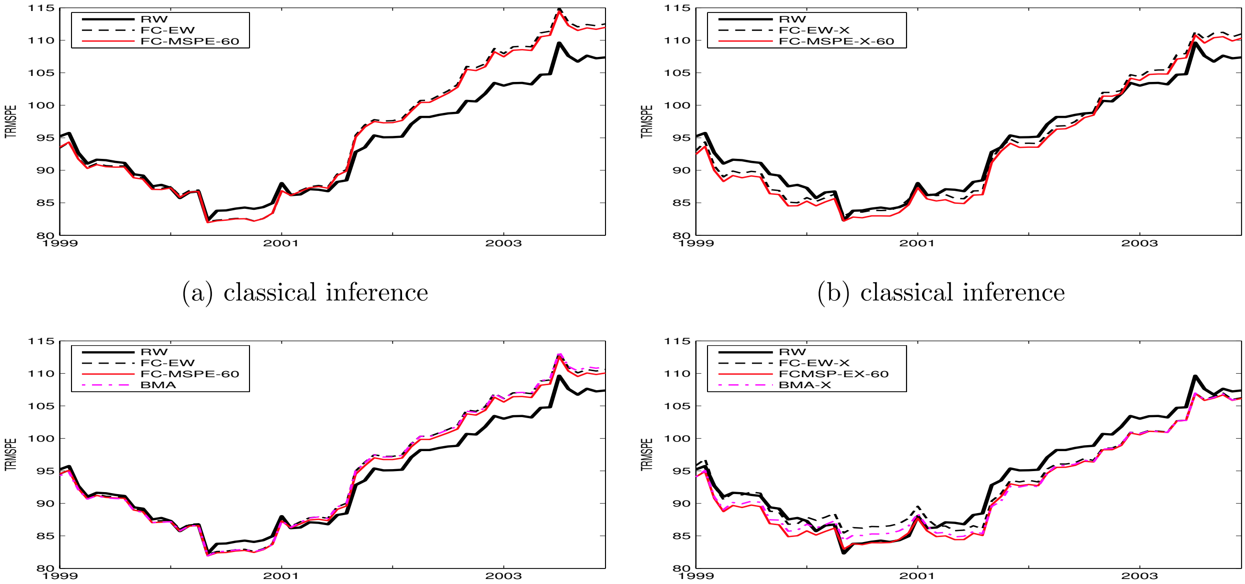
For the 6-month and 12-month horizons, forecast combinations with macroeconomic information, based on the MSPE-weights using a long historical window are the most accurate forecasting methods. It is interesting that whereas for the individual models (except for the VAR model) adding macro factors worsens forecasting performance, for the forecast combination methods adding macro factors is very beneficial.

Panel E of Table 6.9 and 6.10 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. More accurate combination schemes would

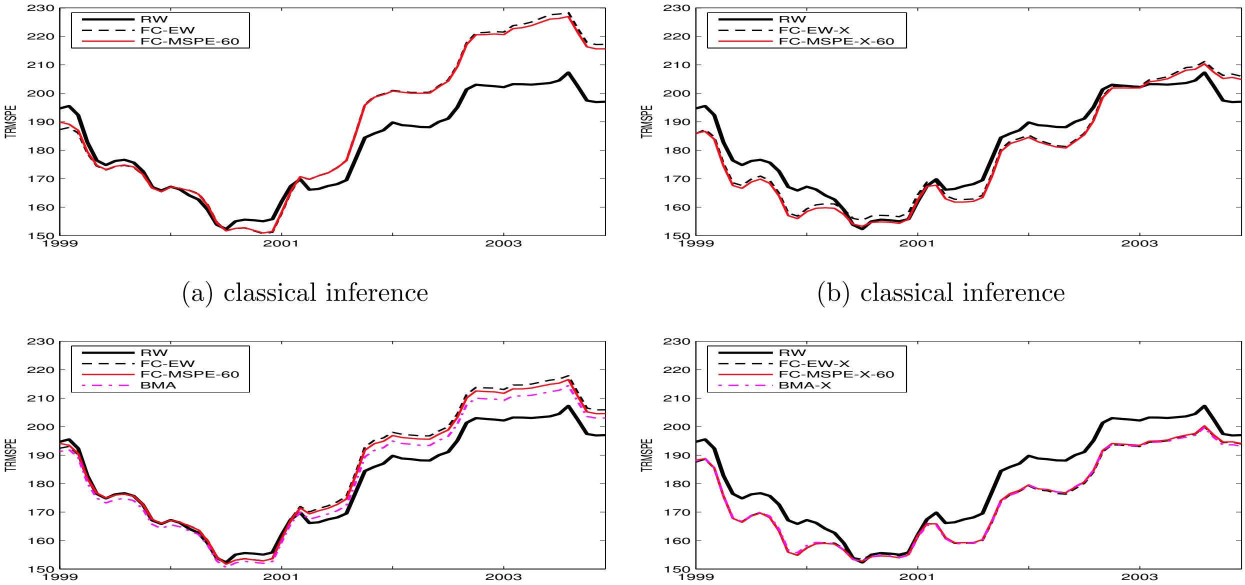
Figure 6.8: 60-month moving TRMSPE: 1-month horizon



(a) classical inference (b) classical inference

(c) Bayesian inference (d) Bayesian inference

Figure 6.9: 60-month moving TRMSPE: 3-month horizon

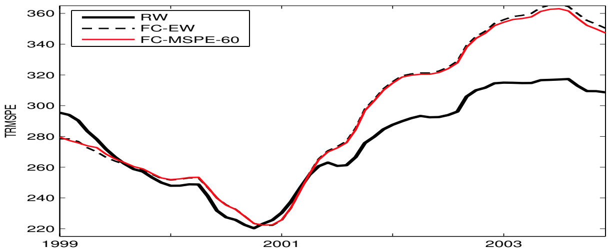


(a) classical inference (b) classical inference

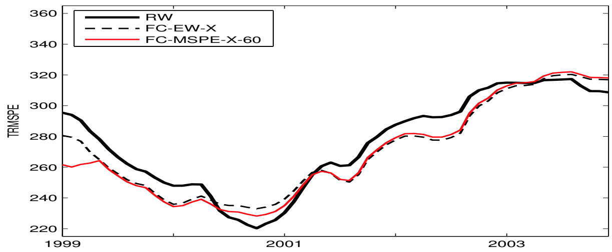
(c) Bayesian inference (d) Bayesian inference

Notes: 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 6.8 and a 3-month horizon in Figure 6.9. 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 whereas otherwise only individual models without macro factors are considered in the forecast combination approaches.

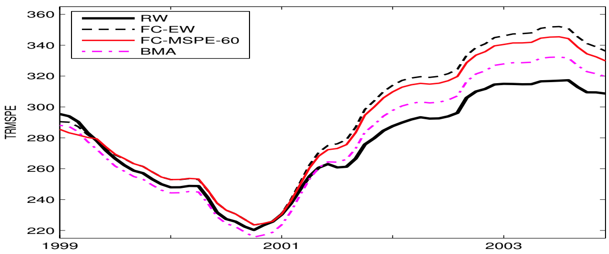
Figure 6.10: 60-month moving TRMSPE: 6-month horizon



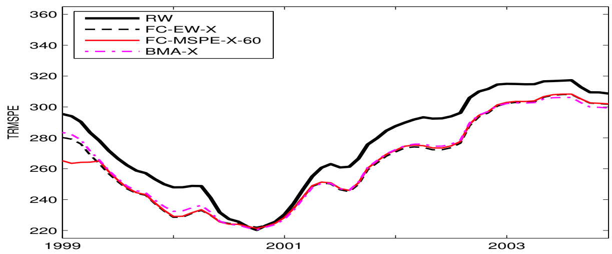
(a) classical inference



(b) classical inference

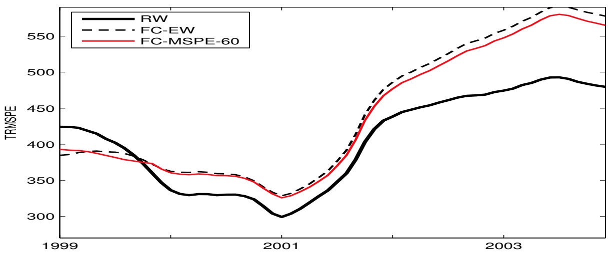


(c) Bayesian inference

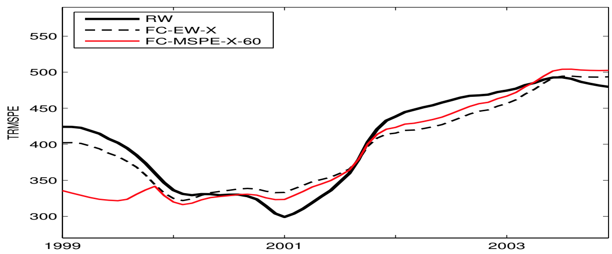


(d) Bayesian inference

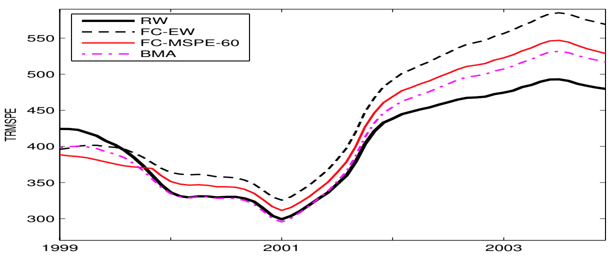
Figure 6.11: 60-month moving TRMSPE: 12-month horizon



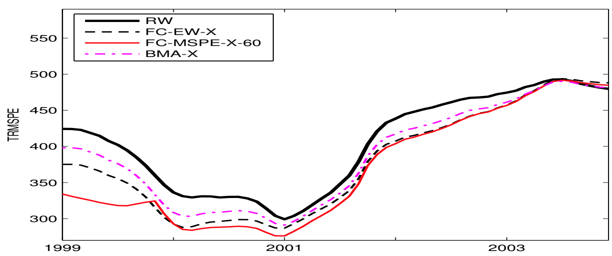
(a) classical inference



(b) classical inference



(c) Bayesian inference



(d) Bayesian inference

Notes: 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 6.10 and a 12-month horizon in Figure 6.11. 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 whereas otherwise only individual models without macro factors are considered in the forecast combination approaches.

ideally be able to account for structural changes[26].

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 6.8-6.10. These graphs show the 60-month rolling TRMSPE for the equally-weighted and MSPE-weighted combination schemes for the period 1999-2003[27], 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[28]. 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 6.9-6.11 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.

6.6 Concluding remarks

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

[26]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 $= \frac{1}{v}\sum_{r=1}^{v}\lambda^{r-1}(\widehat{y}^{(\tau_i)}_{T+h-r|T-r} - y^{(\tau_i)}_{T+h-r})^2$. 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.

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

[28]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.

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 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 provides strong support for the use of 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 chapter 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

6A 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.

6A.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

$$\widehat{y}^{(\tau_i)}_{T+h} = \hat{c}^{(\tau_i)} + \hat{\phi}^{(\tau_i)} \widehat{y}^{(\tau_i)}_{T+h-1} + \hat{\psi}^{(\tau_i)\prime} \widehat{X}_{T+h} \tag{6A.1}$$

with $\widehat{y}^{(\tau_i)}_{T} = y^{(\tau_i)}_{T}$. 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 $(c^{(\tau_i)}, \phi^{(\tau_i)}, \psi^{(\tau_i)}, \sigma^{2(\tau_i)})$,

$$(c^{(\tau_i)}, \phi^{(\tau_i)}, \psi^{(\tau_i)}, \sigma^{2(\tau_i)})' \sim NG(\underline{b}, \underline{v}, \underline{s}^2, \underline{\nu}) \tag{6A.2}$$

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

6A.2 VAR model

Frequentist Inference

We estimate the equation parameters (c, Φ, Ψ) in (6.5) using equation-by-equation OLS. Forecasts are obtained as

$$\widehat{Y}_{T+h} = \widehat{c} + \widehat{\Phi}\widehat{F}_{T+h-1} + \widehat{\Psi}\widehat{X}_{T+h} \tag{6A.3}$$

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[29].

[29] 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 (6.5). For computational tractability we select the following conjugate priors:

$$\Pi|S \sim MN(\underline{B}, S \otimes \underline{V}) \tag{6A.4}$$

and

$$S \sim IW(\underline{S}, \underline{\mu}) \tag{6A.5}$$

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

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

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

$$\begin{aligned} &p(\Pi, S|Y_T, F_{T-1}, X_T) = p(Y_T|F_{T-1}, X_T, \Pi, S)p(\Pi|S)p(S) \\ &\propto |S|^{-(T+N+\nu+1)/2}\exp(-\tfrac{1}{2}tr(S^{-1}(\underline{S} + (Y_T - Z_T\Pi)'(Y_T - Z_T\Pi) + (\Pi - \underline{B})'\underline{V}^{-1}(\Pi - \underline{B})))) \end{aligned} \tag{6A.7}$$

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,

$$p(\Pi|Y_T, F_{T-1}, X_T) \propto |\underline{S} + (W_T - V_T\widehat{\Pi})'(W_T - V_T\widehat{\Pi}) + (\Pi - \widehat{\Pi})'(Z'_T Z_T + \underline{V}^{-1})(\Pi - \widehat{\Pi})|^{-(T+\nu)/2} \tag{6A.8}$$

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

$$S|Y_T, F_{T-1}, X_T \sim IW(\underline{S} + (W_T - V_T\widehat{\Pi})'(W_T - V_T\widehat{\Pi}), T + \nu) \tag{6A.9}$$

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

$$\begin{aligned} &p(Y_{T+h}|Y_T, X_T, F_{T+h-1}, X_{T+h}) = \int\int p(Y_{T+h}, \Pi, S|Y_T, X_T, F_{T+h-1}, X_{T+h})d\Pi dS \\ &= \int\int p(Y_{T+h}|F_{T+h-1}, X_{T+h}, \Pi, S)p(\Pi, S|Y_T, X_T)d\Pi dS \end{aligned} \tag{6A.10}$$

By applying the inverted Wishart step to (6A.10), and integrating with respect to Π, we have:

$$\begin{aligned} &p(Y_{T+h}|Y_T, X_T, F_{T+h-1}, X_{T+h}) \propto [\underline{S} + (W_T - V_T\widehat{\Pi})'(W_T - V_T\widehat{\Pi}) + \\ &(Y_{T+h} - Z_{T+h}\widehat{\Pi})'(I - Z_{T+h}L^{-1}Z'_{T+h})(Y_{T+h} - Z_{T+h}\widehat{\Pi})]^{-(T+\nu+h)/2} \end{aligned} \tag{6A.11}$$

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 (6A.11) as:

$$p(Y_{T+h}|Y_T, X_T, F_{T+h-1}, X_{T+h}) = p(Y_{T+1}|F_T, X_{T+1}) \times \ldots \times p(Y_{T+h}|F_T, X_{T+1}, ..., F_{T+H-1}, X_{T+h}) \tag{6A.12}$$

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 (6A.12) and we apply direct simulation to draw the predictive density of Y_{T+h}, conditional on Y_T and X_T,

$$p(Y_{T+h}|Y_T, X_T) = \iint p(Y_{T+h}|Y_T, X_T, F_{T+h-1}, X_{T+h})p(F_{T+h-1}|F_{T+h-2})p(X_{T+h}|X_{T+h-1})dF_{T+h-1}dX_{T+h} \tag{6A.13}$$

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

6A.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.* (2006b).

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 (6.7) 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 (6.6)-(6.7). 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

$$\widehat{f}_{T+h} = \widehat{a} + \widehat{\Gamma}\widehat{f}_{T+h-1} \tag{6A.14}$$

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:

$$\widehat{y}_{T+h}^{(\tau_i)} = \widehat{\beta}_{1,T+h} + \widehat{\beta}_{2,T+h}\left(\frac{1-\exp(-\tau_i/\widehat{\lambda})}{\tau_i/\widehat{\lambda}}\right) + \widehat{\beta}_{3,T+h}\left(\frac{1-\exp(-\tau_i/\widehat{\lambda})}{\tau_i/\widehat{\lambda}} - \exp(-\tau_i/\widehat{\lambda})\right) \tag{6A.15}$$

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

$$\sigma^{2^{(\tau_i)}} \sim \text{IG-2}(\underline{\nu}^{(\tau_i)}, \underline{\delta}^{(\tau_i)}) \tag{6A.16}$$

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

$$Q_1 \sim \mathrm{IW}(\underline{\mu}_1, \underline{\Delta}_1) \tag{6A.17}$$

$$Q_2 \sim \mathrm{IW}(\underline{\mu}_2, \underline{\Delta}_2) \tag{6A.18}$$

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

$$[a, \Gamma] \sim MN(\underline{\Gamma}, Q \otimes \underline{V}_\Gamma) \tag{6A.19}$$

Finally for λ we assume a uniform distribution,

$$\lambda \sim U(\underline{a}_\lambda, \underline{b}_\lambda) \tag{6A.20}$$

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

$$p(Y_T, F_T|\theta) = \prod_{t=1}^{T}\prod_{i=1}^{18} p(y_t^{(\tau_i)}|f_t, \lambda, \sigma^{2(\tau_i)})p(f_t|f_{t-1}, a, \Gamma, Q) \tag{6A.21}$$

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 (6.6)–(6.7). When we combine (6A.21) with the prior densities $p(\theta)$ in (6A.16)–(6A.20) we obtain the posterior density

$$p(\theta, B_T|Y_T, M_T, M_{T-1}, M_{T-2}) \propto p(Y_T, F_T|\theta)p(\theta) \tag{6A.22}$$

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:

$$p(\lambda|Y_T, F_T, H) \propto \prod_{t=1}^{T}\prod_{i=1}^{N} p(y_t^{(\tau_i)}|f_t, \lambda, \sigma^{2(\tau_i)})p(\lambda) \tag{6A.23}$$

Equation (6A.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[30]. More formally and referring to equation (6A.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.

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

$$p(Y_{T+h}|Y_T, F_T) = \iint p(y^{(\tau_i)}_{T+h}|f_{T+h}, \lambda, \sigma^{2^{(\tau_i)}})p(f_{T+h}|f_{T+h-1}, a, \Gamma, Q)\times$$
$$p(\theta, B_T|Y_T, M_T, M_{T-1}, M_{T-2})df_{T+h}\, d\theta \quad (6A.24)$$

where $p(y^{(\tau_i)}_{T+h}|f_{T+h}, \lambda, \sigma^{2^{(\tau_i)}})$ 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 (6A.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 $y^{(\tau_i)}_{T+h}$.

6A.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 (6.23). 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 (6.13) 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 (6.22), 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

$$\widehat{f}_{T+h} = \widehat{\mu} + \widehat{\Psi}\widehat{f}_{T+h-1} \quad (6A.25)$$

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

$$\widehat{y}^{(\tau_i)}_{T+h} = \widehat{a}^{(\tau_i)} + \widehat{b}^{(\tau_i)}\widehat{f}_{T+h} \quad (6A.26)$$

[30]Mönch (2006b) 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.

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

$$p(y_{T+h}^{(\tau_i)}|Y_T, F_T) = \iint p(y_{T+h}^{(\tau_i)}|f_{T+h}, a^{(\tau_i)}, b^{(\tau_i)}, \sigma^{2\,(\tau_i)})p(f_{T+h}|f_{T+h-1}, \mu, \Psi, Q)p(\theta|Y_T, F_T)df_{T+h}\, d\theta \tag{6A.27}$$

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 (6A.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 $(Y_T,\ F_T)$:

$$p(y_{T+h}^{(\tau_i)}|Y_T, X_T) = \iint p(y_{T+h}^{(\tau_i)}|f_{T+h}, \widehat{a}^{(\tau_i)}, \widehat{b}^{(\tau_i)}, \widehat{\sigma^2}^{(\tau_i)})p(f_{T+h}|f_{T+h-1}, \widehat{\mu}, \widehat{\Psi}, \widehat{Q})df_{T+h}d\theta \tag{6A.28}$$

6B 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

$$p(y_{T+h}^{(\tau_i)}|Y_T, D_T) = \sum_{i=1}^{M} P(m_j^{(\tau_i)}|Y_T, D_T)p(y_{T+h}^{(\tau_i)}|Y_T, D_T, m_j^{(\tau_i)}) \tag{6B.1}$$

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:

$$P(m_j^{(\tau_i)}|Y_T, D_T) = \frac{p(y_{T,o}^{(\tau_i)}|Y_T, D_T, m_j^{(\tau_i)})P(m_j^{(\tau_i)})}{\sum_{s=1}^{k} p(y_{T,o}^{(\tau_i)}|Y_T, D_T, m_s^{(\tau_i)})P(m_s^{(\tau_i)})} \tag{6B.2}$$

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.

6C 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 (6A.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 (6A.4)–(6A.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:

$$\lambda \sim U(3.34, 33.45) \tag{6C.1}$$

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, $P(m_j^{(\tau_i)}) = 1/M$.

Nederlandse Samenvatting

(Summary in Dutch)

Inleiding

Dit proefschrift bestaat uit een bundel van studies op het gebied van twee belangrijke onderwerpen binnen de kwantitatieve financiële analyse: de volatiliteit van rendementen op financiële waarden en de rente termijn structuur. Alvorens een meer gedetailleerde beschrijving van de inhoud en contributies van de verschillende hoofdstukken te geven wordt eerst een beknopte introductie gegeven van beide onderwerpen.

Onzekerheid, of 'volatiliteit', is één van de meest cruciale ingrediënten in vele gebieden binnen de financiële wetenschap. Voorbeelden van gebieden waarbij volatiliteit een belangrijke rol speelt zijn bijvoorbeeld het construeren van aandelenportefeuilles, risicomanagement en het prijzen van derivaten. Een ietwat losse definitie van volatiliteit kan gegeven worden als zijnde de beweeglijkheid van de stochastische, of onvoorspelbare, component van een variabele die varieert door de tijd heen. Een dergelijke variabele komt in de financiële wereld meestal neer op het rendement van een financiële waarde zoals een individueel aandeel of een samengestelde index zoals de Standard & Poor's 500 Index. Volatiliteit is, in tegenstelling tot rendementen, niet rechtstreeks waarneembaar en moet derhalve geschat worden. Het is daarom noodzakelijk om schattingen van de volatiliteit te construeren, allereerst met als doel om volatiliteit te meten en om deze vervolgens daarna te modelleren.

Een nog steeds erg populaire aanpak om volatiliteit te meten is door middel van het gebruik van gekwadrateerde dagrendementen. Ondanks het feit dat het kwadrateerde rendement een inefficiënte *ex-post* maatstaf is voor de daadwerkelijke volatiliteit geeft het wel duidelijk de meest belangrijke eigenschap van volatiliteit weer: namelijk dat zij niet constant is door de tijd heen maar dat zij tijdsvariërend is. Een tweede, hieraan gerelateerde eigenschap, is dat volatiliteit een grote mate van persistentie vertoont: volatiliteit in het verleden verklaart volatiliteit in het heden. De mate van persistentie is belangrijk vanuit economisch oogpunt omdat het aangeeft of schokken in volatiliteit oftewel permanent dan wel tijdelijk zijn hetgeen bijvoorbeeld de hoogte van risico premies zal beïnvloeden. De aanwezigheid van persistentie suggereert eveneens dat, in tegenstelling tot de rendementen welke erg moeilijk voorspelbaar zijn, volatiliteit tot op zekere hoogte veel beter voorspelbaar is.

Een in de praktijk veelgebruikt model voor het modelleren van de conditionele variantie van tijdreeksen is het Generalized Conditional Heteroscedasticity (GARCH) model van Engle (1982) en Bollerslev (1986). Dit model legt een parametrische Autoregressive Moving Average (ARMA) structuur op voor de gekwadrateerde rendementen. Een nadeel van het GARCH model is dat het impliceert dat schokken in het rendement een exponentieel uitdovend effect hebben op de volatiliteit. Onderzoek heeft aangetoond dat de persistentie van schokken mogelijk beter beschreven kan worden door middel van een zogeheten "long-memory" proces waarbij het effect van schokken uitdooft met een langzamer hyperbolisch tempo.

Het is echter nog onduidelijk of volatiliteit daadwerkelijk de long-memory eigenschap bezit. Alhoewel het modelleren hiervan wel degelijk mogelijk is, is het vinden van een bevredigend antwoord op de vraag waarom schokken een dergelijk langdurend effect op volatiliteit hebben, tot nu toe moeilijk gebleken. Een mogelijk alternatief antwoord wordt aangedragen door Granger en Hyung (1999) en Diebold en Inoue (2001). Deze auteurs tonen aan dat tijdreeksen die structurele breuken in het niveau van de volatiliteit vertonen, soortgelijke eigenschappen bevatten als tijdreeksen die daadwerkelijk long-memory zijn. Het gevolg hiervan is dat structurele breuken in de volatiliteit mogelijk long-memory-achtige kenmerken in tijdreeksen kunnen introduceren. Het is daarom van belang om bij het modelleren van volatiliteit te toetsen op de aanwezigheid van structurele breuken en om deze op te nemen in volatiliteitsmodellen.

Ondanks hun populariteit werd er lange tijd aangenomen dat volatiliteitsmodellen die gebaseerd zijn op gekwadrateerde rendementen teleurstellende voorspellingen voor de toekomstige volatiliteit opleveren. Andersen en Bollerslev (1998) laten echter zien dat dit het gevolg is van het gebruik van een onnauwkeurige schatting van de daadwerkelijke *ex-post* volatiliteit. Merton (1980) toonde eerder al aan dat het gebruik van intradag rendementsobservaties theoretisch leidt tot een schatting van de volatiliteit die geen onnauwkeurigheid meer bevat. Andersen en Bollerslev (1998) doen als gevolg hiervan de aanbeveling om een maatstaf voor de dagelijkse volatiliteit te construeren die gebaseerd is op de som van gekwadrateerde intradag rendementen. Zij laten zien dat wanneer de voorspellingen van GARCH modellen geëvalueerd worden met deze nieuwe maatstaf, "gerealiseerde volatiliteit" genaamd, de voorspelkracht van dergelijke modellen veel groter blijkt te zijn.

Gerealiseerde volatiliteit kan niet alleen als maatstaf gebruikt worden om voorspellingen van volatiliteitsmodellen te evalueren, maar kan tevens gebruikt worden om volatiliteit rechtstreeks te modelleren. Andersen, Bollerslev, Diebold en Labys (2003) laten zien dat gerealiseerde volatiliteit, naarmate deze gebaseerd is op intradag rendementen die met een steeds hoger wordende frequentie gemeten worden, een maatstaf is voor volatiliteit die vrijwel geen meetfout meer bevat. Het gevolg hiervan is dat volatiliteit als "observeerbaar" beschouwd kan worden en dat zij dus nu rechtstreeks gemodelleerd kan worden met behulp van standaard technieken voor het modelleren van financiële tijdreeksen. Alhoewel de nadruk in de literatuur vooral ligt op het modelleren van de volatiliteit van individuele tijdreeksen is het tevens mogelijk om de covarianties tussen verschillende tijdreeksen eveneens rechtstreeks te modelleren. Het gebruik van traditionele modellen zoals het multivariate GARCH model werd bemoeilijkt door het grote aantal te schatten parameters in dergelijke modellen. Door het gebruik van gerealiseerde covarianties is het nu echter veel beter mogelijk om een gehele covariantie matrix te modelleren.

De literatuur op het gebied van gerealiseerde volatiliteit heeft zich in het afgelopen decennium sterk ontwikkeld. Echter, een aantal aandachtspunten betreffende het gebruik van gerealiseerde volatiliteitsmaatstaven wordt op dit moment nog grondig onderzocht. Een van de belangrijkste vragen komt naar voren wanneer de theorie toegepast wordt in de praktijk. De theorie impliceert dat de hoogst mogelijke intradag frequentie gebruikt dient te worden om een zo accuraat mogelijk gerealiseerde volatiliteitsschatter te construeren. Echter, het is bekend dat wanneer in de praktijk rendementen waargenomen worden met een hoger wordende frequentie deze in toenemende mate ruis zullen bevatten. Dit is bijvoorbeeld het gevolg van het niet synchroon handelen van verschillende aandelen. Er dient dus een afweging gemaakt te worden tussen nauwkeurigheid enerzijds, hetgeen impliceert dat rendementen bij een zo hoog mogelijke frequentie waargenomen dienen te worden en ruis anderzijds, hetgeen impliceert dat het gebruik van een lagere frequentie mogelijkerwijs een betere volatiliteitsschatter oplevert. Een belangrijke vraag die beantwoord moet worden is daarom ook welke frequentie het beste gebruikt kan worden voor het construeren van gerealiseerde variantie en covariantieschatters.

Het eerste deel van dit proefschrift levert bijdragen aan de literatuur op het gebied van het toetsen op structurele breuken in het niveau van de volatiliteit, het modelleren van gerealiseerde volatiliteit en op het gebied van het kiezen van de optimale intradag frequentie voor het construeren van gerealiseerde volatiliteitsmaatstaven.

In het tweede deel van dit proefschrift ligt de nadruk op het analyseren van de rentetermijnstructuur. De rentetermijnstructuur, of ook wel rentecurve genoemd, geeft de relatie weer tussen rentestanden die gelden voor verschillende looptijden. De rentecurve bepaalt de huidige waarde van toekomstige inkomsten en dient daarom als richtlijn voor het nemen van economische besluiten. De noodzaak om de rentetermijnstructuur te bestuderen wordt verder duidelijk wanneer men zich realiseert dat lange-termijn rentestanden, na een correctie voor risico, opgebouwd zijn uit verwachtingen voor toekomstige korte-termijn rentes. Hierdoor bevatten lange-termijn rentes informatie over korte-termijn rentes in de toekomst. Tevens bevat de rentecurve meer in zijn algemeenheid informatie over de toekomstige stand van de economie. De helling van de rentecurve (welke het verschil tussen lange- en korte-termijn rentestanden weergeeft) is met succes gebruikt voor het voorspellen van de groei van het Bruto Binnenlands Product en voor het voorspellen van economische recessies. Macro-economen richten zich derhalve meer en meer op het trachten te doorgronden van de verbanden tussen rentestanden, monetair beleid en macro-economische indicatoren.

Samenvattend is het duidelijk dat de rentetermijnstructuur van cruciaal belang is bij het waarderen van obligaties, het meten en beheren van renterisicos en monetair beleid. Het is daarom van groot belang dat de rentetermijnstructuur nauwkeurig geschat kan worden, daar deze niet eenduidig waarneembaar is. Als mede is het van belang dat zij accuraat voorspeld kan worden. Dit proefschrift richt zich met name op het tweede punt: het voorspellen van de toekomstige rentetermijnstructuur.

Het voorspellen van de rentecurve waarbij alleen gebruik gemaakt wordt van historische rente-informatie is erg moeilijk doordat eenvoudige modellen, met name het 'random walk' model betere voorspellingen genereren dan meer complexe modellen (zie bijv Duffee, 2002). Recente studies laten echter zien dat de rentecurve tot op zekere hoogte wel degelijk voorspelbaar is, vooral door het toevoegen van macro-economische variabelen aan modellen voor de rentetermijnstructuur. Echter, het kunnen identificeren van een enkel model dat in staat is om consistent toekomstige rentes accuraat te voorspellen blijft een lastige taak.

Het tweede deel van dit proefschrift draagt allereerst bij aan de literatuur door het verder onderzoeken van één klasse van rentemodellen in het bijzonder. Vervolgens wordt de voorspelkwaliteit van een veelvoud aan termijnstructuurmodellen onderzocht met als doel het vaststellen van de stabiliteit van de voorspelkracht van deze modellen. Tevens worden technieken geanalyseerd voor het combineren van voorspellingen van verschillende modellen.

Resultaten

Dit proefschrift is opgebouwd uit twee delen. De hoofdstukken 2, 3, en 4 vormen samen deel A en richten zich op het modelleren van de volatiliteit van financiële tijdreeksen. In hoofdstuk 3 en 4 ligt de nadruk vooral op het gebruik van hoge-frequentie intradag rendementen voor het analyseren van volatiliteiten en covarianties. Deel B bestaat uit de hoofdstukken 5 en 6 en de analyses die hierin uitgevoerd worden richten zich op het schatten en voorspellen van de rentetermijnstructuur.

Deel A: Modelleren en voorspellen van de volatiliteit van aandelen rendementen

Aangetoond is dat de volatiliteit van financiële tijdreeksen sporadische niveauveranderingen ondergaat. Als deze structurele breuken buiten beschouwing gelaten worden bij het modelleren van de volatiliteit kan dit leiden tot een verkeerde inschatting van de persistentie in volatiliteit. Het is daarom van belang dat geïdentificeerd kan worden wanneer breuken hebben plaatsgevonden. Hoofdstuk 2 beschouwt Cumulative Sums of Squares (CUSUM) toetsen voor het identificeren van structurele breuken in het gemiddelde niveau van de volatiliteit van tijdreeksen die conditionele heteroscedasticiteit vertonen. Een belangrijke conclusie van het onderzoek in dit hoofdstuk is dat het toepassen van dergelijke toetsen rechtstreeks op rendement reeksen tot een substantiële overschatting van het aantal breuken kan leiden. Aangetoond wordt dat het noodzakelijk lijkt om eerst de heteroscedasticiteit uit deze reeksen te filteren alvorens de CUSUM toetsen toe te passen. Een uitgebreide Monte Carlo simulatie analyse toont aan dat wanneer de toetsen uitgevoerd worden op rendement reeksen die gestandaardiseerd zijn met behulp van een door een GARCH model verkregen schatting van de volatiliteit, dit leidt tot een gemiddelde correcte identificatie van het gemiddeld correct aantal structurele breuken. Eveneens wordt aangetoond dat het op deze manier toepassen van de CUSUM toetsen een robuuste aanpak blijkt als het volatiliteitsproces afwijkt van het veronderstelde GARCH proces. In dit hoofdstuk wordt tevens een algoritme ontwikkeld voor het sequentieel toepassen van de CUSUM toetsen met als doel het identificeren van meerdere structurele breuken.

Een empirische toepassing van het sequentiële algoritme op rendement reeksen van verschillende groei economieën bevestigt de theoretische eigenschappen van de CUSUM toetsen. Met het aangedragen toetsalgoritme en het toetsen op GARCH gefilterde rendementen worden aanzienlijk minder breuken in de volatiliteit gevonden voor de beschouwde groei economiën in vergelijking met eerdere studies.

Hoofdstuk 3 analyseert een niet-lineair model voor gerealiseerde volatiliteit. In het bijzonder wordt een niet-lineair autoregressive fractionally integrated model (ARFI) ontwikkeld met niveauveranderingen, dag-van-de-week effecten, effecten van nieuwsaankondigingen en leverage effecten. Het model omvat zowel structurele breuken als long-memory waardoor het effect van beide op volatiliteit onderzocht kan worden. Het volledige model, alsmede verscheidene gerestricteerde versies hiervan, worden geschat voor de volatiliteit van de S&P 500 index. Het niet-lineaire model leidt tot een betere beschrijving van de geobserveerde data en alle individuele niet-lineairiteiten zijn duidelijk significant. Het model levert voorspellingen op die, voor voorspelhorizons tot 20 dagen, beter zijn dan voorspellingen van een lineair ARFI model en eveneens beter dan die van GARCH-type modellen. Het toevoegen van de niet-leariteiten aan meer eenvoudige modellen voor gerealiseerde volatiliteit leidt tot soortgelijke verbeteringen in voorspelkracht.

Bij veel financiële toepassingen worden niet alleen schattingen van de volatiliteit vereist maar schattingen van de gehele covariantie matrix. Hierdoor zijn nauwkeurige schattingen voor covarianties en correlaties van even groot belang als voor volatiliteit. Hoofdstuk 4 richt zich op de voordelen van het gebruik van hoge frequentie intradag rendementen voor het meten en voorspellen van de dagelijkse covariantiematrix. In tegenstelling tot hoofdstuk 3 waarin alleen de populaire intradag frequentie van 5 minuten gebruikt wordt voor het construeren van de gerealiseerde volatiliteit, onderzoekt dit hoofdstuk rechtstreeks de optimale keuze voor de te gebruiken intradag frequentie. Deze optimale frequentie wordt vastgesteld door middel van het construeren van mean-variance efficiënte aandelenportefeuilles en het vervolgens evalueren van de performance van deze portefeuilles. De portefeuilles worden samengesteld uit de aandelen die samen de S&P 100 Index vormen. Ondanks het feit dat deze aandelen tot de klasse van meest liquide aandelen behoren blijkt de optimale intradag frequentie veel lager te liggen dan de populaire 5 minuten frequentie; eerder tussen 30 en 65 minuten. De gevonden resultaten worden aangetoond robuust te zijn voor verschillende transactiekosten niveaus en de frequentie waarmee de portefeuille geherbalanceerd wordt.

Deel B: Modelleren en voorspellen van de rente termijn structuur

In Deel B staat het modelleren van de rentetermijnstructuur centraal. In hoofdstuk 5 wordt een specifieke klasse van rentetermijnstructuur modellen beschouwd: de klasse van Nelson-Siegel modellen. Dit hoofdstuk analyseert hoe nauwkeurig deze modellen de termijnstructuur kunnen schatten alsmede hoe goed zij in staat zijn om toekomstige rentestanden te voorspellen. Verscheidene schattingstechnieken worden tevens onderzocht. De resultaten van hoofdstuk 5 tonen aan dat het waardevol is om het oorspronkelijke Nelson en Siegel (1987) model, welke de rentetermijnstructuur modelleert met behulp van drie factoren, uit te breiden met extra factoren waardoor het model beter in staat is om de rentecurve te benaderen. Er wordt tevens aangetoond dat een dergelijke uitbreiding, met name het toevoegen van een vierde factor, eveneens de voorspelkracht van het model verbetert. De voorspelkracht wordt daartoe vergeleken met verschillende benchmarkmodellen. Daarnaast worden voorspellingen in verschillende subperioden geanalyseerd.

Waar in hoofdstuk 5 de nadruk ligt op het identificeren van een enkel model dat in staat is om de termijnstructuur consistent nauwkeurig te voorspellen, wordt in hoofdstuk 6 voor een andere aanpak gekozen. In dit hoofdstuk wordt getracht toekomstige rentestanden te voorspellen aan de hand van een panel van verschillende modellen. In het bijzonder worden verschillende aandachtspunten beschouwd: parameteronzekerheid, modelonzekerheid en het gebruik van macro-economische informatie. Modellen met verschillende complexiteitsniveaus worden beoordeeld op hun voorspelkracht. Dit wordt gedaan door de voorspellingen van de verschillende modellen te vergelijken aan de hand van een 10-jaars voorspelperiode. Door de voorspelkracht te analyseren in verschillende subperioden wordt aangetoond dat de voorspelkracht sterk variëert door de tijd heen. Dit toont aan dat het een moeilijke opgave is om een enkel succesvol voorspelmodel te identificeren. De mogelijke oplossing hiervoor. welke aangedragen wordt in hoofdstuk 6, is om voorspellingen van meerdere modellen te combineren. Verschillende technieken om dit te doen worden onderzocht en er wordt aangetoond dat gecombineerde voorspellingen nauwkeurig en eveneens stabiel zijn door de tijd heen. Een methode waarbij modelgewichten gebaseerd worden op getoonde voorspelkracht in het verleden, lijkt in het bijzonder erg waardevol te zijn.

Bibliography

Aggarwal, R., C. Inclan, and R. Leal (1999), Volatility In Emerging Stock Markets, *Journal of Financial and Quantitative Analysis*, **34**, 33–55.

Aiolfi, M. and A. Timmermann (2006), Persistence in Forecasting Performance and Conditional Combination Strategies., *Journal of Econometrics*, **135**, 31–53.

Aït-Sahalia, Y., P. Mykland, and L. Zhang (2005), How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise, *Review of Financial Studies*, **18**, 351–416.

Andersen, T. G. and T. Bollerslev (1997), Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns, *Journal of Finance*, **52**, 975–1005.

Andersen, T. G. and T. Bollerslev (1998a), Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, **39**, 885–905.

Andersen, T. G. and T. Bollerslev (1998b), Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies, *Journal of Finance*, **53**, 219–265.

Andersen, T. G., T. Bollerslev, and F. X. . Diebold (2002), Parametric and Nonparametric volatility measurement, in Y. Aït-Sahalia and L. P. Hansen (eds.), *Handbook of Financial Econometrics*, Amsterdam: North Holland, *forthcoming.*

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001a), The Distribution of Realized Stock Return Volatility, *Journal of Financial Economics*, **61**, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001b), The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Association*, **96**, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003a), Modeling and Forecasting Realized Volatility, *Econometrica*, **71**, 579–625.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003b), Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange, *American Economic Review*, **93**, 38–62.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2004), Analytic Evaluation of Volatility Forecasts, *International Economic Review*, **45**, 1079–1110.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2005), Correcting the Errors: Volatility Forecast Evaluation Using on High-Frequency Data and Realized Volatilities, *Econometrica*, **73**, 279–296.

Andreou, E. and E. Ghysels (2002a), Detecting Multiple Breaks in Financial Market Volatility Dynamics, *Journal of Applied Econometrics*, **17**, 579–600.

Andreou, E. and E. Ghysels (2002b), Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results, *Journal of Business & Economic Statistics*, **20**, 363–376.

Andreou, E. and E. Ghysels (2003), Tests For Breaks in the Conditional Co-Movements of Asset Returns, *Statistica Sinica*, **13**, 1045–1074.

Andreou, E. and E. Ghysels (2004), The Impact of Sampling Frequency and Volatility Estimators on Change-Point Tests, *Journal of Financial Econometrics*, **2**, 290–318.

Andrews, D. W. K. (1991), Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, **59**, 817–858.

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.

Areal, N. M. P. C. and S. J. Taylor (2002), The Realized Volatility of FTSE-100 Futures Prices, *Journal of Futures Markets*, **22**, 627–648.

Azzalini, A. (1985), A Class of Distributions Which Includes the Normal Ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

Backus, D., S. Foresi, and C. Telmer (1999), Discrete-Time Models of Bond Pricing, in N. Jegadeesh and B. Tuckman (eds.), *Advanced Fixed-Income Valuation Tools*, Wiley and Sons.

Bai, J. (1997), Estimating Multiple Breaks One at a Time, *Econometric Theory*, **13**, 315–352.

Bai, J. (1999), Likelihood Ratio Tests for Multiple Structural Changes, *Journal of Econometrics*, **91**, 299–323.

Bai, J. and P. Perron (1998), Estimating and Testing Linear Models with Multiple Structural Changes, *Econometrica*, **66**, 47–78.

Bai, J. and P. Perron (2003), Computation and Analysis of Multiple Structural Change Models, *Journal of Applied Econometrics*, **18**, 1–22.

Baillie, R. and T. Bollerslev (1989), The Message in Daily Exchange Rates: A Conditional Variance Tale, *Journal of Business and Economic Statistics*, **7**, 279–305.

Baillie, R. T., T. Bollerslev, and H. Mikkelsen (1996), Fractionally Integrated Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics*, **14**, 3–30.

Bandi, F. M. and J. R. Russell (2005), Microstructure Noise, Realized Variance, and Optimal Sampling, *Working paper, Graduate School of Business, University of Chicago*.

Bandi, F. M. and J. R. Russell (2006), Separating Microstructure Noise from Volatility, *Journal of Financial Economics*, **79**, 655–692.

Bandi, F. M., J. R. Russell, and Y. Zhu (2008), Using High-Frequency Data in Dynamic Portfolio Choice, *forthcoming in Econometric Reviews*.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2004), Designing Realised Kernels to Measure the Ex-Post Variation in Equity Prices in the Presence of Noise, *Department of Economics Series, No. 264, University of Oxford*.

Barndorff-Nielsen, O. E. and N. Shephard (2002a), Econometric Analysis of Realised Volatility and its use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society*, **B 64**, 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2002b), Estimating Quadratic Variation Using Realized Variance, *Journal of Applied Econometrics*, **17**, 457–478.

Barndorff-Nielsen, O. E. and N. Shephard (2003), Realised Power Variation and Stochastic Volatility, *Bernoulli*, **9**, 243–265 and 1109–1111.

Barndorff-Nielsen, O. E. and N. Shephard (2004a), Econometric Analysis of Realised Covariation: High Frequency Covariance, Regression and Correlation in Financial Economics, *Econometrica*, **72**, 885–925.

Barndorff-Nielsen, O. E. and N. Shephard (2004b), How Accurate is the Asymptotic Approximation to the Distribution of Realised Volatility, in P. A. R. D. A. Andrews, J. Powell and J. H. Stock (eds.), *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg*, Cambridge: Cambridge University Press.

Barndorff-Nielsen, O. E. and N. Shephard (2004c), Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics*, **2**, 1–48.

Bekaert, G., C. B. Erb, C. R. Harvey, and T. E. Viskanta (1998), Distributional Characteristics of Emerging Market Returns and Asset Allocation, *Journal of Portfolio Management*, **24**, 102–116.

Bekaert, G. and C. R. Harvey (1997), Emerging Equity Market Volatility, *Journal of Financial Economics*, **43**, 29–77.

Bekaert, G. and C. R. Harvey (2000a), Capital Flows and the Behavior of Emerging Equity Market Returns, in S. Edwards (ed.), *Capital Flows and the Emerging Economies: Theory, Evidence and Controversies*, Chicago: University of Chicago Press, pp. 159–194.

Bekaert, G. and C. R. Harvey (2000b), Foreign Speculators and Emerging Equity Markets, *Journal of Finance*, **55**, 565–614.

Bekaert, G. and C. R. Harvey (2002), Research in Emerging Markets Finance: Looking to the Future, *Emerging Markets Review*, **3**, 429–448.

Bekaert, G. and C. R. Harvey (2003), Emerging Markets Finance, *Journal of Empirical Finance*, **10**, 3–55.

Bekaert, G., C. R. Harvey, and C. T. Lundblad (2003), Equity Market Liberalization in Emerging Markets, *Journal of Financial Research*, **26**, 275–299.

Beran, J. (1995), Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models, *Journal of the Royal Statistical Society*, **B 57**, 659–672.

Beran, J. (1999), SEMIFAR Models - A Semiparametric Framework for Modelling Trends, Long-Range Dependence and Stationarity, *Working paper, University of Konstanz*.

Beran, J. and D. Ocker (2001), Volatility of Stock Market Indices - An Analysis Based on SEMIFAR Models, *Journal of Business & Economic Statistics*, **19**, 103 – 116.

Berkes, I. and L. Horváth (2003), Limit Results for the Empirical Process of Squared Residuals in GARCH(p,q) Models, *Stochastic Processes and their Applications*, **105**, 271–298.

Berkes, I., L. Horváth, and P. Kokoszka (2003), GARCH(p,q) Processes: Structure and Estimation, *Bernoulli*, **9**, 201–227.

Bernadell, C., J. Coche, and K. Nyholm (2005), Yield Curve Prediction for the Strategic Investor, *ECB Working paper series*, **472**.

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.

Bikbov, R. and M. Chernov (2005), No-Arbitrage Macroeconomic Determinants of the Yield Curve, *Working paper, Columbia University*.

BIS (2005), Zero-Coupon Yield Curves: Technical Documentation, *Bank for International Settlements, Basle*.

Björk, T. and B. Christensen (1999), Interest Rate Dynamics and Consistent Forward Rate Curves, *Mathematical Finance*, 323–348.

Black, F. (1976), Studies of Stock Market Volatility Changes, *Proceedings of the American Statistical Association, Buinsess and Economic Statistics Section*, 177–181.

Bliss, R. R. (1997), Testing Term Structure Estimation Methods, *Advances in Futures and Options Research*, **9**, 197–231.

Bolder, D. and D. Stréliski (1999), Yield Curve Modelling at the Bank of Canada, *Bank of Canada Technical Report No. 84*.

Bollerslev, T. (1986), Generalized Conditional Autoregressive Heteroscedasticity, *Journal of Econometrics*, **31**, 307–327.

Bollerslev, T., J. Litvinova, and G. Tauchen (2006), Leverage and Volatility Feedback Effects in High-Frequency Data, *Journal of Financial Econometrics*, **4**, 353–384.

Bollerslev, T. and J. H. Wright (2000), Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data, *Journal of Econometrics*, **98**, 81–106.

Bollerslev, T. and B. Y. B. Zhang (2003), Measuring and Modeling Systematic Risk in Factor Pricing Models Using High-Frequency Data, *Journal of Empirical Finance*, **10**, 533–558.

Bomfim, A. N. (2003a), Monetary Policy and the Yield Curve, *FEDS Working Paper No. 2003-15*.

Bomfim, A. N. (2003b), Pre-Announcement Effects, News Effects, and Volatility; Monetary Policy and the Stock Market, *Journal of Banking & Finance*, **27**, 133–151.

Boswijk, H. P. (2004), Asymptotic Theory for Integrated Processes, manuscript in preparation, Oxford University Press, parts available at `http://www.fee.uva.nl/ke/boswijk/book.pdf`.

Caks, J. (1977), The Coupon Effect on Yield to Maturity, *Journal of Finance*, **32**, 103–115.

Carter, C. and R. Kohn (1994), On Gibbs Sampling for State-Space Models, *Biometrika*, **81**, 541–553.

Chambers, D. R., W. T. Carleton, and D. W. Waldman (1984), A New Approach to Estimation of the Term Structure of Interest Rates, *Journal of Financial and Quantitative Analysis*, **19**, 233–252.

Chan, L. K. C., J. Karceski, and J. Lakonishok (1999), On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model, *Review of Financial Studies*, **12**, 937–974.

Chen, J. and A. K. Gupta (1997), Testing and Locating Variance Change-Points with Application to Stock Prices, *Journal of the American Statistical Association*, **92**, 739–747.

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.

Christensen, B. J. and N. R. Prabhala (1998), The Relation Between Implied and Realized Volatility, *Journal of Financial Economics*, **50**, 125–150.

Christoffersen, P. F. (1998), Evaluating Interval Forecasts, *International Economic Review*, **39**, 841–862.

Christoffersen, P. F. and F. X. Diebold (1998), Cointegration and Long-Horizon Forecasting, *Journal of Business & Economic Statistics*, **16**, 450–458.

Chu, C. S. J. (1995), Detecting Parameter Shifts in GARCH Models, *Econometric Theory*, **14**, 241–266.

Cohen, K. J., G. A. Hawawini, S. F. Maier, R. A. Schwartz, and D. K. Whitcomb (1983), Friction in the Trading Process and the Estimation of Systematic Risk, *Journal of Financial Economics*, **12**, 263–278.

Colacito, R. and R. F. Engle (2006), Testing and Valuing Dynamic Correlations for Asset Allocation, *Journal of Business & Economic Statistics*, **24**, 238–253.

Corsi, F. (2004), A Simple Long Memory Model of Realized Volatility, *Working paper, University of Southern Switserland*.

Cox, J., J. E. Ingersoll, and S. A. Ross (1985), A Theory of the Term Structure of Interest Rates, *Econometrica*, **53**, 385–407.

Croushore, D. (2006), Forecasting with Real-Time Macroeconomic Data, in G. Elliot, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1, Elsevier, North-Holland.

Dahlquist, M. and L. E. O. Svensson (1996), Estimating the Term Structure of Interest Rates for Monetary Policy Analysis, *Scandinavian Journal of Economics*, **98**, 163–183.

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.

Dawid, A. P. (1984), Statistical Theory: The Prequential Approach, *Journal of the Royal Statistical Society Series A*, **147**, 278–292.

De Santis, G. and S. Imrohoroglu (1997), Stock Returns and Volatility in Emerging Financial Markets, *Journal of International Money and Finance*, **16**, 561–579.

den Haan, W. J. and A. Levin (1997), A Practioner's Guide to Robust Covariance Matrix Estimation, in C. R. Rao and G. S. Maddala (eds.), *Handbook of Statistics - Vol. 15*, Elsevier Science, Amsterdam, pp. 291–341.

De Jong, F. (2000), Time Series and Cross-section Information in Affine Term-Structure Models, *Journal of Business & Economic Statistics*, **18**, 300–314.

De Jong, F. and T. Nijman (1997), High Frequency Analysis of Lead-Lag Relationships between Financial Markets, *Journal of Empirical Finance*, **4**, 259–277.

De Pooter, M. (2007), Examining the Nelson-Siegel Class of Term Structure Models: Insample fit versus out-of-sample forecasting performance, *Tinbergen Institute Discussion paper. No. 2007-043/4*.

De Pooter, M., M. Martens, and D. van Dijk (2008), Predicting the Daily Covariance Matrix for S&P100 Stocks Using Intraday Data - But which frequency to use?, *forthcoming in Econometric Reviews*.

De Pooter, M., F. Ravazzolo, and D. van Dijk (2007), Predicting the Term Structure of Interest Rates: Incorporating parameter uncertainty, model uncertainty and macroeconomic information, *Tinbergen Institute Discussion paper. No. 2007-028/4*.

De Pooter, M. and D. van Dijk (2004), Testing for Changes in Volatility in Heteroscedastic Time Series - a further examination, *Econometric Insitute research report, 2004-38/A*.

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. (1986), Modeling the Persistence of Conditional Variances: A Comment, *Econometric Reviews*, **5**, 51–56.

Diebold, F. X. and A. Inoue (2001), Long Memory and Structural Change, *Journal of Econometrics*, **105**, 131–159.

Diebold, F. X., L. Ji, and C. Li (2006a), A Three-Factor Yield Curve Model: Non-Affine Structure, Systematic Risk Sources, and Generalized Duration, in L. Klein (ed.), *Macroeconomics, Finance and Econometrics: Essays in Memory of Albert Ando*, Cheltenham, U.K., 240-274.

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., M. Piazzesi, and G. D. Rudebusch (2005), Modeling Bond Yields in Finance and Macroeconomics, *American Economic Review*, **95**, 415–420.

Diebold, F. X., G. D. Rudebusch, and B. Aruoba (2006b), The Macroeconomy and the Yield Curve: a Dynamic Latent Factor Approach, *Journal of Econometrics*, **131**, 309–338.

Dolan, C. P. (1999), Forecasting the Yield Curve Shape: Evidence in Global Markets, *Journal of Fixed Income*, 92–99 (June).

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.

Ebens, H. (1999), Realized Stock Volatility, *unpublished manuscript, Johns Hopkins University*.

Ederington, L. H. and J. H. Lee (1993), How Markets Process Information: News Releases and Volatility, *Journal of Finance*, **48**, 1161–1191.

Eklund, J. and S. Karlsson (2007), Forecast Combination and Model Averaging using Predictive Measures, *Econometric Reviews*, **26**, 329–363.

Engle, R. F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50**, 987–1007.

Engle, R. F. and V. K. Ng (1993), Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, **48**, 1749–1778.

Engle, R. F. and K. Sheppard (2001), Theoretical and Empirical Properties of Dynamic Conditional Correlation MVGARCH, *UCSD Working Paper No. 2001-15*.

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**.

Fabozzi, F. J., L. Martellini, and P. Priaulet (2005), Predictability in the Shape of the Term Structure of Interest Rates, *Journal of Fixed Income*, 40–53 (June).

Fama, E. F. and R. R. Bliss (1987), The Information in Long-Maturity Forward Rates, *American Economic Review*, **77**, 680–692.

Ferguson, R. and S. Raymar (1998), A Comparative Analysis of Several Popular Term Structure Estimation Models, *Journal of Fixed Income*, 17–33.

Fleming, J., C. Kirby, and B. Ostdiek (2001), The Economic Value of Volatility Timing, *Journal of Finance*, **56**, 329–352.

Fleming, J., C. Kirby, and B. Ostdiek (2003), The Economic Value of Volatility Timing Using "Realized" Volatility, *Journal of Financial Economics*, **67**, 473–509.

Foster, D. P. and D. N. Nelson (1996), Continuous Record Asymptotics for Rolling Sample Variance Estimators, *Econometrica*, **64**, 139–174.

Franses, P. H., M. van der Leij, and R. Paap (2002), Modelling and Forecasting Level Shifts in Absolute Returns, *Journal of Applied Econometrics*, **17**, 601–616.

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.

Geweke, J. and C. Whiteman (2006), Bayesian Forecasting, in G. Elliot, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, North-Holland, 3-80.

Gimeno, R. and J. M. Nave (2006), Genetic Algorithm Estimation of Interest Rate Term Structure, *Bank of Spain Working Paper No. 0636*.

Giot, P. and S. Laurent (2004), Modelling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models, *Journal of Empirical Finance*, **11**, 379–398.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993), On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Financ*, **48**, 1779–1801.

Granger, C. W. J. and N. Hyung (1999), Occasional Structural Breaks and Long Memory, discussion paper 99-14, University of California at San Diego.

Guidolin, M. and A. Timmermann (2007), Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach, *forthcoming in Journal of Econometrics*.

Gürkanyak, R. S., B. Sack, and J. H. Wright (2006), The U. S. Treasury Yield Curve: 1961 to the Present, *Finance and Economics Discussion Series No 2006-28, Federal Reserve Board, Washington, D.C.*.

Hafner, C. M. and P. H. Franses (2003), A Generalized Dynamic Conditional Correlation Model for Many Assets, *Econometric Institute Report No. 2003-18*.

Hansen, P. R. and A. Lunde (2005), A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data, *Journal of Financial Econometrics*, **3**, 525–554.

Hansen, P. R. and A. Lunde (2006), Realized Variance and Market Microstructure Noise, *Journal of Business and Economic Statistics*, **24**, 127–218.

Hansen, P. R., A. Lunde, and J. M. Nason (2003), Choosing the Best Volatility Model: The Model Confidence Set Approach, *Oxford Bulletin of Economics and Statistics*, **65**, 839–861.

Hansen, P. R., A. Lunde, and J. M. Nason (2005a), Model Confidence Sets for Forecasting Models, *Federal Reserve Bank of Atlanta, Working paper 2005-7*.

Hansen, P. R., A. Lunde, and V. Voev (2005b), Realized Covariance Estimation Using High-Frequency Data in the Presence of Noise, *working paper, Stanford University*.

Harris, F. H., T. H. McInish, G. L. Shoesmith, and R. A. Wood (1995), Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets, *Journal of Financial and Quantitative Analysis*, **30**, 563–579.

Harvey, C. R. and R. D. Huang (1991), Volatility in the Foreign Currency Futures Market, *Review of Financial Studies*, **4**, 543–569.

Harvey, D. I., S. J. Leybourne, and P. Newbold (1998), Tests for Forecast Encompassing, *Journal of Business and Economic Statistics*, **16**, 254–259.

Hayashi, T. and N. Yoshida (2005), On Covariance Estimation of Non-Synchronously Observed Diffusion Processes, *Bernoulli*, **11**, 359–379.

Hordahl, P., O. Tristani, and D. Vestin (2006), A Joint Econometric Model of Macroeconomic and Term-Structure Dynamics, *Journal of Econometrics*, **131**, 405–444.

Horváth, L., P. Kokoszka, and G. Teyssière (2001), Empirical Process of the Squared Residuals of an ARCH Sequence, *Annals of Statistics*, **29**, 445–469.

Hsu, D. A. (1977), Tests for Variance Shift at an Unknown Time Point, *Applied Statistics*, **26**, 279–284.

Huse, C. (2007), Term Structure Modelling with Observable State Variables, *Working paper, London School of Economics*.

Inclán, C. and G. Tiao (1994), Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance, *Journal of the American Statistical Association*, **89**, 913–923.

Jagannathan, R. and T. Ma (2003), Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps, *Journal of Finance*, **58**, 1651–1683.

Jeffrey, A., O. Linton, and T. Nguyen (2006), Flexible Term Structure Estimation: Which Method is Preferred?, *Metrika*, 1–24.

Jegadeesh, N. and S. Titman (1993), Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance*, **48**, 65–91.

Jorion, P. (1995), Predicting Volatility in the Foreign Exchange Market, *Journal of Finance*, **50**, 507–528.

Kim, C.-J. and C. R. Nelson (1999), *State-Space Models with Regime Switching*, MIT Press, Cambridge, Massachusetts.

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, *Finance and Economics Discussion Series, No 2005-33, Federal Reserve Board, Washington, D.C.*.

Kim, E. H. and V. Singal (2000), Opening Up of Stock Markets: Lessons from Emerging Economies, *Journal of Business*, **73**, 25–66.

Kim, S., S. Cho, and S. Lee (2000), On the CUSUM Test for Parameter Changes in GARCH(1,1) Models, *Communications in Statistics. Theory and Methods*, **29**, 445–462.

Kokoszka, P. and R. Leipus (2000), Change-point Estimation in ARCH Models, *Bernoulli*, **6**, 513–539.

Kokoszka, P. and G. Teyssière (2002), Change-Point Detection in GARCH Models: Asymptotic and Bootstrap Tests, working paper, Utah State University.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley & Sons Ltd, West Sussex, England.

Koopman, S. J., B. Jungbacker, and E. Hol (2005), Forecasting Daily Variability of the S&P 100 Stock Index Using Historical Realised and Implied Volatility Measurements, *Journal of Empirical Finance*, **12**, 445–475.

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*.

Lamoureux, C. G. and W. D. Lastrapes (1990), Persistence in Variance, Structural Change, and the GARCH Model, *Journal of Business and Economic Statistics*, **8**, 225–234.

Lavielle, M. and E. Moulines (2000), Least-Squares Estimation of an Unknown Nummber of Shifts in a Time Series, *Journal of Time Series Analysis*, **21**, 33–59.

Lee, S. and S. Park (2001), The CUSUM of Squares Test for Scale Changes in Infinite Order Moving Order Average Processes, *Scandinavian Journal of Statistics*, **28**, 625–644.

Lee, S., Y. Tokutsu, and K. Maekawa (2003), The Residual CUSUM Test for the Constancy of Parameters in GARCH(1,1) Models, working paper, Seoul National University.

Li, K. (2002), Long Memory Versus Option-Implied Volatility Predictions, *Journal of Derivatives*, **9**, 9–25.

Linton, O., E. Mammen, J. P. Nielsen, and C. Tanggaard (2001), Yield Curve Estimation by Kernel Smoothing Methods, *Journal of Econometrics*, **105**, 185–223.

Litterman, R. and J. Scheinkman (1991), Common Factors Affecting Bond Returns, *Journal of Fixed Income*, 54–61.

Liu, M. (2000), Modeling Long Memory in Stock Market Volatility, *Journal of Econometrics*, **99**, 139–171.

Liu, Q. (2004), On Portfolio Optimization: How do we Benefit from High-Frequency Data, *Working paper, University of Hawaii.*

Lo, A. W. and A. C. MacKinlay (1990), An Econometric Analysis of Nonsynchronous Trading, *Journal of Econometrics*, **45**, 181–212.

Lopez, J. A. (1999), Methods for Evaluating Value-at-Risk Estimates, *Federal Reserve Bank of San Francisco*, **2**, 3–17.

Lundbergh, S. and T. Teräsvirta (2002), Evaluating GARCH Models, *Journal of Econometrics*, **110**, 417–435.

MacKinnon, J. G. (2000), Computing Numerical Distribution Functions in Econometrics, in A. Pollard, D. Mewhort, and D. Weaver (eds.), *High Performance Computing Systems and Applications*, Kluwer, Amsterdam, pp. 455–470.

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.

Maheu, J. M. and T. H. McCurdy (2002), Nonlinear Features of Realized FX Volatility, *Review of Economics and Statistics*, **84**, 668–681.

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.

Martens, M. (2002), Measuring and Forecasting S&P 500 Index-Futures Volatility Using High-Frequency Data, *Journal of Futures Markets*, **22**, 497–518.

Martens, M. (2004), Estimating Unbiased and Precise Realized Covariances, *Working paper, Erasmus University Rotterdam*.

Martens, M., D. van Dijk, and M. de Pooter (2004), Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity, *Tinbergen Institute Discussion Paper No. 067/4*.

Martens, M. and J. Zein (2004), Predicting Financial Volatility: High-Frequency Time-Series Forecasts Vis-à-Vis Implied Volatility, *Journal of Futures Markets*, **24**, 1005–1028.

McCulloch, J. H. (1975), The Tax-Adjusted Yield Curve, *Journal of Finance*, **30**, 811–830.

Merton, R. C. (1980), On Estimating the Expected Return on the Market; An Exploratory Investigation, *Journal of Financial Economics*, **8**, 323–361.

Mikosch, T. and C. Starica (2004), Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects, *Review of Economics and Statistics*, **86**, 378–390.

Mönch, E. (2006a), Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach, *Working paper, Humboldt University Berlin*.

Mönch, E. (2006b), Term Structure Surprises: The Predictive Content of Curvature, Level, and Slope.pdf, *Working paper, Humboldt University Berlin*.

Nelson, C. R. and A. F. Siegel (1987), Parsimonious Modeling Of Yield Curves, *Journal of Business*, **60**, 473–489.

Newey, W. K. and K. D. West (1987), A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, **55**, 703–708.

Oomen, R. C. A. (2002), Modelling Realized Variance when Returns are Serially Correlated, *Working paper, University of Warwick*.

Pagan, A. R. and G. W. Schwert (1990), Alternative Models for Conditional Stock Volatility, *Journal of Econometrics*, **45**, 267–290.

Patton, A. J. (2006), Volatility Forecast Comparison Using Imperfect Volatility Proxies, *Working paper, London School of Economics*.

Piazzesi, M. (2003), Affine Term Structure Models, in Y. Ait-Sahalia and L. Hansen (eds.), *Handbook of Financial Econometrics*, Elsevier, North-Holland, *forthcoming.*

Politis, D. N. and J. P. Romano (1994), The Stationary Bootstrap, *Journal of the American Stastical Association*, **89**, 1303–1313.

Pong, S., M. B. Shackleton, S. J. Taylor, and X. Xu (2004), Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models, *Journal of Banking and Finance*, **28**, 2541–2563.

Richards, A. J. (1996), Volatility and Predictability in National Stock Markets: How Do Emerging and Mature Markets Differ?, *International Monetary Fund Staff Papers*, **43**, 461–501.

Rouwenvorst, K. G. (1998), International Momentum Strategies, *Journal of Finance*, **53**, 267–284.

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**.

Sack, B. (2000), Using Treasury STRIPS to Measure the Yield Curve, *Working paper, Federal Reserve Board, Washington, D.C.*.

Sakata, T. (1988), Detecting a Change in Variances, *Communications in Statistics. Theory and Methods*, **17**, 641–655.

Scholes, M. and J. Williams (1977), Estimating Betas from Nonsynchronous Data, *Journal of Financial Economics*, **5**, 309–327.

Sheppard, K. (2006), Realized Covariance and Scrambling, *Working paper, University of Oxford*.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

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), Combining 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*.

Svensson, L. E. O. (1994), Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994, *NBER Working Paper Series*, **4871**.

Talwar, P. P. and J. E. Gentle (1981), Detecting a Scale Shift in a Random Sequence at an Unknown Time Point, *Applied Statistics*, **30**, 301–304.

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.

Taylor, S. (1986), *Modelling Financial Time Series*, John Wiley, New York.

Basle Committee on Banking Supervision (1996), Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements, *Bank for International Settlements Publication No. 22*.

Thomakos, D. D. and T. Wang (2003), Realized Volatility in the Futures Markets, *Journal of Empirical Finance*, **10**, 321–353.

Timmermann, A. (2006), Forecast Combinations, in G. Elliot, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, North-Holland, 135-196.

Vasicek, O. A. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, **5**, 177–188.

Vasicek, O. A. and H. G. Fong (1982), Term Structure Modeling Using Exponential Splines, *Journal of Finance*, **37**, 339–348.

West, K. D. (2006), Forecast Evaluation, in G. Elliot, C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, North-Holland, 99-134.

White, H. (2000), A Reality Check for Data Snooping, *Econometrica*, **68**, 1097–1126.

Wichern, D. W., R. B. Miller, and D. A. Hsu (1976), Changes of Variance in First-Order Autoregressive Time Series Models, *Applied Statistics*, **25**, 248–256.

Wu, T. (2005), Macro Factors and the Affine Term Structure of Interest Rates, *forthcoming in Journal of Money, Credit and Banking*.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Zhang, L. (2006), Estimating Covariation: Epps Effect and Microstructure Noise, *Working paper, University of Illinois at Chicago*.

Zhang, L., P. Mykland, and Y. Aït-Sahalia (2005), A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, *Journal of the American Statistical Association*, **100**, 1394–1411.

Zhou, B. (1996), High-frequency data and volatility in foreign-exchange rates, *Journal of Business and Economic Statistics*, **14**, 45–52.

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:

364. M. ABREU, *Spatial determinants of economic growth and technology diffusion.*
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*.